

Barbara Olbrych

STATYSTYKA MATEMATYCZNA
PRZYKŁADY

Radom 2023

Recenzent:
Prof. dr hab. inż. Zbigniew Kosma

Procedura recenzowania jest zgodna z wytycznymi Ministerstwa Nauki i Szkolnictwa Wyższego, które zawarto w broszurze *Dobre praktyki w procedurach recenzyjnych w nauce*, Warszawa 2011.

Redakcja językowa: Barbara Jaworska

Projekt okładki: Dagmara Zym

Skład i łamanie: Cezary Majewski

© Copyright by Akademia Handlowa Nauk Stosowanych w Radomiu,
Radom 2023

e - ISBN : 978-83-62491-75-9

Wydawnictwo
Akademii Handlowej Nauk Stosowanych w Radomiu
ul. Mazowieckiego 7a
26-600 Radom
Tel. [48] 360-10-75
e-mail: wydawnictwo@ahns.pl
www.ahns.pl

*Istnieje w naturze wiele zagadnień,
których nie można ani dostrzec z wystarczającą dokładnością,
ani przedstawić odpowiednio jasno, ani wykorzystać dostatecznie zręcznie
bez pomocy matematyki*

Franciszek Bacon

SPIS TREŚCI

Wstęp	6
Rozdział I. Elementy kombinatoryki i rachunku prawdopodobieństwa	8
1.1. Kombinatoryka	8
1.2. Zdarzenia losowe	15
1.3. Rachunek prawdopodobieństwa	20
1.4. Powtórzenie w przykładach	29
Rozdział II. Zmienne losowe	34
2.1. Podstawowe pojęcia	34
2.2. Zmienne losowe skokowe (dyskretne).....	36
2.3. Zmienne losowe ciągłe.....	43
2.4. Parametry rozkładu zmiennej losowej	48
2.4.1. Parametry klasyczne	48
2.4.2. Parametry pozycyjne.....	56
2.5. Powtórzenie w przykładach	58
Rozdział III. Wybrane rozkłady zmiennej losowej	67
3.1. Rozkłady skokowej zmiennej losowej.....	67
3.1.1. Rozkład zero-jedynkowy	67
3.1.2. Rozkład dwumianowy	68
3.1.3. Rozkład Poissona.....	71
3.2. Rozkłady ciągłej zmiennej losowej	73
3.2.1. Rozkład jednostajny.....	74
3.2.2. Rozkład normalny.....	76
3.2.3. Rozkład wykładniczy.....	83
3.2.4. Rozkład χ^2 (chi-kwadrat)	85
3.2.5. Rozkład t-Studenta.....	88
3.3. Nierówność Czebyszewa	91
3.4. Zmienna losowa dwuwymiarowa	92
3.4.1. Rozkłady brzegowe i rozkłady warunkowe.....	94
3.4.2. Niezależność zmiennych losowych X i Y	95
3.5. Powtórzenie w przykładach.....	96
Rozdział IV. Twierdzenia graniczne	102
4.1. Twierdzenia lokalne.....	102
4.2. Twierdzenia integralne.....	104
4.3. Prawa wielkich liczb	107
Rozdział V. Statystyki i ich rozkłady	110
5.1. Sposoby doboru próby	110
5.2. Średnia z próby n-elementowej.....	116
5.3. Wariancja i odchylenie z próby n-elementowej.....	117

5.4. Statystyka Studenta z próby	120
5.5. Rozkład frakcji z próby.....	121
Rozdział VI. Estymacja parametrów	125
6.1. Estymacja parametrów populacji generalnej	125
6.1.1. Estymacja punktowa	129
6.1.2. Estymacja przedziałowa	131
6.2. Powtórzenie w przykładach	144
Rozdział VII. Weryfikacja hipotez statystycznych	151
7.1. Hipotezy parametryczne	152
7.2. Hipotezy nieparametryczne.....	161
7.3. Powtórzenie w przykładach	166
Rozdział VIII. Wybrane pojęcia statystyki matematycznej.....	171
Zadania do samodzielnego rozwiązania	177
Bibliografia	186
Tablice statystyczne.....	187

WSTĘP

Statystyka znajduje coraz szersze zastosowanie prawie we wszystkich dziedzinach wiedzy. Informacje na temat zjawisk i procesów społeczno-ekonomicznych powinny być dostarczane odbiorcom precyzyjnie, a to nie tylko zachęca, ale wręcz zmusza do korzystania z metod analizy ilościowej.

Statystyka matematyczna stanowi podstawę uogólnień wyników otrzymywanych z badań próby losowej na całą populację generalną, z której ta próba pochodzi. Proces uogólniania nosi nazwę wnioskowania statystycznego. Wnioskowanie statystyczne musi być zatem oparte o wyniki uzyskane z badań reprezentatywnej próby losowej, tj. takiej, której struktura ze względu na interesujące badacza cechy zbliżona jest do struktury populacji generalnej. Próbę losową tworzą elementy populacji generalnej pobrane z niej według pewnego modelu postępowania zwanego schematem losowania. Znajomość statystyki matematycznej pozwala na uniknięcie błędu, polegającego na wyciągnięciu fałszywych wniosków z prawdziwych danych statystycznych.

Celem, jaki przyświeca tej inicjatywie wydawniczej jest dostarczenie studentom i praktykom opracowania, które zawiera podstawowe elementy teorii oraz zbiór gotowych przykładów, bogato ilustrowanych. Książka ma charakter podręcznika akademickiego. W podręczniku, który właśnie trafia do rąk czytelnika zamieszczono – oprócz teorii „w pigułce” - znaczną liczbę przykładów, całkowicie rozwiązanych i opatrzonych obszernymi komentarzami. Przedmiot ten jest raczej trudny, dlatego podręcznik został tak napisany, aby mógł go czytać każdy student i praktyk.

Na ogół nie ma kłopotów z brakiem podręczników, ale trudnością jest jednak znalezienie odpowiednich materiałów w formie przykładów. Aby zdobyć wprawę w posługiwaniu się narzędziami statystycznymi należy rozwiązać wiele zadań, posługując się przykładami, które mogą pomóc w zrozumieniu „o co chodzi”. Do opisu zagadnień naukowych często stosuje się symbole matematyczne, korzystając z bogatego i trudnego dla studentów aparatu matematyki. Z uwagi na pracochłonność obliczeń w przykładach i zadaniach występuje niewielka liczba obserwacji. Zrezygowano z pokazywania czasochłonnych technik obliczeniowych ze względu na dostępność komputerowych pakietów statystycznych i arkuszy kalkulacyjnych, zawierających programy komputerowe.

Aby zrozumieć metody statystyczne, trzeba posiadać wiedzę z rachunku prawdopodobieństwa stanowiącego kanwę wnioskowania statystycznego, dlatego pierwszy spośród ośmiu rozdziałów zawiera elementy kombinatoryki i rachunku prawdopodobieństwa. Rozdział drugi ujmuje zagadnienia

związane z pojęciem zmiennej losowej, zaś trzeci prezentuje wybrane rozkłady zmiennych losowych. Kolejne rozdziały wprowadzają do najważniejszych zastosowań praktycznych, występujących jako szacowanie (estymacja) parametrów rozkładu badanej cechy w populacji generalnej oraz jako sprawdzanie (weryfikacja) hipotez statystycznych.

Dynamiczny rozwój statystyki, ciągle wzbogacającej swe metody skłania do tego, aby poznać ze zrozumieniem teorię zmiennej losowej oraz podstawy estymacji i weryfikacji hipotez statystycznych. Należyty efekt dydaktyczny daje analiza różnorodnych treści i prostych rachunkowo zadań. Zawarte w książce przykłady o treści zaczerpniętej z różnych nauk powinny spowodować szeroką przydatność opracowania. Bogaty zbiór metod statystycznych zmusza autora do dokonania wyboru, który nie jest łatwy. Rezygnacja z głębszych wywodów teoretycznych wynika z ograniczonej objętości i skupienia uwagi przede wszystkim na przykładach. Metody statystyczne stanowią narzędzie, które pozwala opisać otaczającą nas rzeczywistość, ale przede wszystkim pomóc w jej zrozumieniu.

Przy opracowaniu niniejszego wydania podręcznika akademickiego została wykorzystana książka „Wnioskowanie statystyczne w przykładach”, napisana przez Autorkę w 2008 roku, uzupełniona. Opinię o niniejszej publikacji wyda, jak zwykle, rzeczywistość.

Barbara Olbrych

ROZDZIAŁ I
ELEMENTY KOMBINATORYKI I RACHUNKU
PRAWDOPODOBIENSTWA

1.1. Kombinatoryka

1. Zbiór składający się z n elementów uporządkowanych i różnych nazywamy *permutacją bez powtórzeń* z n elementów.

$$P_n = n! = 1 \cdot 2 \cdot 3 \cdot \dots \cdot n \quad (\text{przyjmujemy, że } 0! = 1)$$

2. Zbiór składający się z n elementów uporządkowanych, wśród których pewne elementy powtarzają się odpowiednio n_1, n_2, \dots, n_k razy nazywamy n -elementową *permutacją z powtórzeniami*.

$$P_n^{n_1, n_2, \dots, n_k} = \frac{n!}{n_1! \cdot n_2! \cdot \dots \cdot n_k!}$$

3. *Wariacją bez powtórzeń* z n -elementowego zbioru po k elementów nazywamy uporządkowany zbiór składający się z k różnych elementów wybranych spośród n różnych elementów.

$$V_n^k = \frac{n!}{(n-k)!}$$

4. *Wariacją z powtórzeniami* z n elementów po k nazywamy uporządkowany zbiór składający się z k elementów różnych lub nie różniących się między sobą, wybranych spośród n różnych elementów.

$$\bar{V}_n^k = n^k$$

5. *Kombinacją bez powtórzeń* z n elementów po k nazywamy uporządkowany zbiór składający się z k różnych elementów wybranych spośród n różnych elementów, przy czym obojętne jest, w jakim porządku elementy te są rozmieszczone.

$$C_n^k = \binom{n}{k} = \frac{n!}{k!(n-k)!}$$

6. *Kombinacją z powtórzeniami* z n elementów po k nazywamy zbiór składający się z k elementów, różnych lub nie różniących się między sobą, wybranych spośród n różnych elementów, przy czym obojętne jest, w jakim porządku elementy te są rozmieszczone.

$$\bar{C}_n^k = \binom{n+k-1}{k} \text{ lub } \bar{C}_n^k = \binom{n+k-1}{n-1}$$

Przykład 1.1.1

Ile liczb czterocyfrowych mniejszych od 2000 można utworzyć z cyfr 1, 2, 3, 4?

Rozwiązanie

Cyfra 1 musi być na pierwszym miejscu, pozostałe permutują, a zatem utworzymy $3! = 1 \cdot 2 \cdot 3 = 6$ liczb.

Przykład 1.1.2

Baletnice wychodzą na scenę gęsiego. Ile jest wszystkich możliwych ustawień przy wejściu na scenę pięciu baletnic?

Rozwiązanie

Baletnice mogą wejść pojedynczo na scenę na tyle sposobów, ile istnieje permutacji bez powtórzeń zbioru 5-elementowego, czyli na $5! = 1 \cdot 2 \cdot 3 \cdot 4 \cdot 5 = 120$ sposobów.

Przykład 1.1.3

Malarz otrzymał zlecenie pomalowania czterech różnych lokali używając czterech różnych kolorów farb. Na ile sposobów może to uczynić? Zakładamy, że każdy lokal pomalowany będzie na jeden kolor.

Rozwiązanie

Malarz może pomalować lokale na tyle sposobów, ile istnieje permutacji bez powtórzeń zbioru 4-elementowego, a więc na $4! = 1 \cdot 2 \cdot 3 \cdot 4 = 24$ sposoby.

Przykład 1.1.4

Iloma sposobami można rozmieścić na 9 numerowanych miejscach 3 „orły” i 5 „reszek”?

Rozwiązanie

Ze zbioru 9-elementowego tworzymy zbiory 9-elementowe, w których porządek odgrywa rolę i w których elementy odpowiednio powtarzają się. Korzystamy ze wzoru na permutację z powtórzeniami.

$$P_9^{3,5} = \frac{9!}{3! \cdot 5!} = \frac{5! \cdot 6 \cdot 7 \cdot 8 \cdot 9}{6 \cdot 5!} = 504.$$

3 „orły” i 5 „reszek” możemy rozmieścić na 9 numerowanych miejscach na 504 sposoby.

Przykład 1.1.5

Ile różnych wyrazów mających sens lub nie można utworzyć przedstawiając wszelkimi sposobami litery w wyrazie STATYSTYKA?

Rozwiązanie

Tworzymy zbiory 10-elementowe ze zbioru 10-elementowego, bo mamy 10 liter. Kolejność elementów (liter) odgrywa rolę, niektóre elementy powtarzają się: litera S – 2 razy, A – 2 razy, Y – 2 razy, T – 3 razy. Otrzymamy więc tyle różnych wyrazów, ile jest permutacji z powtórzeniami:

$$P_{10}^{2,2,2,3} = \frac{10!}{2! \cdot 2! \cdot 2! \cdot 3!} = \frac{3! \cdot 4 \cdot 5 \cdot 6 \cdot 7 \cdot 8 \cdot 9 \cdot 10}{8 \cdot 3!} = 75600.$$

Możemy utworzyć 75 600 różnych wyrazów.

Przykład 1.1.6

Na campingu jest 12 jednoosobowych wolnych pomieszczeń do wynajęcia. Przybyła grupa czterech turystów. Na ile sposobów można ich rozmieścić?

Rozwiązanie

Bazujemy na 12-elementowym zbiorze. Czterech turystów można rozmieścić na tyle sposobów, ile jest różnych wariacji bez powtórzeń z 12 elementów po 4:

$$V_{12}^4 = \frac{n!}{(n-k)!} = \frac{12!}{(12-4)!} = \frac{8! \cdot 9 \cdot 10 \cdot 11 \cdot 12}{8!} = 11880.$$

Turyści mogą zająć pomieszczenia na 11 880 sposobów.

Przykład 1.1.7

Przedział w pociągu osobowym ma ponumerowane miejsca od 1 do 4 na dwóch kanapach ustawionych naprzeciw siebie. Na jednej kanapie usiadło trzech pasażerów: P1, P2, P3, a na drugiej tylko dwóch: P4 i P5. Na ile różnych sposobów mogą siedzieć pasażerowie, tak aby zawsze dwóch z nich siedziało naprzeciwko siebie?

Rozwiązanie

Pasażerowie P1, P2, P3 mogą usiąść na 4 numerowanych miejscach na tyle sposobów, ile jest różnych wariacji bez powtórzeń z 4 elementów po 3. Oczywiście, dwie osoby nie mogą siedzieć na jednym miejscu, a więc uporządkowanie odgrywa tu rolę. Mamy zatem:

$$V_4^3 = \frac{4!}{(4-3)!} = 24.$$

Dla każdego ustawienia na kanapie pierwszej istnieje tyle ustawień na kanapie drugiej takich, aby dwóch pasażerów siedziało zawsze naprzeciwko siebie, ile jest różnych wariacji bez powtórzeń z 3 elementów po 2. Do 3 pasażerów na kanapie pierwszej rozmieszczamy odpowiednio 2 na kanapie drugiej, czyli mamy:

$$V_3^2 = \frac{3!}{(3-2)!} = 6.$$

Liczba wszystkich możliwych rozmieszczeń spełniających warunki zadania wyraża się iloczynem (łącznie zajście)

$$V_4^3 \cdot V_3^2 = 24 \cdot 6 = 144.$$

Pasażerowie mogą siedzieć na 144 sposoby.

Przykład 1.1.8

Ilu było uczestników narady u prezesa pewnej spółki, jeżeli wiadomo, że można ich posadzić przy stolikach 2-osobowych na 132 sposoby?

Rozwiązanie

Tworzymy zbiory 2-elementowe spośród n elementów. Kolejność elementów odrywa rolę, a elementy nie mogą się powtarzać, a zatem jest to wariacja bez powtórzeń. Korzystając z odpowiedniego wzoru wyliczamy niewiadomą liczbę elementów zbioru. Do rozwiązania mamy równanie”

$$V_n^2 = 132,$$

$$\frac{n!}{(n-2)!} = 132,$$

$$\frac{(n-2)!(n-1) \cdot n}{(n-2)!} = 132,$$

$$(n-1) \cdot n = 132,$$

$$n^2 - n - 132 = 0,$$

$$n_1 = -11 \quad \text{lub} \quad n_2 = 12.$$

Było 12 uczestników narady.

Przykład 1.1.9

Ile liczb pięciocyfrowych można utworzyć z cyfr 1, 2, 3?

Rozwiązanie

Tworzymy zbiory 5-elementowe z trzech cyfr, a elementy mogą się powtarzać, ale kolejność odgrywa rolę. A więc, tyle jest różnych liczb pięciocyfrowych, ile różnych wariacji z powtórzeniami z trzech elementów:

$$\overline{V}_3^5 = 3^5 = 243.$$

Można utworzyć 243 liczby.

Przykład 1.1.10

Abonent zapomniał dwóch ostatnich cyfr numeru telefonu. Jaka jest maksymalna liczba prób do wykonania, aby trafić na właściwy numer?

Rozwiązanie

Tworzymy zbiory 2-elementowe spośród 10-elementowego zbioru cyfr. Kolejność występowania elementów odgrywa rolę, a zatem jest to wariacja z powtórzeniami:

$$\bar{V}_{10}^2 = 10^2 = 100 .$$

Maksymalna liczba prób wynosi 100.

Przykład 1.1.11

Na przyjęcie przybyło 6 osób, które miały do dyspozycji 6 różnokolorowych kieliszków. Jeden z kieliszków został stłuczony, a gospodarz podał dwa różne gatunki wina. Iloma sposobami goście mogą się napić?

Rozwiązanie

Wiadomo, że jedna osoba nie będzie piła, bo jest 5 różnokolorowych kieliszków dla 6 osób. Rozmieszczenie kieliszków możemy obliczyć według wzoru na wariację bez powtórzeń:

$$V_6^5 = \frac{6!}{(6-5)!} = \frac{6!}{1!} = 1 \cdot 2 \cdot 3 \cdot 4 \cdot 5 \cdot 6 = 720 .$$

Wiadomo też, że są dwa gatunki wina, a kieliszków 5, więc na pewno powtórzy się gatunek wina. Istotne jest, którego wina, do jakiego kieliszka gospodarz naleje, a zatem jest to wariacja z powtórzeniami

$$\bar{V}_2^5 = 2^5 = 32$$

Aby goście mogli przystąpić do picia wina, muszą zajść dwa zdarzenia jednocześnie: rozmieszczenie kieliszków i rozlewanie wina, a zatem odpowiedni iloczyn:

$$V_6^5 \cdot \bar{V}_2^5 = 720 \cdot 32 = 23040 .$$

Goście mogą napić się wina na 23 040 sposobów.

Przykład 1.1.12

Centrala telefoniczna pracuje na połączeniach 7-cyfrowych, które uzyskuje z cyfr 0, 1, 2, ..., 9. Ilu abonentów maksymalnie może zarejestrować centrala, jeżeli połączenia zaczynające się od zera nie są brane pod uwagę?

Rozwiązanie

Tworzymy zbiory 7-elementowe z cyfr, których jest dziesięć. Kolejność występowania cyfr odgrywa rolę, bo to jest numer telefonu, a cyfry mogą się

powtarzać. Do obliczeń wykorzystamy wzór na wariację z powtórzeniami następująco:

$$\bar{V}_{10}^7 - \frac{1}{10} \bar{V}_{10}^7 = \frac{9}{10} \bar{V}_{10}^7 = \frac{9}{10} \cdot 10^7 = 9000000.$$

Centrala może zarejestrować maksymalnie 9 000 000 abonentów.

Przykład 1.1.13

Ile nastąpi powitań, gdy jednocześnie spotka się sześciu znajomych?

Rozwiązanie

Ze zbioru 6-elementowego tworzymy zbiory 2-elementowe (witają się dwie osoby). Elementy nie powtarzają się, kolejność przy witaniu nie odgrywa roli, a zatem jest to kombinacja bez powtórzeń:

$$C_6^2 = \binom{6}{2} = \frac{6!}{2!(6-2)!} = \frac{4! \cdot 5 \cdot 6}{2 \cdot 4!} = 15.$$

Nastąpi 15 powitań.

Przykład 1.1.14

W pewnym turnieju szachowym wzięło udział 12 szachistów i każdy zagrał jedną partię z każdym. Ile partii szachowych rozegrano w tym turnieju?

Rozwiązanie

Ze zbioru 12-elementowego tworzymy zbiory 2-elementowe (szachiści grają w parach). Elementy nie powtarzają się, kolejność nie odgrywa roli, bo partnerzy dobierani są losowo. Jest to kombinacja bez powtórzeń:

$$C_{12}^2 = \binom{12}{2} = \frac{12!}{2!(12-2)!} = \frac{10! \cdot 11 \cdot 12}{2 \cdot 10!} = 66.$$

Rozegrano 66 partii.

Przykład 1.1.15

Rozpatrzmy 10 żetonów ponumerowanych liczbami 1, 2, 3, ..., 10 i wylosujemy jeden z nich. Na ile sposobów możemy wylosować liczbę pierwszą?

Rozwiązanie

W rezultacie losowania możemy otrzymać każdy z tych żetonów. Stawiając pytanie, czy wylosowana liczba jest liczbą pierwszą, wyróżniamy pewien zbiór cyfr, który ma tę cechę. Do zbioru tego należą cyfry: 1, 2, 3, 5 i 7. Zatem ze zbioru 10-elementowego losujemy zbiór 5-elementowy, przy czym kolejność nie odgrywa roli. A zatem jest to kombinacja bez powtórzeń:

$$C_{10}^5 = \binom{10}{5} = \frac{10!}{5!(10-5)!} = \frac{5! \cdot 6 \cdot 7 \cdot 8 \cdot 8 \cdot 10}{5! \cdot 5!} = \frac{30240}{120} = 252.$$

Liczbę pierwszą możemy wylosować na 252 sposoby.

Przykład 1.1.16

Pewna spółka przeprowadza konkurs w celu zatrudnienia czterech nowych pracowników. Do konkursu zgłosiło się 10 kandydatów. Iloma sposobami można obsadzić stanowiska pracy?

Rozwiązanie

Ze zbioru 10-elementowego wybieramy czwórki w sposób dowolny. Możemy zastosować wzór na kombinację bez powtórzeń:

$$C_{10}^4 = \binom{10}{4} = \frac{10!}{4!(10-4)!} = \frac{6! \cdot 7 \cdot 8 \cdot 9 \cdot 10}{24 \cdot 6!} = 210.$$

Stanowiska te można obsadzić na 210 sposobów.

Przykład 1.1.17

Mamy do dyspozycji cztery rodzaje owoców południowych: pomarańcze, grapefruity, kiwi i banany. Tworzymy paczki dla dzieci po pięć owoców. Ile różnych paczek możemy utworzyć?

Rozwiązanie

Tworzymy zbiór 5-elementowy ze zbioru 4-elementowego. Oznacza to, że przynajmniej jeden element musi się powtórzyć, a kolejność nie odgrywa roli. A zatem paczek utworzymy tyle, ile jest kombinacji z powtórzeniami, co możemy zapisać:

$$\bar{C}_4^5 = \binom{4+5-1}{5} = \binom{8}{5} = \frac{8!}{5!(8-5)!} = \frac{5! \cdot 6 \cdot 7 \cdot 8}{6 \cdot 5!} = 56.$$

Możemy utworzyć 56 paczek.

Przykład 1.1.18

Pewien klient, który zaciągnął pożyczkę musi wpłacić do kasy 10 rat miesięcznych w ciągu roku. Iloma sposobami może to zrobić?

Rozwiązanie

Klient może wpłacać po jednej racie, po dwie albo jeszcze inaczej. W roku jest 12 miesięcy, a zatem będzie to kombinacja z powtórzeniami:

$$\bar{C}_{12}^{10} = \binom{12+10-1}{10} = \binom{21}{10} = \frac{21!}{10!(21-10)!} = 352716.$$

Klient może wykorzystać 352 716 sposobów.

Przykład 1.1.19

Klub sportowy ma do dyspozycji 10 sprinterów. Obliczyć ile jest możliwych składów sztafet 4×100 m:

- a) z ustaloną kolejnością zmian,

b) bez ustalonej kolejności zmian.

Rozwiązanie

a) gdy kolejność zmian jest ustalona, to sztafeta ABCD jest różna od sztafety BCDA itd., są to więc 4-elementowe wariacje bez powtórzeń, a zatem:

$$V_{10}^4 = \frac{10!}{(10-4)!} = \frac{10!}{(10-4)!} = \frac{6! \cdot 7 \cdot 8 \cdot 9 \cdot 10}{6!} = 5040.$$

b) jeżeli kolejność zmian nie jest ustalona, wówczas sztafetę ABCD uważa się za taką samą co BCDA a więc jest to kombinacja bez powtórzeń:

$$C_{10}^4 = \binom{10}{4} = \frac{10!}{4!(10-4)!} = \frac{10!}{4! \cdot 6!} = \frac{6! \cdot 7 \cdot 8 \cdot 9 \cdot 10}{24 \cdot 6!} = 210.$$

1.2. Zdarzenia losowe

Znajomość praw przyrody nie zawsze wystarcza, aby przewidzieć wynik obserwacji, czy też prowadzonych eksperymentów. Podstawą naszej wiedzy o świecie jest doświadczenie. Powtarzając wielokrotnie to samo doświadczenie, w tych samych warunkach, możemy wyrobić sobie pogląd o badanych zjawiskach oraz o prawach, którym podlegają. Przyczyny wpływające na wynik końcowy doświadczeń są liczne i bardzo złożone. Trudno jest ocenić zmiany w procesie ich przebiegu, które to zmiany często powoduje przypadek, a tym bardziej wyeliminować błąd. Na przykład dokonując kilkakrotnie pomiaru pewniej wielkości w identycznych warunkach otrzymujemy na ogół różne wyniki na skutek popełnianych błędów. Rzucając tą samą monetą zauważamy, że w niektórych rzutach otrzymujemy orła, zaś w innych reszkę. O wyniku decyduje tylko przypadek.

Zdarzeniem losowym (przypadkowym) nazywamy takie zdarzenie, którego zajścia lub niezajścia nie można przewidzieć,

Niektóre zdarzenia są indywidualne, niepowtarzalne, inne zaś mają charakter masowy. Zjawiska masowe powtarzają się nieskończenie wiele razy. Gdy zjawisko ma charakter masowy, to pomimo nieprawidłowości, którym podlegają indywidualne wyniki, przeciętny wynik długiego ciągu doświadczeń losowych wykazuje pewną prawidłowość. Zaobserwowana prawidłowość polega na tym, że częstość zajścia zdarzenia losowego oscyluje wokół pewnej wartości stałej, przy dużej liczbie doświadczeń. Tylko masowe, przypadkowe zdarzenia są przedmiotem rachunku prawdopodobieństwa, który bada prawa rządzące prawidłowościami, jakim one podlegają.

Wybrane właściwości zdarzeń oraz pojęcia z nimi związane

Między zdarzeniami zachodzą następujące związki:

1. Jeżeli zdarzenie A zachodzi wtedy i tylko wtedy, gdy zachodzi zdarzenie B, to mówimy, że zdarzenie A jest równoważne (równe) zdarzeniu B:

$$A = B.$$

2. Jeżeli zdarzenie B zachodzi wtedy, gdy zachodzi zdarzenie A, to mówimy, że zdarzenie A jest sprzyjającym zdarzeniu B:

$$A \subseteq B.$$

3. Jeżeli zdarzenie A zachodzi wtedy i tylko wtedy, gdy zajdzie co najmniej jedno ze zdarzeń A_1, A_2, \dots, A_n , to powiadamy, że zdarzenie A jest sumą zdarzeń A_1, A_2, \dots, A_n . Zapisać to możemy na dwa sposoby:

$$A = A_1 + A_2 + \dots + A_n \quad \text{lub} \quad A = A_1 \cup A_2 \cup \dots \cup A_n.$$

Każde ze zdarzeń A_i ($i = 1, 2, \dots, n$) jest zdarzeniem sprzyjającym A, gdy zajście dowolnego zdarzenia A_i powoduje zajście zdarzenia A.

4. Jeżeli zdarzenia A nie możemy przedstawić w postaci sumy co najmniej dwóch różnych od A zdarzeń, to takie zdarzenie nazywamy **elementarnym** (w teorii prawdopodobieństwa elementarne zdarzenie losowe uznaje się jako pojęcie pierwotne, za pomocą którego określamy zdarzenie losowe).
5. Zbiór wszystkich możliwych (w określonych warunkach) zdarzeń elementarnych nazywamy **przestrzenią zdarzeń elementarnych E**:

$$E = \{e_1, e_2, \dots, e_n\}.$$

Zbiór E nazywa się niekiedy przestrzenią próby.

6. Zdarzenie nazywamy **pewnym**, jeżeli zbiorem zdarzeń sprzyjających jest przestrzeń zdarzeń elementarnych.
7. Zdarzenie nazywamy **niemożliwym**, jeżeli zbiorem zdarzeń sprzyjających jest zbiór pusty.
8. Zdarzenie A polegające na zajściu każdego ze zdarzeń A_1, A_2, \dots, A_n nazywamy iloczynem (koniunkcja, łączne zajście) zdarzeń A_1, A_2, \dots, A_n , co zapisujemy następująco:

$$A = A_1 \cdot A_2 \cdot \dots \cdot A_n \quad \text{lub} \quad A = A_1 \cap A_2 \cap \dots \cap A_n.$$

A czytamy:

$$A = A_1 \text{ i } A_2 \text{ i } \dots \text{ i } A_n$$

9. Zdarzenia A oraz B nazywamy **wykluczającymi się**, jeżeli łączne zajście tych zdarzeń jest zdarzeniem niemożliwym, co możemy zapisać:

$$A \cdot B = 0.$$

10. Powiadamy, że zdarzenia A_1, A_2, \dots, A_n tworzą **układ zupełny** zdarzeń, jeżeli wyłączają się parami, a ich suma jest zdarzeniem pewnym, tzn.

$$A_1 + A_2 + \dots + A_n = E \quad \text{oraz} \quad A_i \cdot A_j = 0,$$

gdzie: $i \neq j$

$$i = 1, 2, \dots, n; \quad j = 1, 2, \dots, n.$$

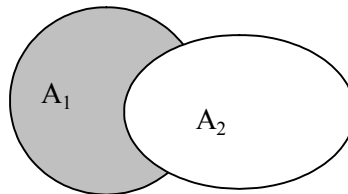
11. Zdarzenia A i \bar{A} nazywamy zdarzeniami **przeciwnymi**, jeżeli tworzą układ zupełny zdarzeń, tzn. łączne zajście tych zdarzeń jest zdarzeniem niemożliwym, co możemy zapisać:

$$A + \bar{A} = E \quad \text{oraz} \quad A \cdot \bar{A} = 0.$$

12. Zdarzenia A polegające na tym, że zajdzie zdarzenie A_1 i nie zajdzie zdarzenie A_2 nazywamy **różnicą zdarzeń** A_1 i A_2 i co zapisujemy:

$$A = A_1 - A_2.$$

Różnicę zdarzeń ilustruje poniższy rysunek.



Z matematycznego punktu widzenia zdarzenia interpretujemy jako zbiory, a zatem można na nich wykonywać działania tak samo, jak na zbiorach.

Przykład 1.2.1

Rzucamy dwiema kostkami. Niech zdarzenie A polega na otrzymaniu nieparzystej sumy oczek, a zdarzenie B polega na wyrzuceniu nieparzystej liczby oczek na jednej kostce i parzystej liczby oczek na drugiej kostce. Łatwo dostrzec, że $A = B$.

Przykład 1.2.2

Rzucamy jeden raz kostką do gry. Niech zdarzenie A polega na otrzymaniu parzystej liczby oczek, zaś zdarzenie A_1 polega na otrzymaniu „4”. Wówczas $A_1 \subseteq A$.

Przykład 1.2.3

Niech A_1 oznacza zdarzenie polegające na wyrzuceniu nieparzystej liczby oczek, zaś A_2 oznacza „orła” przy rzucie monetą. Zdarzenie $A = A_1 + A_2$ polega na otrzymaniu liczby nieparzystej lub „orła”, to znaczy na zajściu co najmniej jednego z tych zdarzeń.

Przykład 1.2.4

W doświadczeniu polegającym na jednokrotnym rzucie monetą określimy przestrzeń zdarzeń elementarnych. Zdarzenie wyrzucenia „orła” jest zdarzeniem elementarnym, zdarzenie wyrzucenia „reszki” jest także zdarzeniem elementarnym. Przestrzeń zdarzeń elementarnych możemy zapisać następująco:

$$E = \{e_1, e_2\} = \{\text{„orzeł”}, \text{„reszka”}\}.$$

Przykład 1.2.5

W doświadczeniu polegającym na jednokrotnym rzucie kostką do gry mamy do czynienia ze skończoną przestrzenią zdarzeń elementarnych.

e_1 – pojawienie się jednego oczka,

e_2 – pojawienie się dwóch oczek,

⋮

⋮

⋮

e_6 – pojawienie się sześciu oczek.

Tak więc $E = \{e_1, e_2, e_3, e_4, e_5, e_6\}$.

Przykład 1.2.6

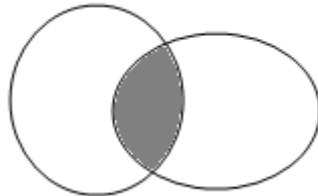
Rzucamy jeden raz kostką do gry. Zdarzenie polegające na wyrzuceniu jakiegokolwiek liczby oczek od 1 do 6 uznajemy za pewne, gdyż jedno ze zdarzeń elementarnych zawsze zajdzie.

Przykład 1.2.7

Rzucamy jeden raz kostką do gry. Zdarzenie polegające na wyrzuceniu „7” jest zdarzeniem niemożliwym, bo na kostce nie ma takiej liczby.

Przykład 1.2.8

Niech zdarzenia A_1 i A_2 oznaczają tarcze o częściowo wspólnym polu. Jeżeli strzelec ma trafić jedną kulą w obie tarcze (p. rys.) to musi strzelić w te części obu tracz, które się pokrywają.



Przykład 1.2.9

Rzucamy jeden raz kostką do gry. Niech zdarzenie A polega na wyrzuceniu parzystej liczby oczek, zaś zdarzenie B polega na wyrzuceniu „5”. Zbiory A i B nie mają wspólnego elementu, a zatem zdarzenia A i B wykluczają się wzajemnie.

Przykład 1.2.10

Rzucamy jeden raz kostką do gry. Niech zdarzenie A polega na wyrzuceniu parzystej liczby oczek, zaś zdarzenie B polega na wyrzuceniu nieparzystej liczby oczek, czyli $A = \{2, 4, 6\}$; $B = \{1, 3, 5\}$. Wszystkie zdarzenia elementarne zdarzeń A i B dają całą przestrzeń zdarzeń elementarnych $E = \{1, 2, 3, 4, 5, 6\}$, przy czym wykluczają się wzajemnie. Wynika stąd, że zdarzenia A i B tworzą układ zupełny zdarzeń.

Przykład 1.2.11

Rzucamy jeden raz monetą. Zdarzenie polegające na wyrzuceniu „orła” oraz zdarzenie polegające na wyrzuceniu „reszki” są zdarzeniami przeciwnymi, ponieważ tworzą układ zupełny zdarzeń.

Przykład 1.2.12

Niektóre skrzynki owoców południowych dostarczane do magazynu są zepsute. Wybrano losowo dwie skrzynki do kontroli. Ile możliwych zdarzeń elementarnych występuje w tym doświadczeniu. Wymienić wszystkie zdarzenia elementarne i określić przestrzeń zdarzeń elementarnych tego doświadczenia losowego.

Rozwiązanie

Doświadczeniem losowym jest wybór skrzynki z owocami i ocena: zepsuta, niezepsuta. Stąd możliwe są tylko dwa zdarzenia elementarne zarówno dla pierwszej, jak i dla drugiej skrzynki. Oznaczając odpowiednio przez S_1 i S_2 przestrzeń zdarzeń elementarnych możemy przedstawić następująco:

Zdarzenie elementarne	S_1	S_2
e_1	zepsute	zepsute
e_2	zepsute	niezepsute
e_3	niezepsute	zepsute
e_4	niezepsute	niezepsute

Przestrzenią zdarzeń elementarnych są wszystkie możliwe zdarzenia elementarne, co możemy oznaczyć:

$$E = \{e_1, e_2, e_3, e_4\}.$$

1.3. Rachunek prawdopodobieństwa

Rachunek prawdopodobieństwa zajmuje się wykrywaniem i badaniem prawidłowości zachodzących pomiędzy zdarzeniami. Podstawowymi pojęciami rachunku prawdopodobieństwa są: zdarzenie elementarne, przestrzeń zdarzeń oraz prawdopodobieństwo.

Prawdopodobieństwo jest charakterystyką liczbową stopnia możliwości realizacji zdarzenia w określonych warunkach. Prawdopodobieństwo realizacji zdarzenia lub krócej prawdopodobieństwo zdarzenia A oznacza się najczęściej jako $P(A)$. Jeżeli przestrzeń zdarzeń elementarnych E zawiera n jednakowo możliwych i wzajemnie wykluczających się zdarzeń elementarnych, spośród których k zdarzeń sprzyja zajściu pewnego zdarzenia A (n, k – skończone), to prawdopodobieństwem zajścia zdarzenia A nazywamy liczbę:

$$P(A) = \frac{k}{n}.$$

Jest to tak zwana **klasyczna definicja prawdopodobieństwa**, którą sformułował Laplace na podstawie obserwacji częstości empirycznych dla długich serii doświadczeń. Ponieważ pojęcie prawdopodobieństwa łączy się z wynikiem obserwacji lub eksperymentu, prawdopodobieństwo można zdefiniować jako funkcję, której wartościami są liczby rzeczywiste, a argumentami zdarzenia.

Definicja prawdopodobieństwa Laplace'a jest praktycznie tożsama z pojęciem częstości empirycznej z dodatkowym warunkiem skończonej liczby jednostek zbiorowości statystycznej przy założeniu jednakowo możliwych zdarzeń. Wynika stąd jej ograniczone zastosowanie tylko do dużych liczebnie zbiorowości statystycznych.

Praktyczne zastosowanie ma **aksjomatyczna definicja prawdopodobieństwa** zdarzenia losowego sformułowana przez Kołmogorowa, która opiera się na trzech niewymagających dowodu założeniach (pewnikach):

1. Każdemu zdarzeniu losowemu A odpowiada liczba $P(A) = p$ zwana prawdopodobieństwem realizacji tego zdarzenia, przy czym liczba ta zawiera się w granicach od $p = 0$ do $p = 1$.
2. Jeżeli z góry wiadomo, że zdarzenie B jest pewne (nie jest losowe), to prawdopodobieństwo $P(B) = 1$.
3. Jeżeli zdarzenia A_1, A_2, \dots, A_n tworzą skończony ciąg zdarzeń losowych parami wykluczających się to:

$$P(A_1 + A_2 + \dots + A_n) = P(A_1) + P(A_2) + \dots + P(A_n) = 1.$$

Dla nieskończonego, ale przeliczalnego ciągu zdarzeń mamy:

$$P(A_1 + A_2 + A_3 + \dots) = P(A_1) + P(A_2) + P(A_3) + \dots \leq 1.$$

Podstawowe własności prawdopodobieństw:

1. Prawdopodobieństwo $P(A)$ zdarzenia pewnego równa się 1:

$$P(A) = 1.$$

2. Prawdopodobieństwo $P(A)$ zdarzenia niemożliwego równa się 0:

$$P(A) = 0.$$

3. Prawdopodobieństwo każdego zdarzenia spełnia nierówność:

$$0 \leq P(A) \leq 1.$$

4. Jeżeli $A \subset B$ tzn. zajście zdarzenia A pociąga za sobą zajście zdarzenia B , to

$$P(A) < P(B).$$

5. Jeżeli zdarzenia A jest równe zdarzeniu B , to

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

6. Prawdopodobieństwo zajścia zdarzenia przeciwnego zdarzeniu A jest równe jedności pomniejszonej o prawdopodobieństwo zdarzenia A :

$$P(\bar{A}) = 1 - P(A).$$

7. Prawdopodobieństwo zdarzenia A polegającego na zajściu jednego z dwu wykluczających się zdarzeń A_1, A_2 równa się sumie prawdopodobieństw tych zdarzeń:

$$P(A) = P(A_1) + P(A_2).$$

8. Prawdopodobieństwo zdarzenia A polegającego na zajściu przynajmniej jednego ze zdarzeń A_1, A_2 równa się sumie prawdopodobieństw tych zdarzeń zmniejszonej o prawdopodobieństwo łącznego ich zajścia:

$$P(A) = P(A_1) + P(A_2) - P(A_1 \cdot A_2).$$

9. Na to, aby zdarzenia A i B były niezależne, potrzeba i wystarcza, żeby prawdopodobieństwo łącznego zajścia tych zdarzeń było równe iloczynowi prawdopodobieństw każdego z tych zdarzeń:

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

W przypadku, gdy zbiór zdarzeń elementarnych jest nieskończony zachodzi potrzeba rozszerzenia definicji prawdopodobieństwa. Pojęcie liczności zbioru zdarzeń elementarnych jest tu nieprzydatne. Możemy wtedy określić **prawdopodobieństwo geometryczne** zdarzenia.

Niech zbiór zdarzeń elementarnych Ω będzie zbiorem punktów prostej, płaszczyzny lub przestrzeni. Przyjmijmy następujące założenia:

- zbiór Ω jest mierzalny o skończonej mierze, tzn. ma skończoną długość, pole lub objętość,
- każdy punkt zbioru Ω ma jednakową szansę wylosowania.

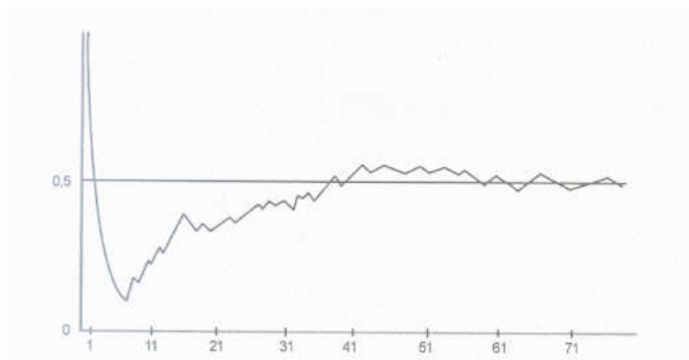
Przy wymienionych założeniach prawdopodobieństwo geometryczne dowolnego zdarzenia A określamy jako iloraz miary zdarzenia A do miary zdarzenia Ω , co wyraża wzór:

$$P(A) = \frac{\text{miara } A}{\text{miara } \Omega},$$

gdzie miara oznacza długość, pole lub objętość w zależności od tego, czy zbiór Ω leży na prostej, płaszczyźnie lub w przestrzeni.

W praktyce nie zawsze znamy liczebność zbioru zdarzeń elementarnych, bądź nie możemy doliczyć się liczby zdarzeń elementarnych sprzyjających zdarzeniom losowym, co potrzebne jest przy wykorzystaniu definicji klasycznej. Znajomości tych wielkości nie wymaga **statystyczna definicja prawdopodobieństwa**. W długiej serii doświadczeń obserwujemy wystąpienie zdarzenia A . Jeżeli częstość $\frac{n}{N}$ zdarzenia A , gdzie N jest długością serii, a n liczbą doświadczeń, w których pojawiło się zdarzenie A , przy wzrastaniu długości serii zbliża się do pewnej liczby p oscylując wokół tej liczby i jeśli wahania częstości zdarzenia przejawiają tendencję malejącą przy rosnącym N , to liczba p nazywana jest prawdopodobieństwem zdarzenia A :

$$P(A) = \lim_{n \rightarrow \infty} \frac{n}{N}.$$



Rys. 1.3.1. Ilustracja statystycznej definicji prawdopodobieństwa [5]

Podane definicje prawdopodobieństwa: klasyczna, geometryczna i statystyczna są szczególnymi przypadkami definicji aksjomatycznej.

Przykład 1.3.1

Przy okrągłym stole ponumerowano miejsca 1, 2, 3, 4. Przybyły dwie osoby, które usiadły przy tym stole. Jakie jest prawdopodobieństwo, że osoby te usiadły na miejscach o rosnących numerach?

Rozwiązanie

Zauważmy, że są trzy sprzyjające przypadki: 1, 2; 2, 3; i 3, 4. Wszystkich możliwych przypadków jest tyle, ile istnieje różnych zbiorów 2-elementowych utworzonych ze zbioru 4-elementowego, w których kolejność odgrywa rolę. Obliczmy więc wariację bez powtórzeń:

$$V_4^2 = \frac{4!}{(4-2)!} = \frac{2! \cdot 3 \cdot 4}{2!} = 12.$$

A zatem $p = \frac{3}{12}$.

Szukane prawdopodobieństwo wynosi 0,25.

Przykład 1.3.2

Babcia dziergała wełniany sweterek dla wnusia. W koszu miała 8 kłębków wełny czerwonej, 5 białej i 4 beżowej. Obliczyć prawdopodobieństwo, że wybierając trzykrotnie z kosza bez zwracania, babcia wyjęła dokładnie 2 kłębki wełny tego samego koloru.

Rozwiązanie

$$p = \frac{\binom{8}{2} \cdot \binom{9}{1} + \binom{5}{2} \cdot \binom{12}{1} + \binom{4}{2} \cdot \binom{13}{1}}{\binom{17}{3}} = \frac{252 + 120 + 78}{680} = \frac{450}{680} = \frac{45}{68} \approx 0,66.$$

Babcia wyjęła z kosza 2 kłębki tego samego koloru z prawdopodobieństwem 0,66.

Przykład 1.3.3

Do sklepu przywieziono owoce: 3 razy tyle bananów, co jabłek, a 8 razy tyle bananów, co pomarańczy. Jakie jest prawdopodobieństwo, że biorąc owoc losowo wybierzemy jabłko?

Rozwiązanie

Wprowadzamy odpowiednie oznaczenia:

y – ilość jabłek,

$3y$ – ilość bananów,

$\frac{3y}{8}$ – ilość pomarańczy.

Obliczamy ilość wszystkich owoców:

$$y + 3y + \frac{3y}{8} = \frac{8y + 24y + 3y}{8} = \frac{35y}{8}.$$

A zatem:

$$p = \frac{y}{\frac{35y}{8}} = \frac{8}{35} = 0,23.$$

Jabłko wylosujemy z prawdopodobieństwem 0,23.

Przykład 1.3.4

Realizując dostawy towaru do zagranicznego kontrahenta z możliwością opóźnień w transporcie oszacowano prawdopodobieństwa terminowości tych dostaw i otrzymano wyniki:

- prawdopodobieństwo terminowej dostawy towaru A – $P(A) = 0,5935$,
- prawdopodobieństwo terminowej dostawy towaru B – $P(B) = 0,3874$,
- prawdopodobieństwo terminowej dostawy towaru C – $P(C) = 0,0191$.

Sprawdź, czy spełnione są aksjomaty Kołmogorowa.

Rozwiązanie

Pierwszy aksjomat jest spełniony, ponieważ wszystkie oszacowane prawdopodobieństwa zawierają się w przedziale od 0 do 1.

Suma prawdopodobieństw wynosi:

$$P(A) + P(B) + P(C) = 0,5935 + 0,3874 + 0,0191 = 1.$$

Jest zdarzeniem pewnym, a zatem spełniony jest drugi aksjomat.

Zdarzenia parami wykluczają się, sprawdzamy trzeci aksjomat:

$$P(A) + P(B) = 0,5935 + 0,3874 = 0,9809 < 1.$$

Przykład 1.3.5

Rzucamy jeden raz kostką do gry. Jakie jest prawdopodobieństwo, że liczba otrzymanych oczek będzie mniejsza od 3?

Rozwiązanie

A – zdarzenie polegające na wyrzuceniu mniej niż 3 oczek,

A_1 – zdarzenie polegające na wyrzuceniu 1 oczka,

A_2 – zdarzenie polegające na wyrzuceniu 2 oczek.

Zdarzenia A_1 i A_2 wykluczają się, korzystamy z właściwości 7, a zatem:

$$P(A) = P(A_1) + P(A_2) = \frac{1}{6} + \frac{1}{6} = \frac{2}{6} = 0,33.$$

Prawdopodobieństwo, że liczba otrzymanych oczek będzie mniejsza od 3 wynosi 0,33.

Przykład 1.3.6

Dwaj myśliwi jednocześnie ujrzeni zająca i jednocześnie strzelili do niego. Każdy z myśliwych na 100 oddanych strzałów trafił 60 razy. Obliczyć prawdopodobieństwo zastrzelenia zająca?

Rozwiązanie

A – zdarzenie polegające na trafieniu zająca,

A_1 – zdarzenie polegające na trafieniu zająca przez I myśliwego,

A_2 – zdarzenie polegające na trafieniu zająca przez II myśliwego,

$A_1 \cdot A_2$ – jednoczesne trafienia zająca przez obu myśliwych.

$$P(A) = P(A_1) + P(A_2) - P(A_1 \cdot A_2) = 0,6 + 0,6 - 0,6 \cdot 0,6 = 1,2 - 0,36 = 0,84.$$

Prawdopodobieństwo zastrzelenia zająca wynosi 0,84.

Przykład 1.3.7

Na loterii jest 10 losów, z których 4 wygrywają. Jakie jest prawdopodobieństwo, że wśród czterech kupionych losów:

a) dokładnie jeden wygrywa,

b) przynajmniej jeden wygrywa.

Rozwiązanie

$$\text{a) } p_1 = \frac{\binom{4}{1} \cdot \binom{6}{3}}{\binom{10}{4}} = \frac{4 \cdot 20}{210} = \frac{80}{210} = 0,38,$$

$$\text{b) } p_2 = 1 - \frac{\binom{4}{0} \cdot \binom{6}{4}}{\binom{10}{4}} = 1 - \frac{15}{210} = 1 - 0,07 = 0,93.$$

Przykład 1.3.8

Trzech fotografów wykonuje zdjęcia w trudnych warunkach atmosferycznych. Zakładamy, że prawdopodobieństwo wykonania dobrego zdjęcia jest dla wszystkich jednakowe i wynosi $P(A) = 0,7$. Jakie jest prawdopodobieństwo p , że przynajmniej jedno zdjęcie będzie udane?

Rozwiązanie

Obliczamy prawdopodobieństwo wykonania nieudanego zdjęcia:

$$P(\bar{A}) = [1 - P(A)]^3 = (0,3)^3 = 0,027.$$

Prawdopodobieństwo p , że przynajmniej jedno zdjęcie będzie udane obliczamy korzystając z własności 6:

$$p = 1 - P(\bar{A}) = 1 - 0,027 = 0,973.$$

Prawdopodobieństwo p zajścia zdarzenia, że przynajmniej jedno zdjęcie będzie nieudane wynosi 0,973.

Prawdopodobieństwem warunkowym $P(A/B)$ zdarzenia A , przy założeniu, że zaszło zdarzenie B , nazywamy iloraz prawdopodobieństwa łącznego zajścia zdarzeń A i B przez prawdopodobieństwa zdarzenia B :

$$P(A/B) = \frac{P(A \cdot B)}{P(B)}; P(B) > 0.$$

Przekształcając powyższy wzór do postaci:

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

otrzymujemy wzór na prawdopodobieństwo łącznego zajścia zdarzeń A i B , które jest równe iloczynowi prawdopodobieństwa zdarzenia B przez prawdopodobieństwo warunkowe zdarzenia A przy założeniu, że zaszło zdarzenie B .

Jeżeli zdarzenia A_1, A_2, \dots, A_n są zdarzeniami wzajemnie wykluczającymi się, z których prawdopodobieństwo żadnego nie wynosi zero i jedno z nich musi zajść (zdarzenia tworzą, tzw. układ zupełny) to prawdopodobieństwo dowolnego zdarzenia B obliczamy ze wzoru na **prawdopodobieństwo całkowite**:

$$\begin{aligned} P(B) &= P(A_1) \cdot P(B/A_1) + P(A_2) \cdot P(B/A_2) + \dots + P(A_n) \cdot P(B/A_n) = \\ &= \sum_{i=1}^n P(A_i) \cdot P(B/A_i). \end{aligned}$$

Wynik wielu doświadczeń (przedsięwzięć) zależy od tego, co dzieje się w różnych pośrednich stadiach jego realizacji, czyli jak poszczególne przyczyny wpływają na wynik końcowy. Prawdopodobieństwo, z jakim przyczyna wpływa na skutek (wynik końcowy) można obliczyć, korzystając z **zasady Bayes'a**.

Jeżeli zdarzenia $A_1, A_2, \dots, A_i, \dots, A_n$ (przyczyny) są zdarzeniami wzajemnie wykluczającymi się, z których prawdopodobieństwo żadnego nie wynosi zero i jedno z nich musi zajść oraz zaszło zdarzenie B (skutek), to warunkowe prawdopodobieństwo dowolnego zdarzenia A_i obliczamy ze wzoru:

$$P(A_i/B) = \frac{P(A_i) \cdot P(B/A_i)}{\sum_{j=1}^k P(A_j) \cdot P(B/A_j)}.$$

Przykład 1.3.9

W grupie 100 osób było 40 mężczyzn. Przebadano tę grupę pod względem wady wzroku i stwierdzono wadę u 2% mężczyzn i 5% kobiet. Wybieramy losowo jedną osobę. Jakie jest prawdopodobieństwo, że osoba ta ma wadę wzroku?

Rozwiązanie

Wprowadzimy oznaczenia:

W – zdarzenie polegające na wylosowaniu osoby z wadą wzroku,

K – zdarzenie, że wylosowana osoba jest kobietą, $P(K) = 0,60$,

M – zdarzenie, że wylosowana osoba jest mężczyzną, $P(M) = 0,40$.

Aby wyznaczyć wszystkie możliwe zdarzenia skorzystamy ze wzoru na prawdopodobieństwo całkowite:

$$P(W) = P(M) \cdot P(W/M) + P(K) \cdot P(W/K) = 0,40 \cdot 0,02 + 0,60 \cdot 0,05 = 0,08 + 0,03 = 0,38.$$

Prawdopodobieństwo, że losowo wybrana osoba ma wadę wzroku wynosi 0,38.

Przykład 1.3.10

W dziesięciu jednakowych koszach są róże. W połowie koszy jest po 15 róż białych i 5 czerwonych, w trzech po 10 róż białych i 10 czerwonych, zaś w pozostałych są 4 róże białe i 16 czerwonych. Sięgamy losowo do jednego kosza i wybieramy jedną różę. Obliczyć prawdopodobieństwo p wylosowania róży czerwonej.

Rozwiązanie

$$P = \frac{5}{10} \cdot \frac{5}{20} + \frac{3}{10} \cdot \frac{10}{20} + \frac{2}{10} \cdot \frac{16}{20} = 0,435.$$

Prawdopodobieństwo wylosowania róży czerwonej wynosi 0,435.

Przykład 1.3.11

W sklepie „Dla pań” na dwóch półkach są wyłożone torebki. Na pierwszej półce jest 6 torebek ze skóry i 4 z tkaniny, a na drugiej jest 8 ze skóry i 16 z tkaniny. Sięgamy losowo do jednej z półek i wybieramy do obejrzenia jedną torebkę. Jakie jest prawdopodobieństwo, że będzie to torebka ze skóry?

Rozwiązanie

$$P = \frac{1}{2} \cdot \frac{6}{10} + \frac{1}{2} \cdot \frac{8}{24} = \frac{6}{20} + \frac{8}{48} = 0,47.$$

Prawdopodobieństwo wybrania torebki ze skóry jest równe 0,47.

Przykład 1.3.12

Z talii 52 kart wylosowano losowo jedną kartę i bez oglądania przełożono do drugiej talii złożonej z 32 kart. Jakie jest prawdopodobieństwo wylosowania damy z otrzymanego zbioru kart?

Rozwiązanie

Wprowadzimy oznaczenia:

D_1 – zdarzenie wylosowania damy z pierwszej talii,

D_2 – zdarzenie wylosowania damy z drugiej talii.

$$\begin{aligned} P(D) &= P(D_1) \cdot P(D_2/D_1) + P(\bar{D}_1) \cdot P(D_2/\bar{D}_1) = \\ &= \frac{4}{52} \cdot \frac{5}{33} + \frac{48}{52} \cdot \frac{4}{33} = \frac{20 + 192}{1716} = \frac{212}{1716} = 0,124. \end{aligned}$$

Prawdopodobieństwo wylosowania damy wynosi 0,124.

Przykład 1.3.13

Okulista przebadał grupę 100 osób, w której było 40 mężczyzn celem wykrycia wady wzroku. W trakcie badania wykryto wady wzroku u 2% mężczyzn i 5% kobiet. Wybieramy losowo jedną osobę, u której wykryto wadę wzroku. Jakie jest prawdopodobieństwo, że jest to mężczyzna?

Rozwiązanie

Wprowadzimy oznaczenia:

W – zdarzenie polegające na wylosowaniu osoby z wadą wzroku,

K – zdarzenie, że wylosowana osoba jest kobietą, $P(K) = 0,60$,

M – zdarzenie, że wylosowana osoba jest mężczyzną, $P(M) = 0,40$.

Aby wyznaczyć wszystkie możliwe zdarzenia skorzystamy ze wzoru na prawdopodobieństwo całkowite

$$\begin{aligned} P(W) &= P(M) \cdot P(W/M) + P(K) \cdot P(W/K) = 0,40 \cdot 0,02 + 0,60 \cdot 0,05 = \\ &= 0,008 + 0,03 = 0,038. \end{aligned}$$

A zatem, w oparciu o zasadę Bayes'a:

$$P(M/W) = \frac{P(M) \cdot P(W/M)}{P(M) \cdot P(W/M) + P(K) \cdot P(W/K)} = \frac{0,008}{0,038} = 0,21.$$

Prawdopodobieństwo, że losowo wybrana osoba z wadą wzroku jest mężczyzną wynosi 0,21.

Przykład 1.3.14

Sklep z żarówkami ma trzech dostawców A, B, C, a wadliwość partii pochodzących od nich wynosi odpowiednio: 2%, 4%, 10%. Sklep zaopatrywany jest w taki sposób, że 25% żarówek pochodzi od dostawcy A, 35% od dostawcy B, reszta od dostawcy C. Klient zakupił w tym sklepie jedną żarów-

kę. Jakie jest prawdopodobieństwo, że żarówka pochodzi od dostawcy C, jeżeli przy pierwszej próbie nie zaświeciła się, a więc była wadliwa.

Rozwiązanie

Wprowadzimy oznaczenia:

W – zdarzenie wylosowania wadliwej żarówki,

A, B, C – zdarzenia wylosowania odpowiedniego dostawcy.

$$\begin{aligned} P(C/W) &= \frac{P(C) \cdot P(W/C)}{P(A) \cdot P(W/A) + P(B) \cdot P(W/B) + P(C) \cdot P(W/C)} = \\ &= \frac{0,40 \cdot 0,1}{0,25 \cdot 0,02 + 0,35 \cdot 0,04 + 0,40 \cdot 0,1} = \frac{0,04}{0,005 + 0,014 + 0,04} = \frac{0,04}{0,059} = 0,68. \end{aligned}$$

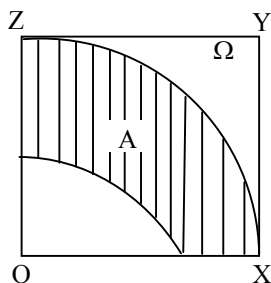
Prawdopodobieństwo, że wylosowana wadliwa żarówka pochodzi od dostawcy C wynosi 0,68.

Przykład 1.3.15

Obliczyć prawdopodobieństwo, że losowo wybrany punkt kwadratu OXYZ o boku 1 jest oddalony od punktu O więcej niż $\frac{1}{3}$ i mniej niż 1.

Rozwiązanie

$$P(A) = \frac{\text{pole } A}{\text{pole } \Omega} = \frac{\frac{1}{4} \pi(1^2 - (\frac{1}{3})^2)}{1^2} = \frac{1}{4} \pi(1^2 - \frac{1}{9}) = \frac{1}{4} \pi \cdot \frac{8}{9} = \frac{2\pi}{9}.$$



Rys. 1.4.1. Interpretacja geometryczna

1.4. Powtórzenie w przykładach

Zadanie 1

Na wyścigach konnych gracz typuje, który koń w gonitwie zajmie pierwsze miejsce i który zajmie drugie miejsce. Ile jest sposobów typowania, jeśli w gonitwie bierze udział 10 koni?

Rozwiązanie

Tworzymy ciągi 2-elementowe spośród 10 elementów. Elementy nie powtarzają się, a kolejność odgrywa rolę. Mamy do czynienia z wariacją bez powtórzeń, stąd:

$$V_{10}^2 = \frac{10!}{(10-2)!} = 90.$$

W tej gonitwie można typować na 90 sposobów.

Zadanie 2

W magazynie znajduje się miejsce na 100 słoików z dżemem owocowym. Słoiki rozmieszczono na trzech półkach: na pierwszej 40, na drugiej 35, na trzeciej 25. Iloza sposobami można to zrealizować?

Rozwiązanie

Sposób ustawienia słoików na każdej półce nie odgrywa roli. Istotne jest tylko to, aby liczba słoików na każdej z trzech półek była równa liczbie słoików możliwej do umieszczenia na każdej półce. Mamy tu do czynienia z kombinacją bez powtórzeń. Łączna liczba sposobów wynosi:

$$C_{100}^4 \cdot C_{60}^{35} \cdot C_{25}^{25} = \frac{100!}{40!(100-40)!} \cdot \frac{60!}{35!(60-35)!} \cdot \frac{25!}{25!(25-0)!} = \frac{100!}{40! \cdot 35! \cdot 25!}.$$

Słoiki można rozmieścić na $\frac{100!}{40! \cdot 35! \cdot 25!}$ sposobów.

Zadanie 3

U handlarza w kartonie znajduje się 7 kompletów lampek choinkowych, w tym 3 wadliwe. Wybieramy bez zwracania trzy komplety. Obliczyć prawdopodobieństwo, że spośród trzech wybranych kompletów lampek choinkowych podczas kontroli jakości dwa z nich nie zaświecą.

Rozwiązanie

Losowanie odbywa się bez zwracania i kolejność wylosowanych elementów nie odgrywa roli, zatem przyjmujemy, że zbiorem zdarzeń elementarnych

jest dowolna kombinacja 5-elementowa $\Omega = \binom{7}{5}$. Dwa komplety wadliwe

można wylosować spośród trzech wadliwych na $\binom{3}{2}$ sposoby, a 3 komplety

dobrych spośród 4 dobrych można wybrać na $\binom{4}{3}$ sposoby. Zatem

$$p = \frac{\binom{3}{2} \binom{4}{3}}{\binom{7}{5}} = \frac{12}{35}.$$

Prawdopodobieństwo wylosowania dwóch wadliwych kompletów lampek choinkowych wynosi $\frac{12}{35}$.

Zadanie 4

Partia towaru liczy 18 sztuk dobrych i 2 wadliwe. Partia zostaje przyjęta, jeśli wśród losowo wybranych 8 sztuk bez zwracania, co najwyżej jedna sztuka będzie wadliwa. Obliczyć prawdopodobieństwo przyjęcia partii.

Rozwiązanie

$$p = \frac{\binom{2}{0} \cdot \binom{18}{8} + \binom{2}{1} \cdot \binom{18}{7}}{\binom{20}{8}} = 0,85.$$

Prawdopodobieństwo przyjęcia partii wynosi 0,85.

Zadanie 5

W skrzynce znajduje się 20 butelek piwa, przy czym 4 butelki mają nieuszczelne kapsle. Klient kupił 3 losowo wybrane przez sprzedawcę butelki. Jakie jest prawdopodobieństwo, że klient kupił butelki szczelnie zamknięte?

Rozwiązanie

$$p = \frac{\binom{16}{3}}{\binom{20}{3}} = 0,49.$$

Prawdopodobieństwo zakupu szczelnie zamkniętej butelki wynosi 0,49.

Zadanie 6

Zawodnik trafia w środek tarczy przy jednym strzale z prawdopodobieństwem 0,5. Ile strzałów powinien oddać, aby z prawdopodobieństwem większym od 0,75 co najmniej raz trafił do celu?

Rozwiązanie

Zdarzenia polegające na trafieniu w dziesiątkę są niezależne. Niech A oznacza zdarzenie, że przy n strzałach zawodnik trafi w środek tarczy co najmniej raz. Wówczas

$$\begin{aligned}P(A) &= 1 - (1 - 0,5)^n > 0,75, \\1 - 0,75 &> (1 - 0,5)^n, \\(1 - 0,5)^n &< 0,25, \\n &> 2.\end{aligned}$$

Zawodnik powinien oddać co najmniej 3 strzały, aby z prawdopodobieństwem 0,75 co najmniej raz trafił w środek tarczy.

Zadanie 7

W dostawie było 96% dzinsów bez wad, z czego 75% stanowiły dzinsy męskie. Z całej dostawy wylosowano jedną parę dzinsów dobrych. Obliczyć prawdopodobieństwo, że są to dzinsy męskie.

Rozwiązanie

Należy skorzystać ze wzoru na prawdopodobieństwo warunkowe, a zatem:

$$p = 0,96 \cdot 0,75 = 0,72.$$

Prawdopodobieństwo wylosowania dzinsów męskich, przy warunku, że nie są wadliwe wynosi 0,72.

Zadanie 8

Na dwóch pochylniach zatrudnieni są robotnicy i uczniowie. Liczba osób zatrudnionych na obu pochylniach jest jednakowa, przy czym udział uczniów na pochylni P1 wynosi 0,2, zaś na pochylni P2 - 0,1. Zdarzył się wypadek i jedna osoba została poszkodowana. Jakie jest prawdopodobieństwo, że był to uczeń?

Rozwiązanie

Niech A oznacza zdarzenie, że wylosowana osoba będzie uczniem. Prawdopodobieństwo wylosowania osoby pracującej na każdej z pochylni jest jednakowe i wynosi 0,5. Stosując wzór na prawdopodobieństwo całkowite otrzymujemy:

$$P(A) = 0,5 \cdot 0,2 + 0,5 \cdot 0,1 = 0,15.$$

Prawdopodobieństwo, że wypadkowi uległ uczeń wynosi 0,15.

Zadanie 9

Dwie grupy studentów pisały pracę kontrolną. W pierwszej grupie było 5 studentów dobrych, 10 średnich i 15 słabych, zaś w drugiej było 10 dobrych, 7 średnich i 3 słabych. Obliczyć prawdopodobieństwo, że losowo wybrany student napisze pracę kontrolną z wynikiem dobrym.

Rozwiązanie

Stosujemy wzór na prawdopodobieństwo całkowite, a stąd:

$$p = \frac{1}{2} \cdot \frac{5}{30} + \frac{1}{2} \cdot \frac{10}{20} = \frac{1}{3}.$$

Prawdopodobieństwo, że losowo wybrany student napisze pracę kontrolną dobrze wynosi $\frac{1}{3}$.

Zadanie 10

Trzech farmerów dostarcza jajka do sklepu spożywczego. Liczba jaj dostarczonych przez poszczególnych farmerów wynosi odpowiednio: I – 1500, II – 3000, III – 1500. Przeciętnie biorąc, procent jaj zepsutych w dostawie pierwszego farmera wynosi 1%, drugiego 5%, trzeciego 1%. Klientka kupiła jedno jajko i okazało się, że jest zepsute. Obliczyć prawdopodobieństwo, że zepsute jajko pochodziło od farmera drugiego.

Rozwiązanie

Z warunków zadania wynika, że pierwszy farmer dostarcza 0,25, drugi 0,5, zaś trzeci 0,25 wszystkich jaj. Obliczymy prawdopodobieństwo zakupu zepsutego jajka (prawdopodobieństwo całkowite) następująco:

$$P(Z) = 0,25 \cdot 0,01 + 0,5 \cdot 0,05 + 0,25 \cdot 0,01 = 0,03.$$

Prawdopodobieństwo, że zepsute jajko pochodzi od drugiego farmera (wzór Bayesa) wynosi:

$$P(Z/II) = \frac{0,5 \cdot 0,05}{0,03} = 0,83.$$

Prawdopodobieństwo, że zepsute jajko pochodziło od drugiego farmera wynosi 0,83.

ROZDZIAŁ II ZMIENNE LOSOWE

2.1. Podstawowe pojęcia i charakterystyki

W różnorodnej działalności człowieka często spotykamy się z eksperymentem czy zdarzeniem przypadkowym. Wyniki badań wyrażane na ogół w postaci danych liczbowych kształtują się w sposób przypadkowy, mamy zatem do czynienia ze zmienną losową. Zmienną losową lub zmienną stochastyczną nazywamy funkcję, której wartość zmienia się w sposób losowy, od doświadczenia do doświadczenia, od przypadku do przypadku. W odróżnieniu od zmiennych zdeterminowanych, dla zmiennych losowych nie możemy dokładnie przewidzieć, jakie wartości przyjmą w danych warunkach. Możemy natomiast podać rozkład prawdopodobieństwa tych zmiennych.

Rozkład zmiennej losowej jest określony, jeżeli znany jest zbiór wszystkich możliwych wartości tej zmiennej oraz znana jest metoda obliczania prawdopodobieństwa przyjmowania przez tę zmienną wartości z dowolnego podzbioru możliwych wartości.

Zmienną losową X możemy zdefiniować jako funkcję o wartościach rzeczywistych, określoną na przestrzeni E zdarzeń elementarnych.

$$X : E \rightarrow \mathbb{R}$$

Oznacza to, że każdemu zdarzeniu z przestrzeni zdarzeń elementarnych można przyporządkować pewną liczbę \mathbb{R} .

$$X(e) \in \mathbb{R}$$

Wartości przyjmowane przez zmienne losowe nazywamy jej realizacjami i oznaczamy najczęściej, w odróżnieniu od zmiennych losowych, małymi literami.

$$X = \{x_1, x_2, \dots, x_i, \dots, x_n\}$$

Przykład 2.1.1

Przestrzenią zdarzeń elementarnych dla jednokrotnego rzutu sześcienną kostką do gry jest zbiór ścianek tej kostki:

$$E = \{e_1, e_2, \dots, e_6\}.$$

Funkcja $X(e)$ przyporządkowująca zdarzeniom elementarnym kolejne liczby od 1 do 6 jest zmienną losową, przyjmującą wartości:

$$X = \{1, 2, 3, 4, 5, 6\}.$$

Przykład 2.1.2

Przy rzucie monetą może wypaść „orzeł” lub „reszka”. W przypadku dwukrotnego rzutu monetą otrzymujemy następujący zbiór zdarzeń:

$$E = \{e_1 \Rightarrow OO, e_2 \Rightarrow OR, e_3 \Rightarrow RO, e_4 \Rightarrow RR\}.$$

Jeżeli wypadnięciu „orła” przyporządkujemy liczbę 1, a wypadnięciu „reszki” liczbę 0, to każdemu ze zdarzeń można przyporządkować liczbę rzeczywistą będącą sumą liczb 1 oraz 0, a zmienna losowa przyjmie wartości:

$$X = \{(1 + 1), (1 + 0), (0 + 1), (0 + 0)\} = \{2, 1, 1, 0\}.$$

Przykład 2.1.3

Jeżeli przedmiotem badań jest przebieg procesu eksploatacji maszyny, to przestrzeń zdarzeń elementarnych mogą tworzyć stany: pracy oraz postójów tej maszyny:

$$E = \{e_{1i} \Rightarrow \text{praca}, e_{2i} \Rightarrow \text{postój}\}.$$

Wówczas dwie zmienne losowe X i Y odwzorowują przebieg tego procesu. Realizacjami tych zmiennych są odcinki czasów pracy i postoju maszyny. Zmienna losowa X przyporządkowuje zdarzeniom elementarnym e_{1i} „praca pomiędzy postojami” wartości czasów pracy, a zmienna losowa Y przyporządkowuje zdarzeniom elementarnym e_{2i} „postój” wartości czasów postoju maszyny:

$$X = \{x_{11}, x_{12}, \dots, x_{1i}, \dots\},$$

$$Y = \{y_{11}, y_{12}, \dots, y_{1i}, \dots\}.$$

Jeżeli zbiór wartości, jakie przyjmuje zmienna losowa jest zbiorem przeliczalnym, to zmienna ta nazywa się **zmienną losową dyskretną** lub **zmienną losową skokową**. Jeżeli natomiast zbiór wartości, jakie przyjmuje zmienna losowa zawiera się w pewnym przedziale liczbowym, to wówczas zmienna ta nazywa się **zmienną losową ciągłą**.

Najczęściej stosowaną metodą określania prawdopodobieństw przyjmowania przez zmienną losową różnych wartości jest podanie dystrybuanty tej zmiennej lub rozkładu jej prawdopodobieństwa. Każda zmienna losowa jest jednoznacznie charakteryzowana za pomocą dystrybuanty.

Dystrybuanta zmiennej losowej X jest to funkcja zdefiniowana następująco:

$$F(x) = P(X < x).$$

Dystrybuanta określa prawdopodobieństwo tego, że zmienna losowa X przyjmie wartości mniejsze od danego x , czyli jest to skumulowana wartość prawdopodobieństwa do punktu, w którym zmienna losowa X przyjmuje

wartość x . Stąd każda zmienna losowa jest jednoznacznie charakteryzowana poprzez jej dystrybuantę.

Dystrybuanta posiada następujące własności:

1. $0 \leq F(x) \leq 1$,
2. $F(x)$ jest funkcją niemalejącą,
3. $F(x)$ jest funkcją co najmniej lewostronnie ciągłą,
4. $\lim_{x \rightarrow -\infty} F(x) = 0$ oraz $\lim_{x \rightarrow +\infty} F(x) = 1$.

Zmienna losowa skokowa jest charakteryzowana za pomocą rozkładu prawdopodobieństwa, a zmienna losowa ciągła za pomocą funkcji gęstości prawdopodobieństwa.

2.2. Zmienne losowe skokowe (dyskretne)

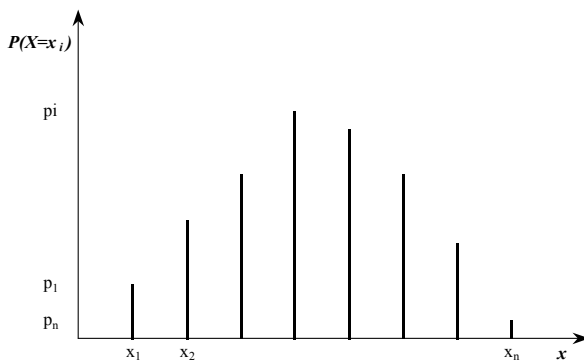
Rozkład prawdopodobieństwa zmiennej losowej skokowej jest to zbiór par (x_i, p_i) , gdzie x_i są to wartości, jakie przyjmuje zmienna losowa X , zaś p_i są to prawdopodobieństwa, z jakimi wartości te są przyjmowane.

Funkcja rozkładu prawdopodobieństwa zmiennej losowej skokowej X przedstawia się następująco:

$$P(X = x_i) = p_i,$$

$$\sum_{i=1}^n p_i = 1 \quad i = 1, 2, \dots, n.$$

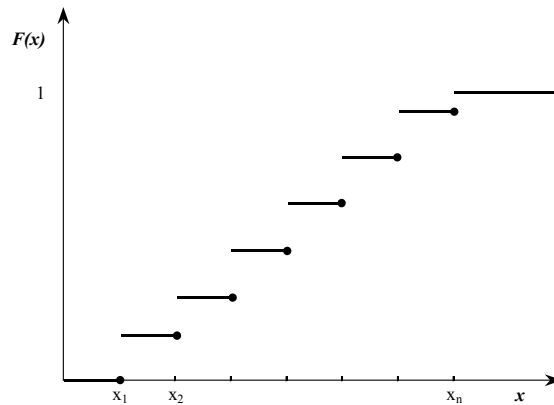
Funkcja rozkładu prawdopodobieństwa przyporządkowuje wartościom zmiennej losowej prawdopodobieństwo wystąpienia tejże wartości.



Rys. 2.2.1. Wykres funkcji rozkładu prawdopodobieństwa skokowej zmiennej losowej

Dystrybuanta zmiennej losowej skokowej X jest to funkcja określona wzorem:

$$F(x) = \sum_{x_i < x} p_i$$



Rys. 2.2.2. Wykres dystrybuanty skokowej zmiennej losowej

Dla zmiennej skokowej X prawdziwe są wzory:

$$P(X < x) = F(x),$$

$$P(a \leq X < b) = F(b) - F(a),$$

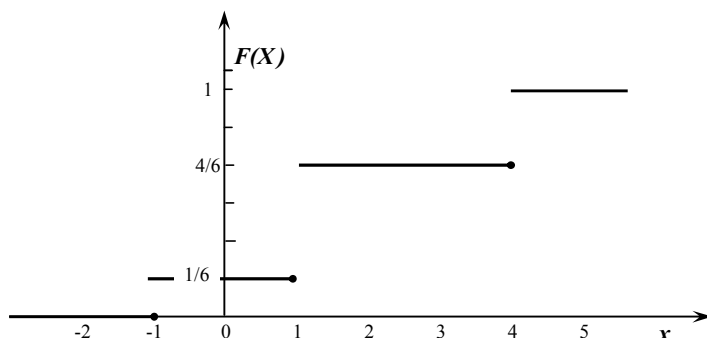
$$P(a \leq X \leq b) = F(b) - F(a) + P(X = b).$$

Przykład 2.2.1

Zmienna losowa X przyjmuje wartości: $x_1 = -1$, $x_2 = 1$, $x_3 = 4$ odpowiednio z prawdopodobieństwami $p_1 = \frac{1}{6}$, $p_2 = \frac{3}{6}$, $p_3 = \frac{2}{6}$. Wyznaczyć dystrybuantę zmiennej losowej X .

Rozwiązanie

$$F(x) = \sum_{x_i < x} p_i = \begin{cases} 0 & \text{dla } x \leq -1 \\ \frac{1}{6} & \text{dla } -1 < x \leq 1 \\ \frac{1}{6} + \frac{3}{6} = \frac{4}{6} & \text{dla } 1 < x \leq 4 \\ \frac{1}{6} + \frac{3}{6} + \frac{2}{6} = 1 & \text{dla } x > 4 \end{cases}$$



Rys. 2.2.3. Dystrybuanta zmiennej losowej X

Przykład 2.2.2

Uрна zawiera losy o numerach zakończonych cyframi 1, 2, 3, 4, 5 w stosunkach odpowiednio jak: 1 : 4 : 8 : 5 : 2. Niech X będzie zmienną losową przyjmującą wartości równe ostatniej cyfrze otrzymanego losu. Podać rozkład tak określonej zmiennej losowej oraz obliczyć prawdopodobieństwo, że zmienna losowa X przyjmie wartości pomiędzy 2 i 3, tj. $P(2 < X < 3)$.

Rozwiązanie

Z podanego w zadaniu stosunku wynika, że suma wszystkich losów w urnie jest wielokrotnością liczby 20. Stąd do obliczenia prawdopodobieństw wystąpienia poszczególnych wartości zmiennej losowej X można przyjąć, że minimalna liczba losów w urnie wynosi 20. Wyznaczony na podstawie klasycznej definicji prawdopodobieństwa, rozkład prawdopodobieństwa zmiennej losowej X podano w tabeli.

$X = x_i$	1	2	3	4	5
$P(X = x_i)$	1/20	1/5	2/5	1/4	1/10

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

Przykład 2.2.3

Uszycie pewnej ilości spódnic przez jedną krawcową w ciągu godziny jest zmienną losową X o następującym rozkładzie prawdopodobieństwa:

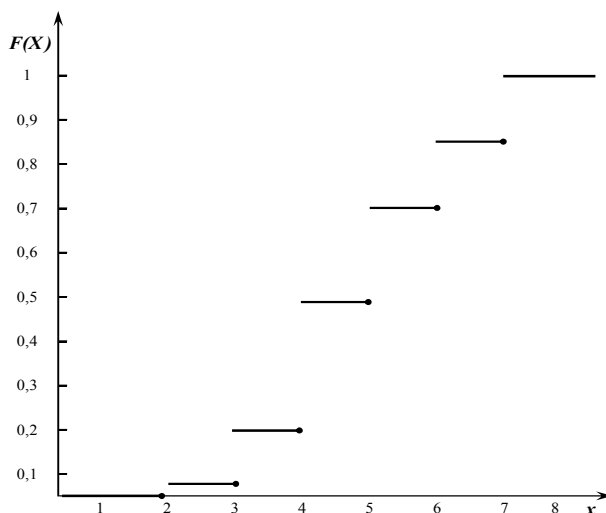
$X = x_i$	2	3	4	5	6	7
$p(x_i)$	0,02	0,18	0,28	0,22	0,16	0,14

Wyznaczyć dystrybuantę tej zmiennej oraz narysować jej wykres.

Rozwiązanie

Dystrybuanta zmiennej losowej X ma postać:

X	...2	...3	...4	...5	...6	...7
$F(X)$	0	0,02	0,20	0,48	0,70	0,86	1



Rys. 2.2.4. Wykres dystrybuanty skokowej zmiennej losowej X

Przykład 2.2.4

Sprawdzić, czy zmienna losowa X przyjmująca wartości $x_1 = -2$, $x_2 = 2$ z prawdopodobieństwami odpowiednio $p_1 = \frac{1}{6}$ i $p_2 = \frac{1}{2}$ jest zmienną losową skokową?

Rozwiązanie

Ponieważ $p_1 + p_2 = \frac{1}{6} + \frac{1}{2} \neq 1$, więc zmienna losowa X nie jest skokowa.

Przykład 2.2.5

Podać rozkład zmiennej losowej, która przyjmuje wartości równe sumie oczek na dwóch kostkach. Obliczyć $P(5 \leq X < 8)$ oraz wykreślić dystrybuantę tej zmiennej.

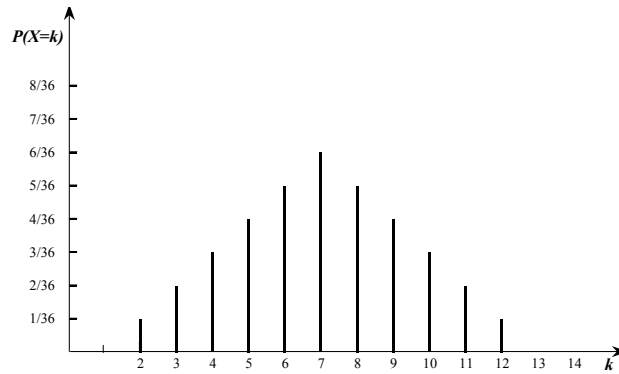
Rozwiązanie

Zmienna losowa może przyjmować wartości będące liczbami naturalnymi z przedziału $[2, 12]$. Obliczmy kolejno prawdopodobieństwa wyrzucenia dwoma kostkami 2, ..., 12 oczek, tj. prawdopodobieństwo iloczynu zdarzeń elementarnych wyrzucenia i ($i = 1, \dots, 6$) oczek na kostce A oraz na kostce B. Podamy sposób obliczania dla wybranych prawdopodobieństw:

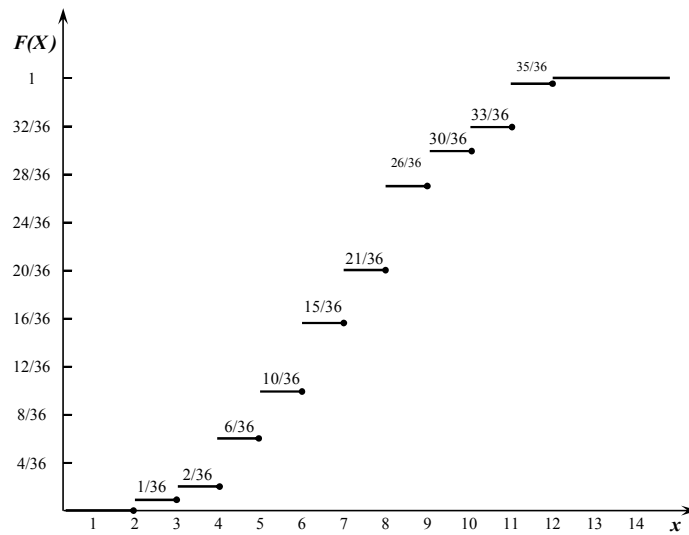
$$P(X = 2) = P(A_1 \cdot B_1) = P(A_1) \cdot P(B_1) = \frac{1}{6} \cdot \frac{1}{6} = \frac{1}{36},$$

$$\begin{aligned}
 P(X = 3) &= P(A_1 \cdot B_2 + A_2 \cdot B_1) = \\
 &= P(A_1) \cdot P(B_2) + P(A_2) \cdot P(B_1) = \frac{2}{36} \\
 P(X = 7) &= P(A_6 \cdot B_1 + A_5 \cdot B_2 + A_4 \cdot B_3 + \\
 &+ A_3 \cdot B_4 + A_2 \cdot B_5 + A_1 \cdot B_6) = \frac{6}{36}
 \end{aligned}$$

k	2	3	4	5	6	7	8	9	10	11	12
$P(X = k)$	$\frac{1}{36}$	$\frac{2}{36}$	$\frac{3}{36}$	$\frac{4}{36}$	$\frac{5}{36}$	$\frac{6}{36}$	$\frac{5}{36}$	$\frac{4}{36}$	$\frac{3}{36}$	$\frac{2}{36}$	$\frac{1}{36}$



Rys. 2.2.5. Wykres rozkładu prawdopodobieństwa skokowej zmiennej losowej X



Rys. 2.2.6. Wykres dystrybuanty skokowej zmiennej losowej X

Dane do wykreślenia dystrybuanty $F(x)$:

k	...2	...3	...4	...5	...6	...7	...8	...9	...10	...11	...12
F(x)	0	$\frac{1}{36}$	$\frac{3}{36}$	$\frac{6}{36}$	$\frac{10}{36}$	$\frac{15}{36}$	$\frac{21}{36}$	$\frac{26}{36}$	$\frac{30}{36}$	$\frac{33}{36}$	$\frac{35}{36}$	1

Prawdopodobieństwo, że zmienna losowa przyjmie wartości z przedziału $(5 \leq X < 8)$ wynosi:

$$P(5 \leq X < 8) = \sum_{k=5}^7 P(X = k) = P(X = 5) + P(X = 6) + P(X = 7) = \frac{15}{36}.$$

Przykład 2.2.6

Zorganizowano następującą grę: rzucamy dwiema kostkami do gry; jeżeli suma oczek jest równa 2, otrzymujemy 10 zł, jeżeli suma oczek jest równa 3, otrzymujemy 5 zł. W pozostałych przypadkach płacimy 1 zł. Podać rozkład zmiennej losowej.

Rozwiązanie

Zmienna losowa X przyjmuje wartości 10, 5, -1 z następującymi prawdopodobieństwami:

$$P(X = 10) = 1/36, P(X = 5) = 2/36, \\ P(X = -1) = 1 - [P(X=10) + P(X=5)] = 33/36.$$

K	10	5	-1
P(X = k)	1/36	2/36	33/36

Przykład 2.2.7

Pewna gra polega na rzucie trzema monetami i otrzymaniu wygranej 10 zł w przypadku wyrzucenia trzech orłów i przegraniu 6 zł w pozostałych przypadkach. Traktując wygraną jako zmienną losową X , podać jej funkcję prawdopodobieństwa i dystrybuantę.

Rozwiązanie

Oznaczamy przez A zdarzenie wygrania 10 zł oraz przez B zdarzenie przegrania 6 zł. Prawdopodobieństwa tego, że zmienna losowa X przyjmie wartości plus 10 (wygrana) lub minus 6 (przegrana) można wyznaczyć w sposób podany poniżej.

Zakładając, że masa monet jest rozłożona równomiernie (monety nie są fałszywe), wyrzucenie „orła” oraz „reszki” realizuje się z takim samym prawdopodobieństwem równym 0,5. Aby wygrać 10 zł należy równocześnie wyrzucić trzy „orły”, stąd prawdopodobieństwo zajścia zdarzenia A jest iloczynem prawdopodobieństw, że na każdej monecie pojawi się „orzeł”, stąd:

$$P(A) = \frac{1}{2} \cdot \frac{1}{2} \cdot \frac{1}{2} = \frac{1}{8}.$$

W pozostałych przypadkach następuje przegrana. Stąd prawdopodobieństwo zajścia zdarzenia B wynosi:

$$P(B) = 1 - P(A) = \frac{7}{8}.$$

Prawdopodobieństwo zajścia zdarzenia A można również wyznaczyć korzystając z klasycznej definicji prawdopodobieństwa Laplace'a.

Przestrzeń zdarzeń elementarnych składa się 2^3 elementów, tj. 2^3 , czyli jest to przestrzeń ośmioelementowa. Zdarzenie sprzyjające wygraniu (układ wygrywający trzech „orłów”) jest tylko jedno.

Stąd prawdopodobieństwo zdarzenia A wynosi:

$$P(A) = \frac{1}{8}.$$

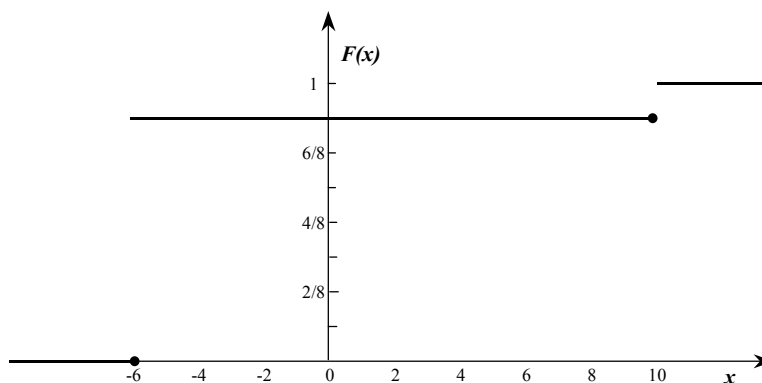
Prawdopodobieństwo zajścia zdarzenia B wynosi $1 - P(A)$.

Funkcję rozkładu prawdopodobieństwa zawiera tabela.

$X = k$	-6	10
$P(X = k)$	$7/8$	$1/8$

Dystrybuanta ma postać:

X - 6	10
$F(x)$	0	$7/8$	1

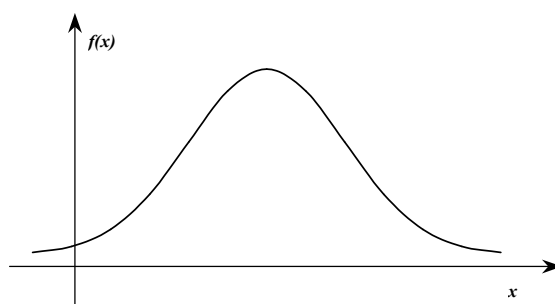


Rys. 2.2.7. Wykres dystrybuanty skokowej zmiennej losowej X

2.3. Zmienne losowe ciągłe

Funkcja gęstości prawdopodobieństwa zmiennej losowej ciągłej przedstawia się następująco:

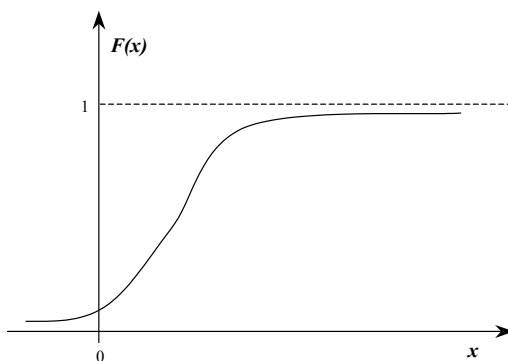
$$\int_{-\infty}^{+\infty} f(x)dx = 1$$



Rys. 2.3.1. Wykres funkcji gęstości prawdopodobieństwa ciągłej zmiennej losowej

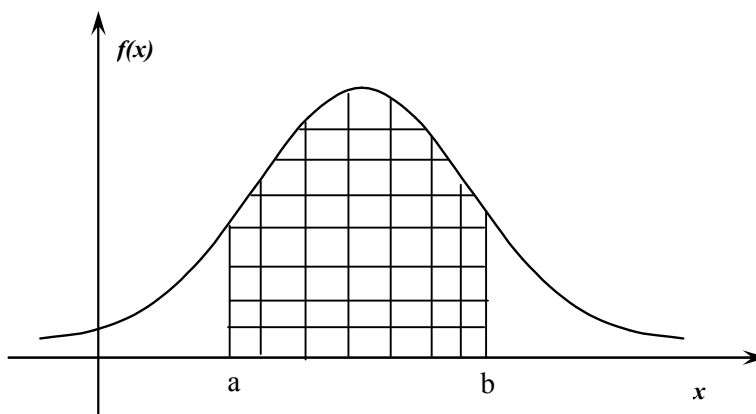
Dystrybuanta zmiennej losowej ciągłej dana jest wzorem:

$$P(X < x) = F(x) = \int_{-\infty}^x f(x)dx$$



Rys. 2.3.2. Wykres dystrybuanty ciągłej zmiennej losowej

Prawdopodobieństwo, że zmienna losowa X przyjmie wartości z dowolnego przedziału $\langle a, b \rangle$ jest równe polu powierzchni obszaru ograniczonego wykresem funkcji gęstości, osią OX oraz prostymi $x = a$ i $x = b$.



Rys. 2.3.3. Wykres funkcji gęstości prawdopodobieństwa ciągłej zmiennej losowej X o wartościach z dowolnego przedziału $\langle a, b \rangle$

$$P(a \leq x \leq b) = P(a < x < b) = P(a < x \leq b) = P(a \leq x < b) =$$

$$F(b) - F(a) = \int_a^b f(x) dx$$

Dla zmiennej losowej ciągłej X zachodzą zależności (a, b stałe):

- $P(X = a) = 0$,
- $P(a \leq X \leq b) = P(a < X < b) = F(b) - F(a)$; $a < b$,
- $P(X < a) = F(a)$,
- $P(X > a) = 1 - F(a)$.

Przykład 2.3.1

Dla jakiej wartości C funkcja:

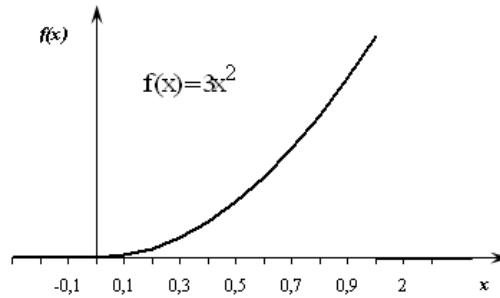
$$f(x) = \begin{cases} 0 & \text{dla } x < 0 \\ Cx^2 & \text{dla } 0 \leq x \leq 1 \\ 0 & \text{dla } x > 1 \end{cases}$$

jest funkcją gęstości prawdopodobieństwa pewnej zmiennej losowej X .

Rozwiązanie

$$\int_{-\infty}^{+\infty} f(x) dx = \int_{-\infty}^0 0 dx + \int_0^1 Cx^2 dx + \int_1^{\infty} 0 dx = C \frac{x^3}{3} \Big|_0^1 = \frac{1}{3}$$

$$\frac{1}{3}C = 1 \text{ stąd } C = 3.$$



Rys. 2.3.4. Wykres funkcji gęstości ciągłej zmiennej losowej X

Przykład 2.3.2

Zmienna losowa X podlega rozkładowi według gęstości danej wzorem:

$$f(x) = \begin{cases} 0 & \text{dla } x < 0 \\ C \sin x & \text{dla } 0 \leq x \leq \frac{1}{3}\pi \\ 0 & \text{dla } x > \frac{1}{3}\pi \end{cases}$$

Obliczyć stałą C . Podać dystrybuantę zmiennej losowej X . Obliczyć

$$P\left(\frac{1}{6}\pi \leq X \leq \frac{1}{4}\pi\right).$$

Rozwiązanie

Stałą obliczamy z zależności $\int_{-\infty}^{+\infty} f(x)dx = 1$.

$$\int_{-\infty}^{+\infty} f(x)dx = \int_{-\infty}^0 0dx + \int_0^{\frac{\pi}{3}} C \sin x + \int_{\frac{\pi}{3}}^{+\infty} 0dx = -C \cos x \Big|_0^{\frac{\pi}{3}} = -C \frac{1}{2} - (-C) = \frac{1}{2}C$$

$$\frac{1}{2}C = 1 \rightarrow C = 2.$$

Dystrybuantę wyznaczamy z zależności:

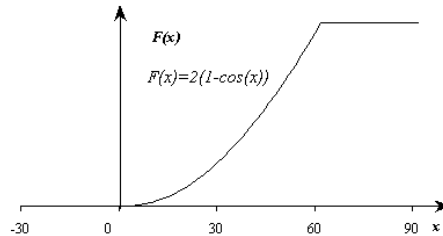
$$F(x) = \int_{-\infty}^x f(x)dx = \int_0^x 2 \sin x dx = -2 \cos x \Big|_0^x = -2 \cos x + 2,$$

stąd:

$$F(x) = \begin{cases} 0 & \text{dla } x < 0 \\ 2(1 - \cos x) & \text{dla } 0 \leq x \leq \frac{\pi}{3} \\ 1 & \text{dla } x > \frac{\pi}{3} \end{cases}$$

Szukane prawdopodobieństwo obliczamy ze wzoru:

$$P\left(\frac{\pi}{6} \leq X \leq \frac{\pi}{4}\right) = \int_{\frac{\pi}{6}}^{\frac{\pi}{4}} 2 \sin x dx = -2 \cos x \Big|_{\frac{\pi}{6}}^{\frac{\pi}{4}} = -2 \frac{\sqrt{2}}{2} + 2 \cdot \frac{\sqrt{3}}{2} = \sqrt{3} - \sqrt{2}.$$



Rys. 2.3.5. Wykres dystrybuanty ciągłej zmiennej losowej X

Przykład 2.3.3

Zmienna losowa X ma gęstość prawdopodobieństwa postaci:

$$f(x) = \begin{cases} 0 & \text{dla } x \leq 0 \\ 3x^2 & \text{dla } 0 < x \leq 1 \\ 0 & \text{dla } x > 1 \end{cases}$$

Wyznaczyć dystrybuantę zmiennej losowej X.

Rozwiązanie

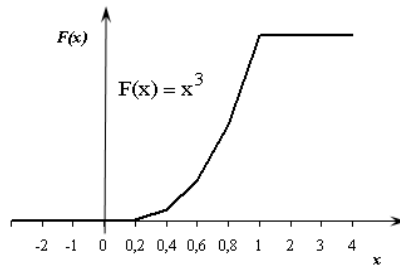
Korzystamy ze wzoru:

$$P(X < x) = F(x) = \int_{-\infty}^x f(x) dx$$

$$F(x) = \int_0^x 3x^2 dx = 3 \int_0^x x^2 dx = 3 \frac{x^3}{3} \Big|_0^x = x^3$$

Ostatecznie:

$$F(x) = \begin{cases} 0 & \text{dla } x \leq 0 \\ x^3 & \text{dla } 0 \leq x \leq 1 \\ 1 & \text{dla } x > 1 \end{cases}$$



Rys. 2.3.6. Wykres dystrybuanty ciągłej zmiennej losowej X

Przykład 2.3.4

Zmienna losowa ciągła X ma gęstość prawdopodobieństwa postaci:

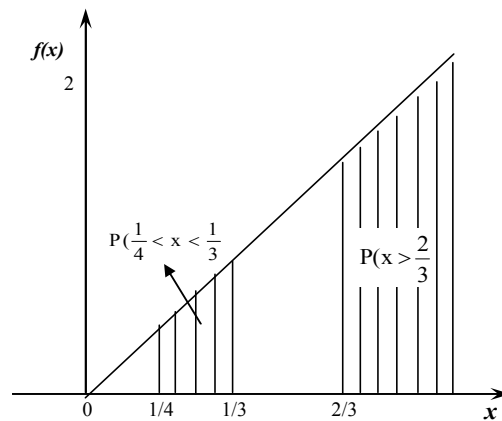
$$f(x) = \begin{cases} 2x & \text{dla } 0 < x < 1 \\ 0 & \text{dla pozostałych } x \end{cases}$$

Obliczyć następujące prawdopodobieństwa: a) $P(\frac{1}{4} < X < \frac{1}{3})$, b) $P(X > \frac{2}{3})$.

Rozwiązanie

$$\text{a) } P(\frac{1}{4} < X < \frac{1}{3}) = \int_{\frac{1}{4}}^{\frac{1}{3}} 2x dx = x^2 \Big|_{\frac{1}{4}}^{\frac{1}{3}} = \left(\frac{1}{3}\right)^2 - \left(\frac{1}{4}\right)^2 = \frac{16}{144} - \frac{9}{144} = \frac{7}{144}.$$

$$\text{b) } P(X > \frac{2}{3}) = \int_{\frac{2}{3}}^1 2x dx = x^2 \Big|_{\frac{2}{3}}^1 = 1^2 - \left(\frac{2}{3}\right)^2 = 1 - \frac{4}{9} = \frac{5}{9}.$$



Rys. 2.3.7. Graficzna prezentacja rozwiązania.

2.4. Parametry rozkładu zmiennej losowej

Zmienne losowe są charakteryzowane poprzez parametry klasyczne oraz pozycyjne.

Wartości stałe, które charakteryzują zmienną losową i są obliczane na podstawie wszystkich wartości tej zmiennej, nazywa się **parametrami klasycznymi** rozkładu zmiennej losowej X . Do grupy tych parametrów należą wartość oczekiwana, odchylenie standardowe oraz współczynnik zmienności.

Wartości stałe charakteryzujące zmienną losową, które są wyznaczone na podstawie jednej lub kilku wartości tej zmiennej, nazywa się **parametrami pozycyjnymi** rozkładu zmiennej losowej X . Do grupy tych parametrów należą mediana oraz modalna (dominanta).

2.4.1. Parametry klasyczne

Wartość oczekiwana $E(X)$ (przeciętna, średnia, nadzieja matematyczna) jest to wartość, wokół której skupiają się realizacje zmiennej losowej, uzyskiwane w wyniku wielokrotnie powtarzanego eksperymentu.

Wartość oczekiwaną oblicza się z zależności:

$$E(X) = \sum_{i=1}^n x_i \cdot p_i \quad \text{– dla zmiennej losowej skokowej,}$$

$$E(X) = \int_{-\infty}^{+\infty} x \cdot f(x) dx \quad \text{– dla zmiennej losowej ciągłej,}$$

gdzie: x_i – i -ta wartość zmiennej losowej X ,

p_i – prawdopodobieństwo wystąpienia wartości x_i ,

$f(x)$ – funkcja gęstości prawdopodobieństwa zmiennej losowej X .

Własności wartości oczekiwanej:

- istnieje zawsze i zawsze jest jedna,
- wartość oczekiwana stałej równa się tej stałej $E(C) = C$,
- wartość oczekiwana sumy dwóch zmiennych losowych X i Y równa się sumie wartości oczekiwanych tych zmiennych:

$$E(X + Y) = E(X) + E(Y),$$

- wartość iloczynu dwóch niezależnych zmiennych losowych jest równa iloczynowi wartości oczekiwanych tych zmiennych:

$$E(X \cap Y) = E(X) \cdot E(Y),$$

- stałą można wyciągnąć przed znak wartości oczekiwanej

$$E(C \cap Y) = E(C) \cdot E(Y) = C \cdot E(Y).$$

Przykład 2.4.1.1

Obliczyć wartość oczekiwaną eksperymentu polegającego na wielokrotnym rzucie kostką do gry.

Rozwiązanie

Zmienna losowa X może przyjmować każdą z wartości wynoszących 1 lub 2, lub 3, lub 4, lub 5, lub 6 z tym samym prawdopodobieństwem wynoszącym $1/6$, jeżeli kostka nie jest fałszywa, tzn. masa kostki jest rozłożona równomiernie na wszystkich ściankach.

$$E(X) = \sum_{i=1}^n x_i \cdot p_i = 1 \cdot \frac{1}{6} + 2 \cdot \frac{1}{6} + 3 \cdot \frac{1}{6} + 4 \cdot \frac{1}{6} + 5 \cdot \frac{1}{6} + 6 \cdot \frac{1}{6} = 3 \frac{1}{2}.$$

Przykład 2.4.1.2

Obliczyć wartość oczekiwaną zmiennej losowej X o następującej funkcji gęstości prawdopodobieństwa:

$$f(x) = \begin{cases} 0 & \text{dla } x \leq 0 \\ 2x & \text{dla } 0 < x \leq 1 \\ 0 & \text{dla } x > 1 \end{cases}$$

Rozwiązanie

$$E(X) = \int_{-\infty}^{+\infty} x \cdot f(x) dx = \int_{-\infty}^0 0 dx + \int_0^1 x \cdot 2x dx + \int_1^{+\infty} 0 dx = 0 + 2 \cdot \frac{x^3}{3} \Big|_0^1 + 0 = \frac{2}{3}.$$

Przykład 2.4.1.3

Zmienna losowa X przyjmuje następujące wartości: $x_1 = -1$, $x_2 = 0$, $x_3 = 1$. Wartość oczekiwana tej zmiennej wynosi $E(X) = 0,1$, natomiast wartość oczekiwana kwadratu zmiennej losowej X wynosi $E(X^2) = 0,9$. Obliczyć prawdopodobieństwa p_1 , p_2 , p_3 odpowiadające możliwym wartościom x_1 , x_2 , x_3 zmiennej losowej X .

Rozwiązanie

Z treści zadania wynika, że możemy utworzyć następujący układ trzech równań:

$$\begin{cases} p_1 + p_2 + p_3 = 1 \\ \sum_{i=1}^3 x_i \cdot p_i = 0,1 \\ \sum_{i=1}^3 (x_i)^2 \cdot p_i = 0,9 \end{cases}$$

Podstawiając dane otrzymujemy:

$$\begin{cases} p_1 + p_2 + p_3 = 1 \\ (-1) \cdot p_1 + 0 \cdot p_2 + 1 \cdot p_3 = 0,1 \\ (-1)^2 \cdot p_1 + 0^2 \cdot p_2 + 1^2 \cdot p_3 = 0,9 \end{cases}$$

$$\begin{cases} p_1 + p_2 + p_3 = 1 \\ -p_1 + p_3 = 0,1 \\ p_1 + p_3 = 0,9 \end{cases}$$
$$\begin{cases} p_1 + p_2 + 0,9 - p_1 = 1 \\ -p_1 + 0,9 - p_1 = 0,1 \\ p_3 = 0,9 - p_1 \end{cases}$$
$$\begin{cases} p_1 = 0,4 \\ p_2 = 0,1 \\ p_3 = 0,5 \end{cases}$$

Przykład 2.4.1.4

Dane są rozkłady niezależnych zmiennych losowych X i Y:

x_i	2	3	5
$P(X = x_i)$	0,5	0,3	0,2

y_i	4	6
$P(Y = y_i)$	0,75	0,25

Rozwiązanie

Obliczyć wartość oczekiwaną każdej ze zmiennych, $E(XY)$ oraz $E(X+Y)$.

$$E(X) = \sum_{i=1}^3 x_i \cdot p_i = 2,9,$$

$$E(Y) = \sum_{i=1}^2 x_i \cdot p_i = 4,5,$$

$$E(X \cdot Y) = E(X) \cdot E(Y) = 13,05,$$

$$E(X + Y) = E(X) + E(Y) = 7,4.$$

Wariancja zmiennej losowej X jest miarą rozproszenia kwadratów odchyleń poszczególnych wartości zmiennej wokół wartości średniej. Im wariancja jest mniejsza, tym bardziej wartości zmiennej losowej skupiają się wokół wartości przeciętnej $E(X)$.

Ogólnie wariancja zmiennej losowej X jest określona za pomocą jednego ze wzorów:

$$V(X) = E[X - E(X)]^2$$

lub

$$V(X) = E(X^2) - [E(X)]^2,$$

czyli wariancja równa się wartości oczekiwanej kwadratu zmiennej losowej pomniejszonej o kwadrat wartości oczekiwanej tej zmiennej.

Własności wariancji:

- wariancja stałej równa się 0: $V(C) = 0$,
- wariancja iloczynu stałej i zmiennej losowej równa się:

$$V(C \cdot X) = C^2 \cdot V(X),$$

- wariancja sumy (różnicy) dwóch niezależnych zmiennych losowych równa się:

$$V(X \pm Y) = V(X) + V(Y),$$

gdy X i Y są niezależne.

Wariancję $V(X)$ oblicza się z zależności:

- dla zmiennej losowej skokowej:

$$V(X) = \sum_{i=1}^n (x_i - E(X))^2 \cdot p_i$$

lub

$$V(X) = \frac{1}{n} \sum_{i=1}^n (x_i^2 \cdot p_i - E(X))^2,$$

- dla zmiennej losowej ciągłej:

$$V(X) = \int_{-\infty}^{+\infty} [x - E(X)]^2 \cdot f(x) dx$$

lub

$$V(X) = \int_{-\infty}^{+\infty} x^2 f(x) dx - [E(X)]^2.$$

Miarą zmienności wartości zmiennej losowej X jest odchylenie standardowe σ obliczane jako pierwiastek z wariancji $V(X)$:

$$\sigma = \sqrt{V(X)} \quad (\text{dla zmiennej losowej skokowej i ciągłej}).$$

Przykład 2.4.1.5

W autobusie zgasło światło w momencie, gdy jeden z pasażerów szukał biletu. Pasażer miał 10 biletów, w tym 5 biletów po 2,40 zł, 3 po 1,20 zł oraz 2 po 0,60 zł. Pasażer wyciągnął losowo jeden bilet i skasował go. Jaka jest wartość oczekiwana jego opłaty za przejazd, wariancja oraz odchylenie standardowe? Przyjmujemy, że prawdopodobieństwo wyciągnięcia każdego z biletów jest takie samo.

Rozwiązanie

Zdarzeniem losowym jest wyciągnięcie biletu. Rozpatrywana zmienna losowa X może przyjmować wartości odpowiadające nominałom biletów, tj. 2,40 lub 1,20 lub 0,60 zł.

Rozkład prawdopodobieństwa zmiennej losowej X przedstawia tabela:

x_i	2,4	1,2	0,6
p_i	0,5	0,3	0,2

Wartość oczekiwana $E(X)$ wynosi:

$$E(X) = \sum_{i=1}^3 x_i \cdot p_i = 2,4 \cdot 0,5 + 1,2 \cdot 0,3 + 0,6 \cdot 0,2 = 1,68 \text{ zł.}$$

Wariancja $V(X)$ wynosi:

$$\begin{aligned} V(X) &= \sum_{i=1}^3 [x_i - E(X)]^2 \cdot p_i = \\ &= [2,4 - 1,68]^2 \cdot 0,5 + [1,2 - 1,68]^2 \cdot 0,3 + [0,6 - 1,68]^2 \cdot 0,2 = 0,562. \end{aligned}$$

Odchylenie standardowe σ wynosi:

$$\sigma = \sqrt{V(X)} = \sqrt{0,562} = 0,75 \text{ zł.}$$

Przykład 2.4.1.6

Obliczyć wariancję zmiennej losowej X eksperymentu polegającego na wielokrotnym rzucie kostką do gry. $E(X) = 3,5$.

Rozwiązanie

$$V(X) = \sum_{i=1}^n (x_i - E(X))^2 \cdot p_i = \frac{1}{6} [(1 - 3,5)^2 + (2 - 3,5)^2 + (3 - 3,5)^2 + (4 - 3,5)^2 + (5 - 3,5)^2 + (6 - 3,5)^2] = 2,92.$$

Przykład 2.4.1.7

Dla zmiennej losowej X o funkcji gęstości prawdopodobieństwa podanej w przykładzie 2.4.1.2 obliczyć wariancję. $E(X) = \frac{2}{3}$.

Rozwiązanie

$$\begin{aligned} V(X) &= \int_{-\infty}^{+\infty} [x - E(X)]^2 \cdot f(x) dx = \int_0^1 \left(x - \frac{2}{3}\right)^2 \cdot 2x dx = \\ &= \int_0^1 \left(x^2 - \frac{4}{3}x + \frac{4}{9}\right) \cdot 2x dx = \int_0^1 2x^3 dx - \int_0^1 \frac{8}{3}x^2 dx + \int_0^1 \frac{8}{9}x dx = \\ &= 2 \cdot \frac{x^4}{4} \Big|_0^1 - \frac{8}{3} \cdot \frac{x^3}{3} \Big|_0^1 + \frac{8}{9} \cdot \frac{x^2}{2} \Big|_0^1 = \frac{1}{18}. \end{aligned}$$

Przykład 2.4.1.8

Dla jakiej wartości a funkcja $f(x)$ jest gęstością zmiennej losowej X ?

$$f(x) = \begin{cases} 0 & \text{dla } x \leq 1 \\ a \cdot x^2 & \text{dla } 1 < x \leq 2 \\ 0 & \text{dla } x > 2 \end{cases}$$

Znaleźć dystrybuantę oraz wartość oczekiwaną.

Rozwiązanie

Funkcja $f(x)$ jest gęstością, więc muszą być spełnione warunki:

1. $f(x) \geq 0$,
2. $\int_{-\infty}^{+\infty} f(x) dx = 1$.

Ad. 1. Warunek oczywisty.

$$\text{Ad. 2. } \int_{-\infty}^1 0 dx + \int_1^2 a \cdot x^2 dx + \int_2^{+\infty} 0 dx = a \left(\frac{x^3}{3} \right) \Big|_1^2 = a \left(\frac{2^3}{3} - \frac{1^3}{3} \right) = \frac{7}{3} a.$$

Z własności funkcji gęstości $f(x)$ wynika, że: $\frac{7}{3} a = 1 \Rightarrow a = \frac{3}{7}$, czyli

$$f(x) = \begin{cases} 0 & \text{dla } x \leq 1 \\ \frac{3}{7} \cdot x^2 & \text{dla } 1 < x \leq 2 \\ 0 & \text{dla } x > 2 \end{cases}$$

Dystrybuanta zmiennej losowej X wynosi:

$$F(x) = \begin{cases} 0 & \text{dla } x \leq 1 \\ \frac{1}{7}(x^3 - 1) & \text{dla } 1 < x \leq 2 \\ 1 & \text{dla } x > 2 \end{cases}$$

Ponieważ:

$$\begin{aligned} F(X) &= \int_{-\infty}^x f(x) dx = \int_{-\infty}^x \frac{3}{7} x^2 dx = \int_{-\infty}^1 0 dx + \int_1^x \frac{3}{7} x^2 dx + \int_2^x 0 dx = \\ &= \int_1^x \frac{3}{7} \cdot \frac{x^3}{3} \Big|_1^x = \frac{3}{7} \left(\frac{x^3}{3} - \frac{1^3}{3} \right) = \frac{1}{7} (x^3 - 1). \end{aligned}$$

Wartość oczekiwana $E(X)$ wynosi:

$$\begin{aligned} E(X) &= \int_{-\infty}^{+\infty} x f(x) dx = \int_1^2 x \cdot \frac{3}{7} x^2 dx = \frac{3}{7} \int_1^2 x^3 dx = \frac{3}{7} \cdot \frac{x^4}{4} \Big|_1^2 = \\ &= \frac{3}{7} \left(\frac{2^4}{4} - \frac{1^4}{4} \right) = \frac{45}{28}. \end{aligned}$$

Miarą zmienności względnej zmiennej losowej X jest współczynnik zmienności V_σ , który wyraża odchylenie standardowe jako procent wielkości średniej (określa wielkość zmienności w stosunku do wielkości będącej przedmiotem pomiarów):

$$V_{\sigma} = \frac{\sigma}{E(X)} \cdot 100\% \text{ (dla zmiennej losowej skokowej i ciągłej).}$$

Za pomocą współczynnika zmienności można porównywać dyspersję kilku zbiorów danych wyrażanych w odmiennych jednostkach miary (metrach, złotych, tonach).

Przykład 2.4.1.9

Obliczyć współczynnik zmienności dla zmiennej losowej X z przykładu 2.4.1.5.

Rozwiązanie

$$V_{\sigma} = \frac{\sigma}{E(X)} \cdot 100\% = \frac{0,75}{1,68} \cdot 100\% = 44,6\% .$$

Przykład 2.4.1.10

Wykonano pięć pomiarów długości pewnego przedmiotu. Obliczona średnia długość mierzonego przedmiotu wynosi $E(X) = 240$ cm, a odchylenie standardowe σ tych pomiarów wynosi 0,2 cm. Obliczyć współczynnik zmienności.

Rozwiązanie

$$V_{\sigma} = \frac{\sigma}{E(X)} \cdot 100\% = \frac{0,2}{240} \cdot 100\% = 0,083\% .$$

Przykład 2.4.1.11

Sprawdzić, czy funkcja określona poniższym wzorem jest funkcją gęstości zmiennej losowej X . Jeśli tak wyznaczyć dystrybuantę, wartość oczekiwaną, odchylenie standardowe oraz współczynnik zmienności.

$$f(x) = \begin{cases} 0 & \text{dla } x \leq 0 \\ 4x^3 & \text{dla } 0 < x \leq 1 \\ 0 & \text{dla } x > 1 \end{cases}$$

Rozwiązanie

Funkcja $f(x)$ jest funkcją nieujemną dla każdego x oraz $\int_{-\infty}^{+\infty} f(x)dx = 1$.

$$\text{Obliczamy wartość całki } \int_{-\infty}^{+\infty} 4x^3 dx = 4 \int_0^1 x^3 dx = 4 \cdot \frac{x^4}{4} \Big|_0^1 = 1 .$$

Ponieważ wartość ta jest równa jedności, funkcja $f(x) = 4x^3$ jest funkcją gęstości prawdopodobieństwa.

Dystrybuanta:

$$F(x) = \begin{cases} 0 & \text{dla } x \leq 0 \\ x^4 & \text{dla } 0 < x \leq 1 \\ 1 & \text{dla } x > 1 \end{cases}$$

ponieważ $F(x) = \int_0^x f(x)dx = 4 \int_0^x x^3 dx = 4 \cdot \frac{x^4}{4} \Big|_0^x = x^4$.

Wartość oczekiwana:

$$E(X) = \int_{-\infty}^{+\infty} xf(x)dx = \int_0^1 x \cdot 4x^3 dx = 4 \cdot \frac{x^5}{5} \Big|_0^1 = 0,8.$$

Wariancja:

$$V(X) = E(X^2) - [E(X)]^2 = \frac{4}{6} - 0,8^2 = \frac{2}{75}.$$

Ponieważ: $E(X^2) = \int_{-\infty}^{+\infty} x^2 \cdot f(x)dx = \int_0^1 x^2 \cdot 4x^3 dx = 4 \int_0^1 x^5 dx = 4 \cdot \frac{x^6}{6} \Big|_0^1 = \frac{4}{6}$ lub

$$\begin{aligned} V(X) &= \int_{-\infty}^{+\infty} [x - E(X)]^2 \cdot f(x)dx = \\ &= \int_0^1 x^2 \cdot 4x^3 dx - \int_0^1 1,6x \cdot 4x^3 dx + \int_0^1 0,64 \cdot 4x^3 dx = \\ &= \left[4 \cdot \frac{x^6}{6} - 6,4 \cdot \frac{x^5}{5} + 2,56 \cdot \frac{x^4}{4} \right]_0^1 = \frac{4}{6} - 6,4 \cdot \frac{1}{5} + 2,56 \cdot \frac{1}{4} = \frac{2}{75}. \end{aligned}$$

Odchylenie standardowe $\sigma = \sqrt{V(x)} = \sqrt{\frac{2}{75}} = 0,163$.

Współczynnik zmienności $V_\sigma = \frac{\sigma}{E(X)} \cdot 100\% = \frac{0,163}{0,8} \cdot 100\% = 20,4\%$.

2.4.2. Parametry pozycyjne

Mediana opisuje środek zbioru danych. W przypadku nieparzystej liczby elementów zbioru jest to wartość elementu środkowego zbioru danych uporządkowanych w kolejności malejącej lub rosnącej (lub średnia z wartości

dwóch środkowych elementów dla parzystej liczby elementów). Położenie mediany w zbiorze n elementowym wyznacza się ze wzoru:

$$\frac{n+1}{2}.$$

Mediana zawsze istnieje, zawsze jest jedna i nie ulega wpływom wartości skrajnych.

Medianą M_e zmiennej losowej skokowej X nazywamy wartość x spełniającą układ równań:

$$P(X \leq x) \geq \frac{1}{2} \text{ i } P(X \geq x) \geq \frac{1}{2}.$$

Mediana M_e zmiennej losowej ciągłej X jest to taka wartość x , dla której spełniona jest równość $F(x) = \frac{1}{2}$.

Modalna (dominanta, moda) jest wartością zmiennej losowej występującą z największą częstotliwością. Zbiór danych może mieć więcej niż jedną dominantę.

Modalną M_o skokowej zmiennej losowej X nazywamy wartość x , której odpowiada największe prawdopodobieństwo realizacji tej zmiennej.

Modalna zmiennej losowej ciągłej X jest to wartość x , dla której funkcja gęstości $f(x)$ osiąga wartość maksymalną.

Przykład 2.4.2.1

Wyznaczyć medianę dla zmiennej losowej skokowej X opisującej rzut kostką do gry.

Rozwiązanie

Rozkład prawdopodobieństwa zmiennej losowej X przedstawia się następująco:

x_i	1	2	3	4	5	6
$P(X = x_i)$	1/6	1/6	1/6	1/6	1/6	1/6

Dla zmiennej losowej skokowej X polegającej na rzucie kostką do gry mediana nie jest jednoznacznie określona. Medianą jest każda liczba zawarta w przedziale domkniętym $[3, 4]$, gdyż wtedy spełnione są nierówności:

$$P(X \leq 3) = \frac{3}{6} = \frac{1}{2} \text{ i } P(X \geq 3) = \frac{4}{6} > \frac{1}{2}$$

oraz

$$P(X \leq 4) = \frac{4}{6} > \frac{1}{2} \text{ i } P(X \geq 4) = \frac{3}{6} = \frac{1}{2}.$$

Przykład 2.4.2.2

Rozkład zmiennej losowej przedstawia się następująco:

x_i	1	2	5
$P(X = x_i)$	0,25	0,5	0,25

Znaleźć dystrybuantę, wartość oczekiwaną, wariancję, medianę i modalną.

Rozwiązanie

Dystrybuanta:

$$F(x) = \begin{cases} 0 & \text{dla } x \leq 1 \\ 0,25 & \text{dla } 1 < x \leq 2 \\ 0,75 & \text{dla } 2 < x \leq 5 \\ 1 & \text{dla } x > 5 \end{cases}$$

Wartość oczekiwana: $E(X) = \sum_{i=1}^3 x_i \cdot p_i = 2,5$.

Wariancja: $V(X) = \sum_{i=1}^3 [x_i - E(X)]^2 \cdot p_i = 2,25$.

Mediana: $M_e = 2$ ponieważ $P(X \leq 2) \geq \frac{1}{2}$ i $P(X \geq 2) \geq \frac{1}{2}$.

Modalna: $M_o = 2$ ponieważ: $P(X = 2) = 0,5$.

2.5. Powtórzenie w przykładach

Zadanie 1

Podać rozkład zmiennej losowej, która przyjmuje wartości równe sumie oczek na trzech kostkach do gry oraz wykreślić dystrybuantę tej zmiennej.

Rozwiązanie

Zmienna losowa może przyjmować wartości będące liczbami naturalnymi z przedziału $\langle 3, 18 \rangle$. Należy obliczyć prawdopodobieństwa przyjęcia przez zmienną losową określonej wartości, tj. wyrzuceniu określonej liczby oczek trzema kostkami.

Przy rzucie trzema kostkami ilość wszystkich możliwych układów liczb oczek wyrzuconych na tych kostkach jest równa:

$$n^k = 6^3 = 216,$$

gdzie: k – liczba kostek,

n – liczba oczek możliwych do wyrzucenia jedną kostką.

Oznaczając przez m_x liczbę możliwych układów, w których suma oczek równa jest x otrzymujemy wyniki umieszczone poniżej w tabeli.

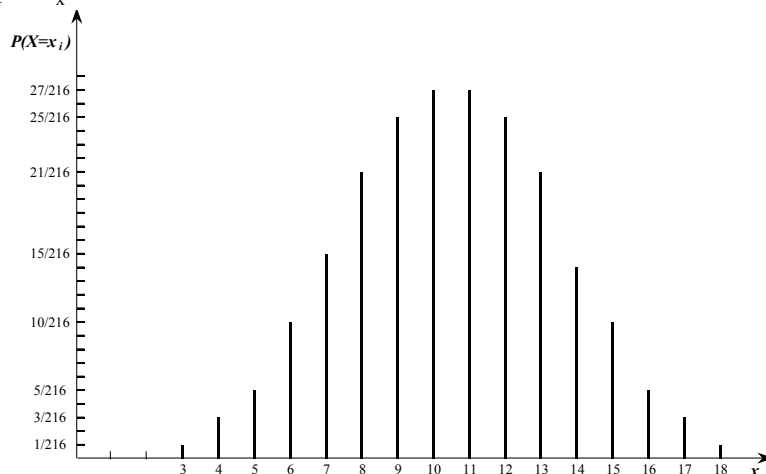
x	3	4	5	6	7	8	9	10
m_x	1	3	6	10	15	21	25	27
$P(X = x)$	1/216	3/216	6/216	10/216	15/216	21/216	25/216	27/216

x	11	12	13	14	15	16	17	18
m_x	27	25	21	15	10	6	3	1
$P(X = x)$	27/216	25/216	21/216	15/216	10/216	6/216	3/216	1/216

Uwaga! Rzucając równocześnie trzema kostkami do gry, sumę trzech oczek otrzymamy tylko wtedy, gdy na każdej kostce wyrzucimy „1”. Taki układ jest tylko jeden, stąd $m_x = 1$. Cztery oczka możemy uzyskać na trzy następujące sposoby:

Kostka nr 1	Kostka nr 2	Kostka nr 3	Suma oczek
1	1	2	4
1	2	1	4
2	1	1	4

stąd $m_x = 3$ itd.

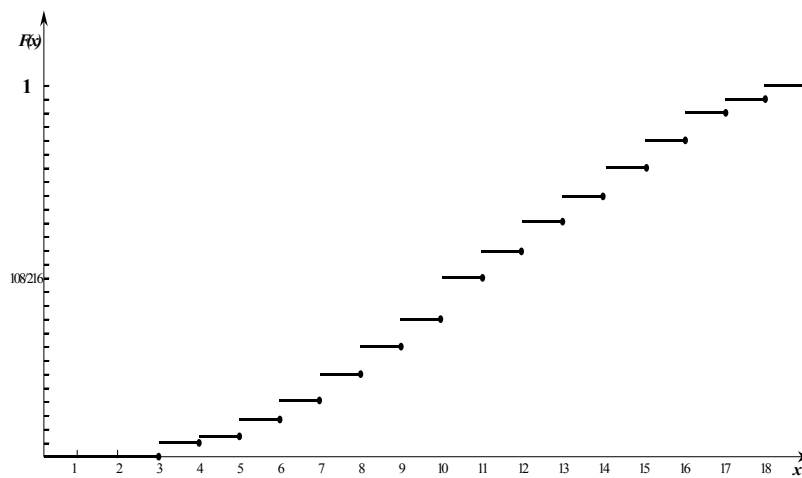


Rys. 2.5.1. Wykres funkcji rozkładu prawdopodobieństwa skokowej zmiennej losowej X

W celu wykreślenia dystrybuanty obliczamy wartości skumulowanego prawdopodobieństwa:

x	...3	...4	...5	...6	...7	...8	...9	...10	...11
F(x)	0	1/216	4/216	10/216	20/216	35/216	56/216	81/216	108/216

x	...12	...13	...14	...15	...16	...17	...18	...
F(x)	135/216	160/216	181/216	196/216	206/216	212/216	215/216	1



Rys. 2.5.2. Wykres dystrybuanty skokowej zmiennej losowej X

Zadanie 2

Zysk osiągnięty przez przedsiębiorstwo P jest zmienną losową X, której rozkład ma następującą funkcję gęstości:

$$f(x) = \begin{cases} 0 & \text{dla } x < 0 \\ Cx & \text{dla } 0 \leq x \leq 4 \\ 0 & \text{dla } x > 4 \end{cases}$$

1. Obliczyć stałą C i wykreślić funkcję gęstości.
2. Wyznaczyć dystrybuantę rozkładu.
3. Obliczyć prawdopodobieństwo, że przedsiębiorstwo osiągnie zysk od 2 do 4 tysięcy złotych.

Rozwiązanie

Ad.1. Stałą C obliczamy z własności:

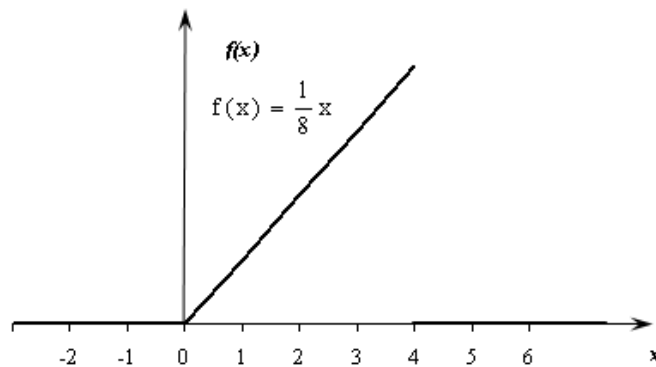
$$\int_{-\infty}^{+\infty} f(x) dx = 1$$

$$\int_{-\infty}^{+\infty} f(x)dx = \int_{-\infty}^0 0dx + \int_0^4 cx dx + \int_4^{+\infty} 0dx = \int_0^4 Cx dx = C \frac{x^2}{2} \Big|_0^4 = C \frac{16}{2} = 8C ;$$

$$8C = 1 \rightarrow C = \frac{1}{8} .$$

Funkcja gęstości ma więc postać:

$$f(x) = \begin{cases} 0 & \text{dla } x < 0 \\ \frac{1}{8}x & \text{dla } 0 \leq x \leq 4 \\ 0 & \text{dla } x > 4 \end{cases}$$



Rys. 2.5.3. Wykres funkcji gęstości ciągłej zmiennej losowej X

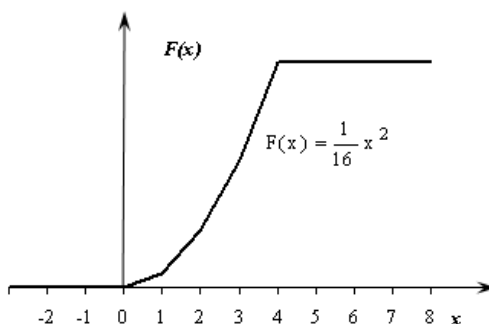
Ad.2. Dystrybuanta

$$P(X < x) = F(x) = \int_{-\infty}^x f(x)dx ,$$

$$F(x) = \int_0^x \frac{1}{8}x dx = \frac{1}{16}x^2 .$$

Ostatecznie postać dystrybuanty jest następująca:

$$F(x) = \begin{cases} 0 & \text{dla } x < 0 \\ \frac{1}{16}x^2 & \text{dla } 0 \leq x \leq 4 \\ 1 & \text{dla } x > 4 \end{cases}$$



Rys. 2.5.4. Wykres dystrybuanty ciągłej zmiennej losowej X

Ad. 3

$$P(2 \leq X \leq 4) = F(4) - F(2) = \frac{1}{16} 4^2 - \frac{1}{16} 2^2 = 1 - \frac{4}{16} = \frac{12}{16} = \frac{3}{4}.$$

Zadanie 3

Wyznaczyć medianę i modalną zmiennej losowej skokowej X o następującym rozkładzie:

x_i	1	2	3	4
$P(X = x_i)$	1/8	2/8	4/8	1/8

Rozwiązanie

Mediana $M_e = 3$, ponieważ liczba ta spełnia nierówności:

$$P(X \leq 3) = \frac{7}{8} > \frac{1}{2} \text{ i } P(X \geq 3) = \frac{5}{8} > \frac{1}{2}.$$

Z rozkładu prawdopodobieństwa zmiennej losowej X wynika bezpośrednio, że $M_o = 3$, ponieważ ta wartość zmiennej losowej występuje z największym prawdopodobieństwem. Zatem w rozpatrywanym przykładzie $M_e = M_o = 3$.

Zadanie 4

Funkcja gęstości stażu pracy osób zatrudnionych na kierowniczych stanowiskach jest następująca:

$$f(x) = \begin{cases} 0 & \text{dla } x < 0 \\ \frac{1}{72} x^2 & \text{dla } 0 \leq x < 6 \text{ lat} \\ 0 & \text{dla } x > 6 \text{ lat} \end{cases}$$

Obliczyć wartość oczekiwaną stażu pracy, wariancję, odchylenie standardowe i współczynnik zmienności.

Rozwiązanie

Wartość oczekiwana wynosi:

$$\begin{aligned} E(X) &= \int_{-\infty}^{+\infty} x \cdot f(x) dx = \int_0^6 x \cdot \frac{1}{72} x^2 dx = \frac{1}{72} \int_0^6 x^3 dx = \frac{1}{72} \cdot \frac{x^4}{4} \Big|_0^6 = \\ &= \frac{1}{72} \left(\frac{6^4}{4} - \frac{0^4}{4} \right) = 4,5. \end{aligned}$$

Przeciętny staż pracy osób zatrudnionych na kierowniczym stanowisku wynosi 4,5 roku.

Wariancja $V(X)$ wynosi:

$$V(X) = E(X^2) - [E(X)]^2 = 21,6 - 4,5^2 = 1,35,$$

gdzie:

$$\begin{aligned} E(X^2) &= \int_{-\infty}^{+\infty} x^2 \cdot f(x) dx = \int_0^6 x^2 \cdot \frac{1}{72} x^2 dx = \frac{1}{72} \int_0^6 x^4 dx = \frac{1}{72} \cdot \frac{x^5}{5} \Big|_0^6 = \\ &= \frac{1}{72} \left(\frac{6^5}{5} - \frac{0^5}{5} \right) = \frac{1}{72} \cdot \frac{7776}{5} = 21,6 \text{ lat.} \end{aligned}$$

Odchylenie standardowe σ wynosi:

$$\sigma = \sqrt{V(X)} = \sqrt{1,35} = 1,16 \text{ lat.}$$

Staż pracy osób na stanowiskach kierowniczych odchyła się średnio od przeciętnego stażu pracy o 1,16 lat.

$$V_{\sigma_7} = \frac{\sigma_7}{E(X)} \cdot 100\% = \frac{1,16}{4,50} \cdot 100\% = 25,8\%.$$

Średnie odchylenie stanowi 25,8% przeciętnego stażu pracy.

Zadanie 5

Dany jest rozkład prawdopodobieństwa zmiennej losowej X

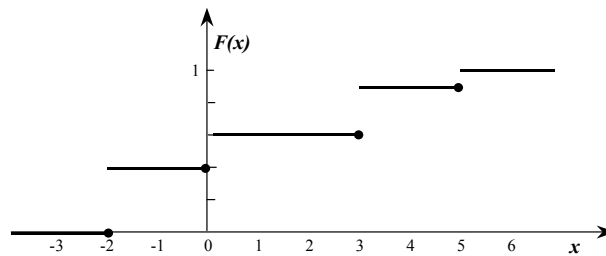
x_i	-2	0	3	5
p_i	c	0,2	0,3	0,1

Wyznaczyć c, dystrybuantę i jej wykres, wartość oczekiwaną, odchylenie standardowe.

Rozwiązanie

$$\sum_{i=1}^4 p_i = 1 \text{ stąd } c = 0,4.$$

$$F(x) = \begin{cases} 0 & \text{dla } x \leq -2 \\ 0,4 & \text{dla } -2 < x \leq 0 \\ 0,6 & \text{dla } 0 < x \leq 3 \\ 0,9 & \text{dla } 3 < x \leq 5 \\ 1 & \text{dla } x > 5 \end{cases}$$



Rys.2.5.5. Wykres dystrybuanty

$$E(X) = \sum_{i=1}^4 x_i \cdot p_i = 0,6.$$

$$V(X) = \sum_{i=1}^4 (x_i - E(X))^2 \cdot p_i = 6,44.$$

$$\sigma = \sqrt{V(X)} = 2,54.$$

Zadanie 6

Zbadano 100 gospodarstw domowych ze względu na liczbę osób w gospodarstwie. Wyniki badania przedstawiono w szeregu punktowym.

Liczba osób w gospodarstwie domowym x_i	Liczba gospodarstw n_i
1	15
2	20
3	30
4	25
5	6
6	4
Razem	100

Niech zmienna losowa X oznacza liczbę osób w gospodarstwie domowym. Wyznaczyć wartość oczekiwaną tej zmiennej losowej.

Rozwiązanie

Zmienna losowa X przyjmuje wartości x_i , $i=1, \dots, 6$ z prawdopodobieństwem

$p_i = \frac{n_i}{n}$. Funkcję prawdopodobieństwa zmiennej losowej X przedstawia

tabela

i	1	2	3	4	5	6
p_i	0,15	0,20	0,30	0,25	0,06	0,04

A zatem:

$$E(X) = \sum_{i=1}^6 x_i p_i = 1 \cdot 0,5 + 2 \cdot 0,2 + 3 \cdot 0,3 + 4 \cdot 0,25 + 5 \cdot 0,06 + 6 \cdot 0,04 = 2,99.$$

Zauważmy, że wzór na wartość oczekiwaną przybiera postać:

$$\sum_{i=1}^6 x_i p_i = \sum_{i=1}^6 x_i \frac{n_i}{n} = \frac{1}{n} \sum_{i=1}^6 x_i n_i,$$

a więc wartość oczekiwana jest równa średniej ważonej wariantów zmiennej X (średniej arytmetycznej wszystkich danych statystycznych).

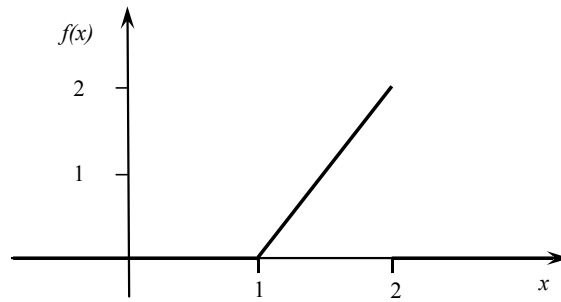
Zadanie 7

Wyznaczyć funkcję gęstości prawdopodobieństwa wiedząc, że jej dystrybucję przedstawia wzór:

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

Rozwiązanie

$$f(x) = F'(x) = \begin{cases} 0 & \text{dla } x < 1 \\ 2x - 2 & \text{dla } 1 \leq x < 2 \\ 0 & \text{dla } x \geq 2 \end{cases}$$



Rys.2.5.6. Wykres funkcji gęstości

Funkcja $f(x)$ jest funkcją gęstości, ponieważ:

- 1) $f(x) \geq 0$ (co widać na rysunku),
- 2) $\int_1^2 (2x - 2) dx = \left[2 \cdot \frac{x^2}{2} - 2x \right]_1^2 = x^2 - 2x \Big|_1^2 = 1$.

ROZDZIAŁ III WYBRANE ROZKŁADY ZMIENNEJ LOSOWEJ

3.1. Rozkłady skokowej zmiennej losowej

Rozkład skokowej zmiennej losowej X jest znany, jeżeli określone są prawdopodobieństwa p_i , z jakimi zmienna ta przyjmuje poszczególne wartości, czyli gdy znany jest zbiór par (x_i, p_i) . Wartości x_i przyjmowane przez skokową zmienną losową X nazywamy punktami skokowymi, a prawdopodobieństwa p_i skokami tej zmiennej. Funkcję, która przyporządkowuje realizacjom skokowej zmiennej losowej X , odpowiadające im prawdopodobieństwa, nazywamy rozkładem prawdopodobieństwa zmiennej losowej X . Rozkład prawdopodobieństwa skokowej zmiennej losowej można przedstawić za pomocą wzoru, tabeli lub wykresu.

3.1.1. Rozkład zero-jedynkowy

Zmienna losowa przyjmująca tylko dwie wartości, którym możemy przyporządkować cyfry zero oraz jeden nazywa się zmienną zero-jedynkową. Doświadczenie polegające na rzucie monetą można opisać zmienną losową zero-jedynkową, przyjmując za sukces wyrzucenie „orła”, któremu to zdarzeniu przypiszemy liczbę „jeden” oraz przyjmując za porażkę wyrzucenie „reszki” i przypisując temu zdarzeniu liczbę „zero”.

Zmienna losowa X ma rozkład zero-jedynkowy, jeśli jej funkcja rozkładu prawdopodobieństwa określona jest wzorem:

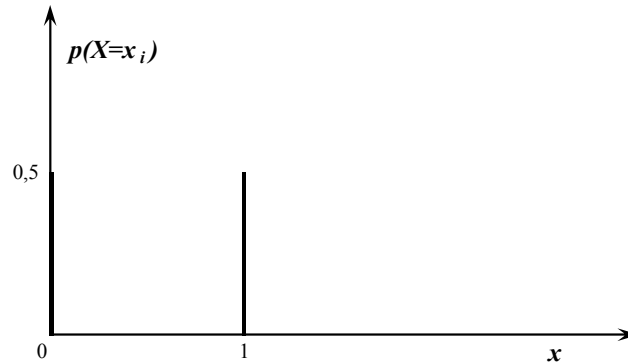
$$\begin{aligned}P(X = 1) &= p, \\P(X = 0) &= q, \\p + q &= 1.\end{aligned}$$

gdzie: p – sukces, q – porażka.

W przypadku rzutu monetą niefałszywą (której masa jest równomiernie rozłożona na obu powierzchniach) rozkład zmiennej losowej zero-jedynkowej ma postać:

$$\begin{aligned}P(X = 1) &= 0,5, \\P(X = 0) &= 0,5,\end{aligned}$$

co graficznie można przedstawić następująco:



Rys. 3.1.1.1. Rozkład zmiennej losowej zero-jedynkowej opisującej rzut monetą

Dystrybuanta rozkładu zero-jedynkowego ma postać:

$$F(x) = \begin{cases} 0 & \text{dla } x \leq 0 \\ q & \text{dla } 0 < x \leq 1. \\ 1 & \text{dla } x > 1 \end{cases}$$

Wartość oczekiwana jest równa prawdopodobieństwu sukcesu p :

$$E(X) = \sum_{i=1}^n x_i \cdot p_i = 1 \cdot p + 0 \cdot q = p.$$

Wariancja jest równa:

$$\begin{aligned} V(X) &= \sum_{i=1}^n x_i^2 \cdot p_i - [E(X)]^2 = \\ &= 1^2 \cdot p + 0^2 \cdot q - p^2 = p - p^2 = p(1 - p) = p \cdot q. \end{aligned}$$

Odchylenie standardowe wynosi:

$$\sigma = \sqrt{V(X)} = \sqrt{p \cdot q}.$$

3.1.2. Rozkład dwumianowy

W wielu praktycznych zagadnieniach należy określić prawdopodobieństwo kilkakrotnego zajścia zdarzenia w wielokrotnie powtarzonym eksperymencie, gdy zdarzenie to zachodzi ze stałym prawdopodobieństwem sukcesu oraz porażki. Eksperyment taki nosi nazwę schematu Bernoulliego.

Jeżeli pewne zdarzenie zachodzi wielokrotnie (k razy) w określonej liczbie $n \geq 2$ niezależnych doświadczeń (tj. takich, że wynik poprzedniego do-

świadczania nie wpływa na wynik następnego), ze stałym prawdopodobieństwem sukcesu p , wtedy zmienna losowa opisująca to zdarzenie ma rozkład prawdopodobieństwa dwumianowy, nazywany też rozkładem Bernoulliego. Rozkład dwumianowy może mieć zastosowanie między innymi w badaniach sympatii politycznych obywateli, jeżeli chcemy określić prawdopodobieństwo tego, że z próby liczącej n wyborców k osób poprze kandydata danej partii lub, że na 50 wysłanych ankiet badawczych otrzymamy 35 odpowiedzi.

Zmienna losowa X ma rozkład dwumianowy, jeśli jej funkcja rozkładu prawdopodobieństwa określona jest następująco:

$$P(X = k) = \binom{n}{k} \cdot p^k \cdot q^{n-k}; \quad k = 0, 1, 2, 3, \dots, n$$

$$\binom{n}{k} = \frac{n!}{k!(n-k)!},$$

przy czym: $n! = 1 \cdot 2 \cdot 3 \cdot \dots \cdot n$; $0! = 1$; $p + q = 1$,

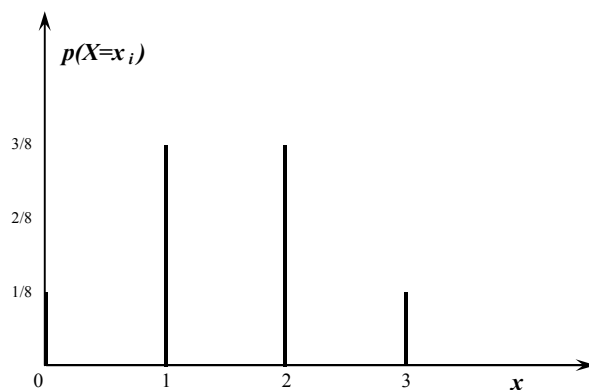
gdzie: n – liczba doświadczeń,

k – liczba doświadczeń zakończonych sukcesem,

p – prawdopodobieństwo sukcesu,

q – prawdopodobieństwo porażki.

Postać funkcji rozkładu dwumianowego zależy od liczby n oraz prawdopodobieństwa sukcesu p . Jeżeli $p = q = 0,5$ rozkład jest symetryczny. Dla $p \neq q$ rozkład ten jest asymetryczny, przy czym: dla $p < 0,5$ rozkład jest prawostronnie asymetryczny, a dla $p > 0,5$ lewostronnie asymetryczny.



Rys. 3.1.2.1. Wykres rozkładu prawdopodobieństwa dla $p = 0,5$; $k = 3$

Dystrybuantę rozkładu dwumianowego oblicza się z zależności:

$$F(x) = \sum_{k < x} \binom{n}{k} \cdot p^k \cdot q^{n-k}.$$

Wartość oczekiwana wynosi:

$$E(X) = n \cdot p.$$

Wariancja jest równa iloczynowi:

$$V(X) = n \cdot p \cdot q.$$

Przykład 3.1.2.1

Prawdopodobieństwo trafienia do celu w jednym strzale wynosi $\frac{2}{3}$. Strzelec oddał niezależnie 4 strzały. Obliczyć prawdopodobieństwo, że cel został trafiony:

- a) jeden raz,
- b) ani razu,
- c) co najmniej raz,
- d) co najwyżej raz.

Rozwiązanie

Sukcesem jest trafienie do celu w jednym strzale, a X jest zmienną losową oznaczającą liczbę celnych spośród 4 strzałów. Zmienna losowa ma rozkład dwumianowy z parametrami: $n = 4$ i $p = \frac{2}{3}$, stąd $q = \frac{1}{3}$.

a) $P(X=1) = \binom{4}{1} \cdot \left(\frac{2}{3}\right) \cdot \left(\frac{1}{3}\right)^3 = 4 \cdot \frac{2}{3} \cdot \frac{1}{3^3} = \frac{8}{3} \cdot \frac{1}{27} = \frac{8}{81} = 0,0988.$

b) $P(X=0) = \binom{4}{0} \cdot \left(\frac{2}{3}\right)^0 \cdot \left(\frac{1}{3}\right)^4 = 1 \cdot 1 \cdot \frac{1}{3^4} = \frac{1}{81} = 0,0123.$

c) $P(X \geq 1) = 1 - P(X < 1) = 1 - P(X = 0) = 1 - 0,0123 = 0,9877.$

d) $P(X \leq 1) = P(X = 1) + P(X = 0) = 0,0988 + 0,0123 = 0,1111.$

Przykład 3.1.2.2

Prawdopodobieństwo, że statystyczny student nie jest przygotowany do ćwiczeń wynosi $\frac{1}{3}$. Obliczyć prawdopodobieństwo $P(X = 3)$, że wśród czterech osób wywołanych do odpowiedzi trzy są nieprzygotowane do ćwiczeń.

Rozwiązanie

Zmienna losowa ma rozkład dwumianowy: $n = 4$, $k = 3$, $p = \frac{1}{3}$, $q = \frac{2}{3}$.

$$P(X = k) = \binom{n}{k} \cdot p^k \cdot q^{n-k},$$

$$P(X = 3) = \binom{4}{3} \cdot \left(\frac{1}{3}\right)^3 \cdot \left(\frac{2}{3}\right)^{4-3} = \frac{4!}{3! \cdot 1!} \cdot \frac{1}{27} \cdot \frac{2}{3} = \frac{8}{81}.$$

Przykład 3.1.2.3

Wiadomo, że 1% skrzynek bananów psuje się w czasie transportu. Z transportu w sposób losowy wybrano 3 skrzynki. Niech X oznacza liczbę skrzynek z zepsutymi bananami spośród trzech wybranych. Znaleźć rozkład prawdopodobieństwa zmiennej losowej X .

Rozwiązanie

Zmienna losowa ma rozkład dwumianowy $n=3$, $k=0,1,2,3$, $p=0,01$, $q=0,99$.

$$P(X = k) = \binom{n}{k} \cdot p^k \cdot q^{n-k},$$

$$P(X = 0) = \binom{3}{0} \cdot 0,01^0 \cdot 0,99^3 = 1 \cdot 1 \cdot 0,99^3 = 0,970299,$$

$$P(X = 1) = \binom{3}{1} \cdot 0,01^1 \cdot 0,99^2 = 3 \cdot 0,01 \cdot 0,99^2 = 0,029403,$$

$$P(X = 2) = \binom{3}{2} \cdot 0,01^2 \cdot 0,99^1 = 3 \cdot 0,0001 \cdot 0,99^1 = 0,000297,$$

$$P(X = 3) = \binom{3}{3} \cdot 0,01^3 \cdot 0,99^0 = 1 \cdot 0,0001 \cdot 0,9^0 = 0,000001.$$

3.1.3. Rozkład Poissona

Rozkład Poissona jest szczególnym przypadkiem rozkładu dwumianowego, zachodzącym wtedy, gdy prawdopodobieństwo p sukcesu jest małe, a liczba realizacji n na tyle duża, że $n \cdot p = \lambda$ jest wielkością stałą. Przyjmuje się, że zmienna losowa ma rozkład Poissona wtedy, gdy $n > 100$ i $p < 0,2$. Rozkład ten dobrze opisuje doświadczenia losowe (eksperymenty), w których obserwujemy dużą serię zdarzeń, zachodzących z małym prawdopodobieństwem sukcesu w pojedynczych obserwacjach, np. przy statystycznej kontroli jakości produkcji.

Zmienna losowa ma rozkład Poissona, jeśli jej rozkład prawdopodobieństwa określony jest wzorem:

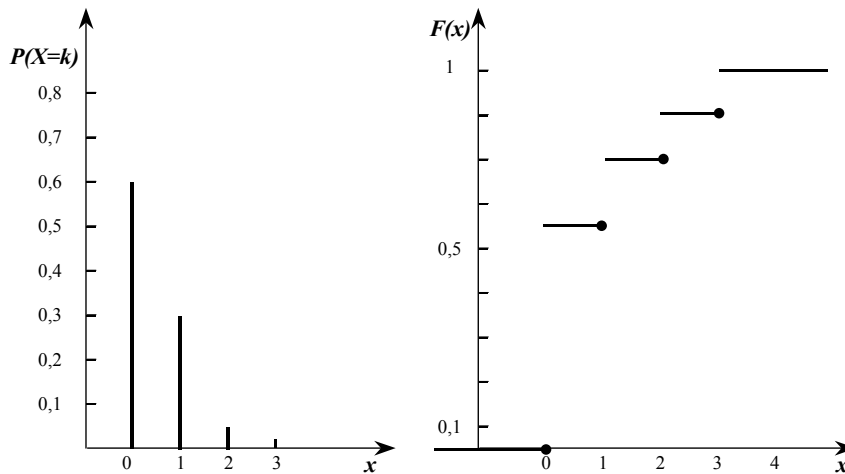
$$P(X = k) = \frac{\lambda^k}{k!} \cdot e^{-\lambda}; \quad k = 0, 1, 2, 3, \dots; \quad \lambda = n \cdot p; \quad e \approx 2,7.$$

Dystrybuantę rozkładu Poissona oblicza się z zależności:

$$F(x) = \sum_{k < x} \frac{\lambda^k}{k!} \cdot e^{-\lambda}.$$

Wartość oczekiwana $E(X)$ jest równa wariancji i wynosi:

$$E(X) = V(X) = \lambda.$$



Rys. 3.1.3.1. Wykres rozkładu prawdopodobieństwa i dystrybuanty dla $\lambda = 0,5; k = 0,1,2,3$

Przykład 3.1.3.1

W pewnym przedsiębiorstwie zaobserwowano, że w ciągu miesiąca zdarzają się średnio dwa wypadki i że rozkład liczby wypadków może być opisany za pomocą rozkładu Poissona. Obliczyć prawdopodobieństwo, że:

- nie będzie wypadków,
- liczba wypadków nie przekroczy dwóch.

Rozwiązanie

Oznaczamy przez X zmienną losową, która oznacza liczbę wypadków w losowo wybranym miesiącu i może przyjmować wartości $0, 1, 2, 3, \dots$

a) nie będzie wypadków ($k = 0$); przypomnijmy, że $0! = 1$.

$$P(X = k) = \frac{\lambda^k}{k!} \cdot e^{-\lambda} = \frac{\lambda^0}{0!} \cdot e^{-2} = 0,135.$$

b) liczba wypadków nie przekroczy dwóch $k \leq 2$

$$\begin{aligned} P(k \leq 2) &= P(k = 0) + P(k = 1) + P(k = 2) = \\ &= \frac{\lambda^0}{0!} \cdot e^{-2} + \frac{\lambda^1}{1!} \cdot e^{-2} + \frac{\lambda^2}{2!} \cdot e^{-2} = 0,675 \end{aligned}$$

Przykład 3.1.3.2

Prawdopodobieństwo, że produkt poddawany próbie nie wytrzyma tej próby wynosi 0,02. Obliczyć prawdopodobieństwo, że wśród 250 takich produktów trzy nie wytrzymają próby.

Rozwiązanie

Oznaczamy przez X zmienną losową, która określa liczbę produktów nie wytrzymujących próby. Intensywność uszkodzeń produktów w czasie prób wynosi: $\lambda = n \cdot p = 250 \cdot 0,02 = 5$.

Prawdopodobieństwo, że uszkodzą się trzy produkty wynosi:

$$P(X = 3) = \frac{\lambda^k}{k!} \cdot e^{-\lambda} = \frac{5^3}{3!} \cdot e^{-5} = 0,145.$$

3.2. Rozkłady ciągłej zmiennej losowej

Rozkład ciągłej zmiennej losowej X opisuje funkcja gęstości prawdopodobieństwa $f(x)$, określona na zbiorze liczb rzeczywistych, o następujących własnościach:

$$\int_a^b f(x)dx = P(a < X \leq b) \text{ dla dowolnych } a < b,$$

$$F(x) = \int_{-\infty}^x f(x)dx = P(X \leq x) \geq 0.$$

Oznacza to, że gdy znana jest postać funkcji $f(x)$, to dla ciągłej zmiennej losowej X można wyznaczyć prawdopodobieństwo, z jakim zmienna ta przyjmuje dowolną wartość x ze zbioru liczb rzeczywistych.

3.2.1. Rozkład jednostajny

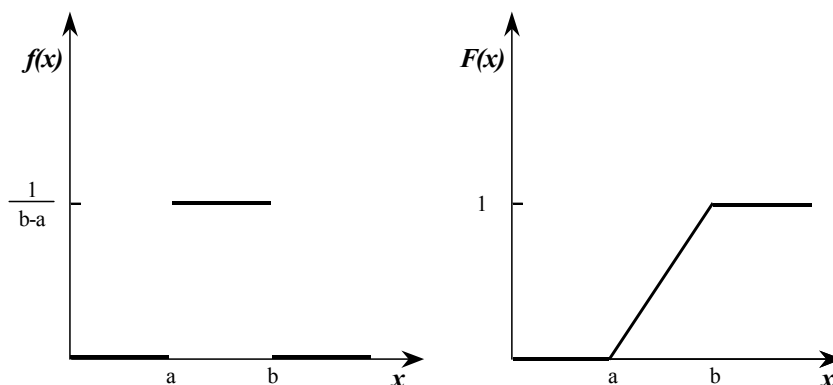
Zmienna losowa X ma rozkład jednostajny (prostokątny, równomierny) w przedziale $[a, b]$, jeżeli funkcja gęstości prawdopodobieństwa tej zmiennej ma następującą postać:

$$f(x) = \begin{cases} 0 & \text{dla } x \leq a \\ \frac{1}{b-a} & \text{dla } a < x \leq b \\ 0 & \text{dla } x > b \end{cases}$$

Dystrybuanta zmiennej losowej o jednostajnej funkcji gęstości prawdopodobieństwa ma postać:

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

Wykresy funkcji gęstości prawdopodobieństwa i dystrybuanty przedstawiono na rysunku 3.2.1.1.



Rys. 3.2.1.1. Wykres funkcji gęstości prawdopodobieństwa i dystrybuanty dla rozkładu jednostajnego

Wartość oczekiwana, wariancja i mediana wynoszą:

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

$$\begin{aligned} V(X) &= E(X)^2 - [E(X)]^2 = \int_a^b x^2 \cdot \frac{1}{b-a} - \frac{(b+a)^2}{4} = \\ &= \frac{1}{b-a} \cdot \frac{x^3}{3} \Big|_a^b - \frac{(b+a)^2}{4} = \frac{b^3 - a^3}{3(b-a)} - \frac{(b+a)^2}{4} = \frac{(b-a)^2}{12}. \end{aligned}$$

$$M_e = \frac{a+b}{2} \text{ (jest to wartość } x, \text{ dla której dystrybuanta ma wartość } 0,5).$$

Zmienna ta nie ma modalnej, ponieważ funkcja gęstości nie ma maksimum.

Przykład 3.2.1.1

Zmienna losowa opisująca okres oczekiwania na wykonanie prototypu pewnego urządzenia ma rozkład jednostajny. Wyznaczyć funkcję gęstości prawdopodobieństwa oraz parametry rozkładu tej zmiennej dla okresu oczekiwania od sześciu do dwunastu miesięcy.

Rozwiązanie

Funkcja gęstości prawdopodobieństwa ma postać:

$$f(x) = \begin{cases} 0 & \text{dla } x \leq 6 \\ \frac{1}{6} & \text{dla } 6 < x \leq 12. \\ 0 & \text{dla } x > 12 \end{cases}$$

Dystrybuanta ma postać:

$$F(x) = \begin{cases} 0 & \text{dla } x \leq 6 \\ \frac{x-6}{6} & \text{dla } 6 < x \leq 12. \\ 1 & \text{dla } x > 12 \end{cases}$$

Wartość oczekiwana, odchylenie standardowe oraz mediana wynoszą:

$$E(X) = \frac{a+b}{2} = \frac{6+12}{2} = 9.$$

$$\sigma = \sqrt{\frac{(b-a)^2}{12}} = \sqrt{\frac{(12-6)^2}{12}} = 1,73.$$

$$M_e = \frac{a+b}{2} = \frac{6+12}{2} = 9.$$

Średni czas oczekiwania na wykonanie prototypu wynosi 9 miesięcy. Jest to również najczęściej występujący czas oczekiwania na wykonanie prototypu. Odchylenie standardowe od średniego czasu oczekiwania na wykonanie prototypu wynosi około 1,73 miesiąca.

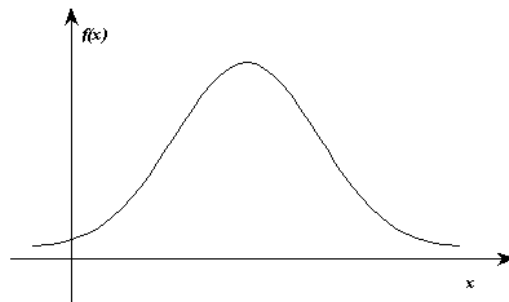
3.2.2. Rozkład normalny

Wśród rozkładów prawdopodobieństwa ciągłych zmiennych losowych największe znaczenie w statystyce ma **rozkład normalny (Gaussa-Laplace'a)**. Rozkład ten ma najszerze zastosowanie praktyczne, ponieważ wiele cech statystycznych charakteryzuje rozkład normalny lub można je opisać za pomocą tego rozkładu, aproksymując inne rozkłady prawdopodobieństwa, gdy rzeczywisty rozkład cechy jest znany, ale obliczenia według rzeczywistego rozkładu są uciążliwe. Istotną cechą rozkładu normalnego jest to, że równanie matematyczne rozkładu pozwala określić prawdopodobieństwo zaistnienia danego zdarzenia jako pole pod krzywą pomiędzy dwoma punktami na osi poziomej (zdarzeń), gdy znana jest wartość średnia m oraz odchylenie standardowe σ . Dalsze udogodnienie polega na tym, że wartości tego pola dla różnych wartości średnich i odchylenia standardowego są podawane w tablicach statystycznych.

Zmienna losowa X ma rozkład normalny z parametrami m i σ , co oznaczamy $N(m, \sigma)$, jeśli jej funkcja gęstości jest określona wzorem:

$$f(x) = \frac{1}{\sigma\sqrt{2\pi}} \cdot e^{-\frac{(x-m)^2}{2\sigma^2}}; \text{ dla } x \in (-\infty, +\infty).$$

Wykres rozkładu normalnego ma kształt dzwonu, symetrycznego wokół wartości średniej, którego ramiona zbliżają się asymptotycznie do osi poziomej w obu kierunkach, tj. od $-\infty$ do $+\infty$.



Rys. 3.2.2.1. Wykres funkcji gęstości prawdopodobieństwa dla rozkładu normalnego

Własności funkcji $f(x)$:

Jest symetryczna względem prostej $x = m$, co znaczy, że spełniona jest zależność:

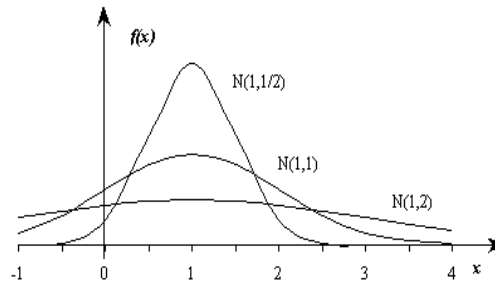
$$P(X > m) = P(X < m) = 0,5.$$

W punkcie m osiąga wartość maksymalną, która wynosi:

$$f(m) = \frac{1}{\sigma\sqrt{2\pi}}.$$

Prawdopodobieństwo, że zmienna losowa X przyjmie wartości z przedziału $[m - 3\sigma, m + 3\sigma]$ jest w przybliżeniu równe 1.

Wartość odchylenia standardowego σ decyduje o skupieniu wartości zmiennej losowej wokół jej średniej. Im mniejsza wartość σ , tym wartości zmiennej losowej układają się bliżej wartości średniej. Wykres funkcji rozkładu prawdopodobieństwa dla różnych wartości odchylenia standardowego σ przedstawiono na rysunku 3.2.2.2.



Rys. 3.2.2.2. Wykresy funkcji rozkładu prawdopodobieństwa dla przypadków $N(1,1)$, $N(1,2)$ oraz $N(1,0,5)$

Dystrybuantę rozkładu normalnego zmiennej losowej X oblicza się z zależności:

$$F(x) = \frac{1}{\sigma\sqrt{2\pi}} \int_{-\infty}^x e^{-\frac{(x-m)^2}{2\sigma^2}} dx.$$

Pozostałe charakterystyki liczbowe dla rozkładu normalnego zmiennej losowej wynoszą:

- wartość średnia: $E(X) = m$,
- wariancja: $V(X) = \sigma^2$,
- mediana: $M_e = m$,
- modalna: $M_o = m$.

Standaryzowana zmienna losowa U

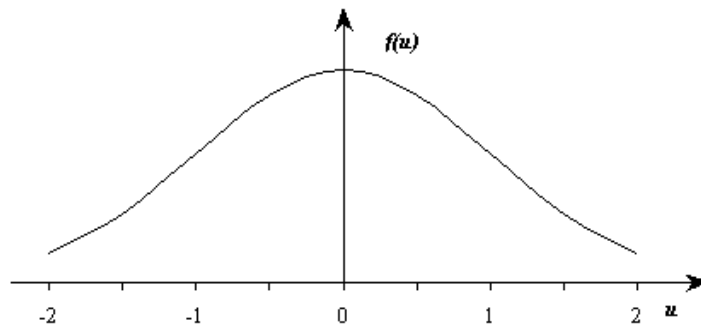
Standaryzowaną zmienną losową U otrzymuje się z przekształcenia zmiennej losowej X o normalnym rozkładzie prawdopodobieństwa, w następujący sposób:

$$U = \frac{X - m}{\sigma},$$

której funkcja gęstości ma następującą postać:

$$f(u) = \frac{1}{\sqrt{2\pi}} \cdot e^{-\frac{u^2}{2}}; u \in (-\infty, +\infty),$$

o następujących parametrach $N(0, 1)$.



Rys. 3.2.2.3. Wykres funkcji gęstości prawdopodobieństwa dla rozkładu normalnego standaryzowanego

Dystrybuanta zmiennej losowej U ma postać:

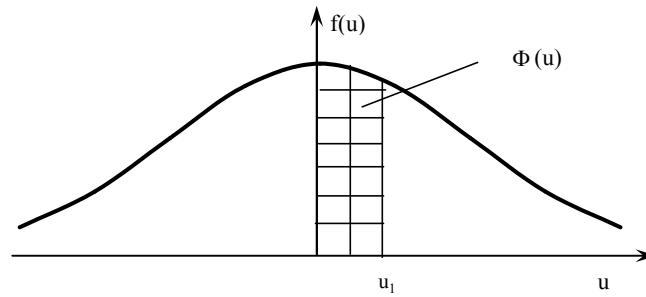
$$F(u) = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^u e^{-\frac{t^2}{2}} dt.$$

Wartości dystrybuanty $F(u)$ są podawane w tablicach statystycznych, przy czym najczęściej podawana jest wartość dystrybuanty w postaci skróconej w formie następującej funkcji:

$$\Phi(u) = \frac{1}{\sqrt{2\pi}} \int_0^u e^{-\frac{t^2}{2}} dt,$$

która określona jest tylko dla dodatnich wartości u.

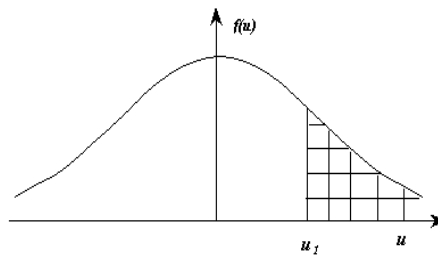
Graficznie wartość funkcji $\Phi(u)$ można przedstawić w postaci pola pod krzywą funkcji gęstości dla $0 < t < u$.



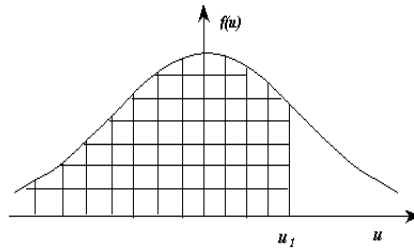
Rys. 3.2.2.4. Wykres funkcji gęstości prawdopodobieństwa dla rozkładu normalnego standardyzowanego

Wartości prawdopodobieństw, że zmienna losowa U przyjmie konkretną wartość obliczamy korzystając z tego, że funkcja $f(u)$ jest symetryczna względem prostej przechodzącej przez punkt $u = 0$ i pole pod krzywą gęstości prawdopodobieństwa jest równe 1, wówczas:

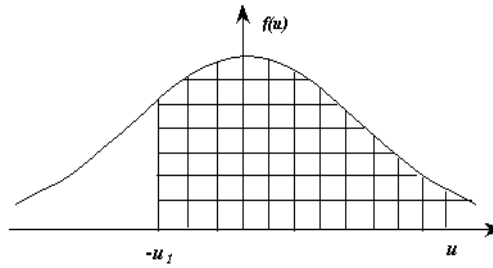
$$\begin{aligned} P(U < +u) &= 0,5 + \Phi(u), \\ P(U > +u) &= 0,5 - \Phi(u), \\ P(U < -u) &= 0,5 - \Phi(u), \\ P(U > -u) &= 0,5 + \Phi(u). \end{aligned}$$



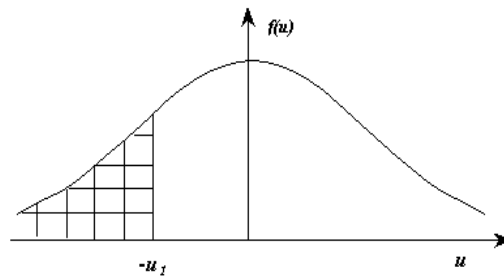
Rys. 3.2.2.5. Wykres dla $P(U > u)$



Rys. 3.2.2.6. Wykres dla $P(U < u)$



Rys. 3.2.2.7. Wykres dla $P(U > -u)$



Rys. 3.2.2.8. Wykres dla $P(U < -u)$

Tablice funkcji $\Phi(u)$ są wykorzystywane do obliczania prawdopodobieństw dla zmiennych losowych o rozkładzie normalnym z dowolnymi parametrami, gdyż każdą zmienną losową o normalnym rozkładzie prawdopodobieństwa można sprowadzić do zmiennej losowej o standaryzowanym rozkładzie $N(0, 1)$.

Dla zmiennej losowej standaryzowanej U przedział $[m - 3\sigma, m + 3\sigma]$ jest następujący: $[-3, +3]$, ponieważ: $m = 0$, a $\sigma = 1$.

Przykład 3.2.2.1

Zmienna losowa X podlega rozkładowi $N(5, 2)$. Obliczyć prawdopodobieństwo $P(X < 3,6)$. Wynik zaznaczyć na wykresie.

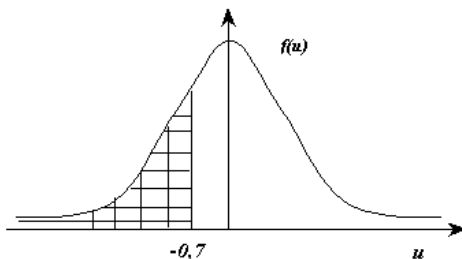
Rozwiązanie

Obliczamy wartość zmiennej losowej standaryzowanej u dla $m = 5$ oraz $\sigma = 2$:

$$u = \frac{X - m}{\sigma} = \frac{3,6 - 5}{2} = -0,7.$$

Odczytujemy z tablic statystycznych wartość dystrybuanty standaryzowanego rozkładu normalnego dla $u = 0,7$:

$$P(X < 3,6) = P(U < -0,7) = 0,5 - \Phi(0,7) = 0,5 - 0,2580 = 0,2420.$$



Rys. 3.2.2.9. Obszar pod krzywą rozkładu normalnego dla $P(X < 3,6)$

Przykład 3.2.2.2

Czas oczekiwania tirów na odprawę na przejściu granicznym (w godzinach) jest zmienną losową X o rozkładzie normalnym $N(8,2)$. Obliczyć prawdopodobieństwo oczekiwania:

- a) mniej niż 5 godzin,
- b) od 6 do 9 godzin,
- c) powyżej 10 godzin.

Rozwiązanie

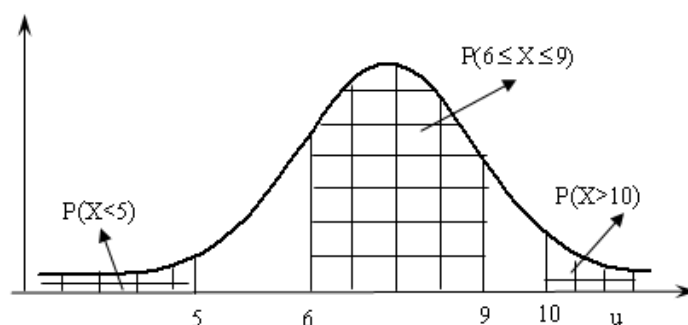
a) $P(X < 5) = P\left(\frac{X - 8}{2} < \frac{5 - 8}{2}\right) = P(U < -1,5) = \Phi(-1,5) = 1 - \Phi(1,5) = 1 - (0,5 + 0,4332) = 1 - 0,7332 = 0,0668$ (6,7%).

Około 6,7% tirów oczekuje na odprawę mniej niż 5 godzin.

b) $P(6 \leq X \leq 9) = P\left(\frac{6 - 8}{2} \leq X \leq \frac{9 - 8}{2}\right) = P(-1 \leq U \leq 0,5) = \Phi(0,5) - \Phi(-1) = \Phi(0,5) + \Phi(1) - 1 = (0,5 + 0,1515) + (0,5 + 0,3413) - 1 = 0,1515 + 0,3413 = 0,5328$ (53,28%).

Około 53,28% tirów oczekuje na odprawę od 6 do 9 godzin.

- c) $P(X > 10) = P\left(\frac{X-8}{2} > \frac{10-8}{2}\right) = P(U > 1) = 1 - \Phi(1) =$
 $= 1 - 0,8413 = 0,1587$ (15,87%).
 Około 15,87% tirów oczekuje na odprawę ponad 10 godzin.



Rys. 3.2.2.10. Graficzna prezentacja rozwiązania

Przykład 3.2.2.3

Długość pewnego detalu jest zmienną losową X o rozkładzie $N(10 \text{ cm}; 0,5 \text{ cm})$. Obliczyć prawdopodobieństwo, że długość detalu zawiera się między 9,7 cm a 10,2 cm.

Rozwiązanie

$$P(9,7 < X < 10,2) = P\left(\frac{9,7-10}{0,5} < X < \frac{10,2-10}{0,5}\right) = P(-0,6 < U < 0,4) =$$

$$\Phi(0,4) - \Phi(-0,6) = \Phi(0,4) - [1 - \Phi(0,6)] = \Phi(0,4) + \Phi(0,6) - 1 =$$

$$= 0,6554 + 0,7257 - 1 = 0,3811.$$

Przykład 3.2.2.4

Zmienna losowa X ma rozkład normalny $N(0, 1)$. Obliczyć prawdopodobieństwa:

- $P(X < -2)$,
- $P(-2 \leq X < 3)$,
- $P(X \geq 3)$.

Rozwiązanie

- $P(X < -2) = \Phi(-2) = 1 - \Phi(2) = 1 - (0,5 + 0,4772) = 1 - 0,9772 = 0,0228$.
- $P(-2 \leq X < 3) = \Phi(3) - \Phi(-2) = \Phi(3) + \Phi(2) - 1 = (0,5 + 0,4987) + (0,5 + 0,772) - 1 = 0,4987 + 0,4772 = 0,9759$.
- $P(X \geq 3) = 1 - P(x < 3) = 1 - \Phi(3) = 1 - (0,5 + 0,4987) = 1 - 9987 = 0,013$.

Przykład 3.2.2.5

Zmienna losowa X ma w populacji rozkład $N(m, 40)$. Znaleźć m wiedząc, że $P(X < 100) = 0,7734$.

Rozwiązanie

Wartość średnią m rozkładu normalnego obliczamy z przekształcenia zależności na zmienną losową standaryzowaną do postaci: $m = X - u \cdot \sigma$.

Ponieważ:

$$P(X < 100) = 0,7734 = 0,5 + 0,2734 = 0,5 + \Phi(u).$$

Wartość u odczytujemy z tablicy dystrybuanty rozkładu normalnego dla $\Phi(u) = 0,2734$ i otrzymujemy: $u = 0,75$.

Stąd $m = X - u \cdot \sigma = 100 - 0,75 \cdot 40 = 70$.

3.2.3. Rozkład wykładniczy

Zmienna losowa X ma wykładniczy rozkład prawdopodobieństwa z parametrem λ , jeżeli jej funkcja gęstości prawdopodobieństwa ma postać:

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

gdzie: λ – średnia intensywność.

Dystrybuanta zmiennej losowej X ma postać:

$$F(x) = \begin{cases} 0 & \text{dla } x \leq 0 \\ 1 - e^{-\lambda \cdot x} & \text{dla } x > 0 \end{cases}$$

Wartość oczekiwana, wariancja oraz mediana wynoszą:

$$E(X) = \frac{1}{\lambda},$$

$$V(X) = \frac{1}{\lambda^2},$$

$$M_e = -\frac{\ln 0,5}{\lambda}.$$

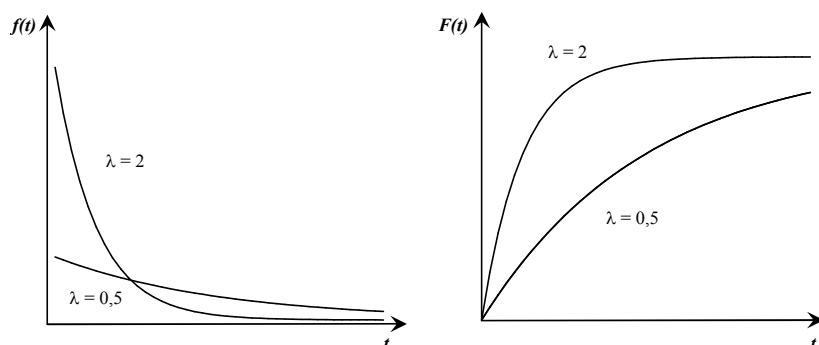
Rozkład wykładniczy znajduje głównie zastosowanie do oceny niezawodności obiektów technicznych. Celowi temu służy funkcja niezawodności $R(t)$, która określa prawdopodobieństwo tego, że czas T bezawaryjnej pracy obiektu nie będzie krótszy niż założona wartość czasu t .

$$R(t) = P(T \geq t) = 1 - P(T < t).$$

Stąd dla rozkładu wykładniczego mamy:

$$R(t) = 1 - F(t) = 1 - (1 - e^{-\lambda \cdot t}) = e^{-\lambda \cdot t}.$$

Wykresy funkcji gęstości prawdopodobieństwa i dystrybuanty przedstawiono na rysunku 3.2.3.1.



Rys. 3.2.3.1. Wykres funkcji gęstości prawdopodobieństwa i dystrybuanty dla rozkładu wykładniczego

Przykład 3.2.3.1

Na podstawie analizy dokumentacji eksploatacyjnej stwierdzono, że czas bezawaryjnej pracy komputerów w pewnej pracowni komputerowej wynosił średnio 200 godzin, oraz że rozkład zmiennej losowej opisującej czas pracy komputerów pomiędzy kolejnymi awariami najbardziej zbliżony jest do rozkładu wykładniczego. Wyznaczyć funkcję gęstości prawdopodobieństwa, dystrybuantę, prawdopodobieństwo, że czas pracy komputera nie będzie krótszy niż 100 godzin oraz podstawowe parametry rozkładu.

Rozwiązanie

Wartość oczekiwana czasu pracy komputera wynosi $E(X) = 200$ godzin. Parametr λ oznaczający intensywność uszkodzeń wynosi:

$$\lambda = \frac{1}{E(X)} = \frac{1}{200} = 0,005.$$

Funkcja gęstości prawdopodobieństwa oraz dystrybuanta wynoszą:

$$f(x) = \begin{cases} 0 & \text{dla } x < 0 \\ 0,005 \cdot e^{-0,005x} & \text{dla } x \geq 0 \end{cases},$$

$$F(x) = \begin{cases} 0 & \text{dla } x \leq 0 \\ 1 - e^{-0,005x} & \text{dla } x > 0 \end{cases}$$

Prawdopodobieństwo bezawaryjnej pracy komputera w czasie 100 godzin wynosi: $R(t) = e^{-\lambda \cdot t} = e^{-0,005 \cdot 100} = e^{-0,5} = 0,6$.

$$\text{Wariancja wynosi: } V(X) = \frac{1}{\lambda^2} = \frac{1}{0,005^2} = 4000.$$

$$M_e = -\frac{\ln 0,5}{\lambda} = -\frac{\ln 0,5}{0,005} = 136,8.$$

3.2.4. Rozkład χ^2 (chi kwadrat)

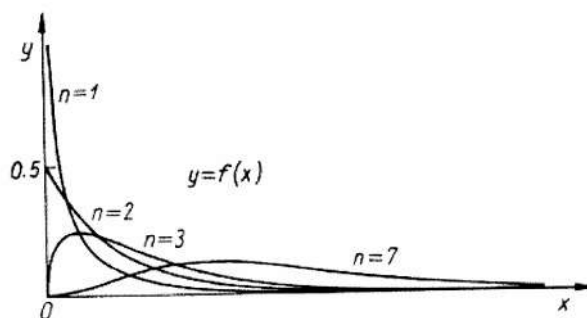
Jeżeli zmienne X_1, X_2, \dots, X_k są niezależnymi zmiennymi losowymi o rozkładzie normalnym $N(0,1)$ dla $i = 1, 2, \dots, k$, to suma kwadratów tych zmiennych $X_1^2 + X_2^2 + \dots + X_k^2$ jest ciągłą zmienną losową o rozkładzie chi-kwadrat χ^2 z $n = k$ stopniami swobody. Funkcja gęstości prawdopodobieństwa tej zmiennej ma postać:

$$f(x) = \begin{cases} 0 & x \leq 0 \\ \frac{1}{2^{0,5n} \cdot \Gamma(0,5n)} \cdot x^{0,5n-1} \cdot e^{-0,5x} & \text{dla } x > 0 \end{cases}$$

gdzie:

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

Postać funkcji gęstości rozkładu χ^2 zależy od liczby stopni swobody n . Dla małych wartości n rozkład jest silnie asymetryczny. Ze wzrostem liczby stopni swobody rozkład jest bardziej symetryczny, zbliżony do rozkładu normalnego.



Rys. 3.2.4.1. Wykres funkcji gęstości rozkładu χ^2 o $n = 1, 2, 3$ stopniach swobody

Wartość oczekiwana i wariancja zmiennej losowej o rozkładzie χ^2 wynoszą: $E(\chi^2) = n$, $V(\chi^2) = 2n$.

Z uwagi na problemy rachunkowe powstające przy wyznaczaniu wartości funkcji gęstości rozkładu χ^2 opracowano tabele, z których dla określonej liczby stopni swobody $n \leq 30$ oraz określonego prawdopodobieństwa α odczytuje się wartość χ_α^2 , dla której spełniony jest warunek:

$$P(\chi_n^2 \geq \chi_\alpha^2) = \alpha.$$

Właściwość ta wykorzystywana jest przy wnioskowaniu statystycznym podczas wyznaczania przedziałów ufności, testowaniu i weryfikacji hipotez statystycznych.

Jeżeli liczba stopni swobody przekracza 30, wówczas należy korzystać z rozkładu zmiennej losowej $\sqrt{2 \cdot \chi_n^2}$, mającej rozkład zbliżony do rozkładu normalnego o parametrach $N(\sqrt{2n-1}, 1)$.

Przykład 3.2.4.1

Wiedząc, że zmienna losowa X ma rozkład chi-kwadrat z trzema stopniami swobody (χ_3^2) obliczyć:

- a) prawdopodobieństwa: $P(X > 0,584)$ i $P(X \leq 4,642)$,
- b) wartości x_α zmiennej losowej X dla przyjętej wartości α :

$$P(X < x_\alpha) = 0,80 \text{ i } P(X > x_\alpha) = 0,05.$$

Rozwiązanie

a) wartość prawdopodobieństwa $P(X > 0,584) = P(\chi_3^2 > 0,584)$ odczytujemy bezpośrednio z tabeli rozkładu. W wierszu dla trzech stopni swobody

należy odnaleźć wartość $\chi_3^2 = 0,584$. Z główki kolumny, w której występuje wartość 0,584, odczytujemy prawdopodobieństwo $\alpha = 0,90$.

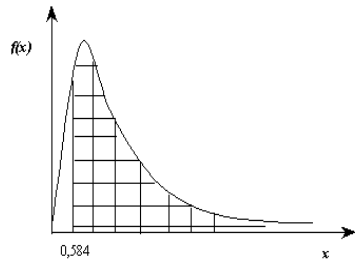
Wartość prawdopodobieństwa

$$P(X \leq 4,462) = 1 - P(X > 4,462) = 1 - P(\chi_3^2 > 4,642).$$

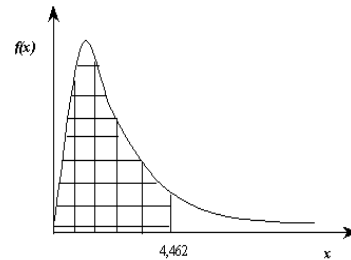
Z tabeli rozkładu χ^2 odczytujemy wartość $P(\chi_3^2 > 4,642) = 0,20$.

Stąd:

$$P(X \leq 4,642) = 1 - 0,20 = 0,80.$$



Rys. 3.2.4.2. Wykres $P(X > 0,584)$



Rys. 3.2.4.3. Wykres $P(X \leq 4,642)$

b) wartości zmiennej losowej X większe od x_α , które zmienna ta przyjmuje z prawdopodobieństwem $\alpha = 0,8$ odczytujemy z tablic rozkładu chi-kwadrat na przecięciu wiersza dla $n = 3$ oraz kolumny dla $\alpha = 0,8$. Otrzymujemy $x_\alpha = 1,005$.

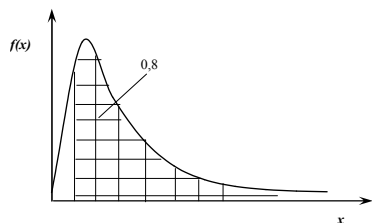
Aby wyznaczyć wartość x_α , dla której $P(X > x_\alpha) = 0,05$ korzystamy z przekształcenia:

$$P(X < x_\alpha) = 1 - P(X \geq x_\alpha) = 1 - P(\chi_3^2 \geq x_\alpha).$$

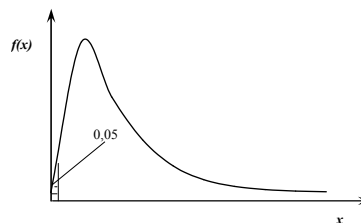
Stąd:

$$P(\chi_3^2 \geq x_\alpha) = 0,95.$$

Odczytując z tabeli rozkładu χ^2 wartość χ_3^2 dla trzech stopni swobody oraz $\alpha = 0,95$, otrzymujemy szukaną wartość $x_\alpha = 0,352$.



Rys. 3.2.4.4. Wykres
 $P(X < x_\alpha) = 0,80$



Rys. 3.2.4.5. Wykres
 $P(X > x_\alpha) = 0,05$

3.2.5. Rozkład t-Studenta

Ciągła zmienna losowa T określona następująco:

$$T = \frac{U}{\sqrt{\chi_n^2}} \sqrt{n},$$

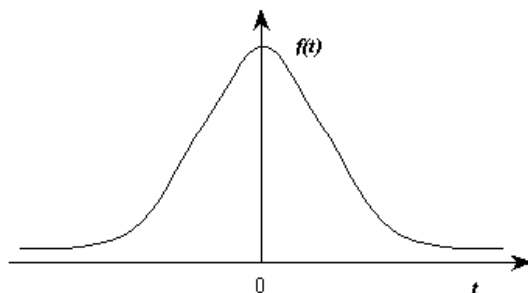
gdzie:

- U oraz χ_n^2 są niezależnymi zmiennymi losowymi,
- zmienna losowa U ma rozkład $N(0,1)$,
- zmienna losowa χ_n^2 ma rozkład chi-kwadrat z n stopniami swobody ma rozkład Studenta z n stopniami swobody.

Funkcja gęstości prawdopodobieństwa tej zmiennej określona jest wzorem:

$$f(t) = \frac{\Gamma(0,5(n+1))}{\sqrt{n} \cdot \pi \cdot \Gamma(0,5n)} \cdot \left(1 + \frac{t^2}{n}\right)^{-0,5(n+1)}.$$

Wykres funkcji gęstości $f(t)$ zmiennej losowej o rozkładzie Studenta, zwanym często rozkładem t-Studenta (rys. 3.2.5.1), jest symetryczny względem osi rzędnych ($t = 0$). Dla stopni swobody $n > 30$ rozkład t-Studenta można aproksymować rozkładem normalnym, ponieważ funkcja gęstości $f(t)$ dla $n \rightarrow \infty$ dąży do rozkładu normalnego standaryzowanego $N(0,1)$.



Rys. 3.2.5.1. Wykres funkcji gęstości $f(t)$ rozkładu t-Studenta

Wartość oczekiwana i wariancja zmiennej losowej o rozkładzie t-Studenta wynosi odpowiednio:

$$E(T) = 0,$$

$$V(T) = \frac{n}{n-2}.$$

Wartości t_α zmiennej losowej T o rozkładzie t-Studenta spełniające warunek:

$$P(|T_n| > t_\alpha) = \alpha$$

zostały tablicowane dla liczby stopni swobody $n \leq 30$ oraz wybranych wartości prawdopodobieństwa α . Jeżeli liczba stopni swobody jest większa od 30, wówczas należy korzystać z tablic rozkładu normalnego standaryzowanego $N(0,1)$.

Przykład 3.2.5.1

Na podstawie wyników eksperymentu badawczego procesu użytkowania 13 obiektów ustalono, że proces ten opisuje zmienna losowa T o rozkładzie t-Studenta. Obliczyć:

- prawdopodobieństwa: $P(|T| > 0,394)$, $P(T > 0,128)$, $P(|T| \leq 1,079)$;
- wartości t_α zmiennej losowej T dla przyjętej wartości α : $P(|T| > t_\alpha) = 0,20$ oraz $P(T > t_\alpha) = 0,80$.

Rozwiązanie

a) prawdopodobieństwo, że wartość bezwzględna zmiennej losowej T będzie większa od 0,394 wynosi:

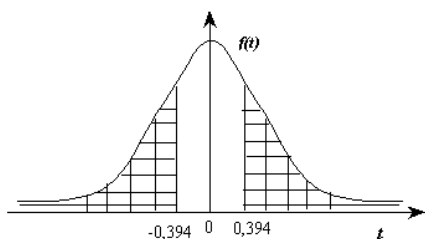
$$P(|T| > 0,394) = P|T_{13}| > 0,394).$$

Wartość prawdopodobieństwa α odczytujemy z tablicy rozkładu t-Studenta w wierszu, który odpowiada 13 stopniom swobody dla wartości $t_\alpha = 0,394$. Stąd $\alpha = 0,7$ (rys. 3.2.5.2).

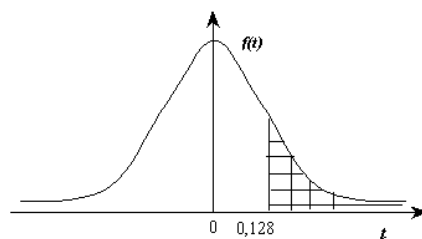
Prawdopodobieństwo $P(T > 0,128)$ obliczamy w następujący sposób. Najpierw z tablicy rozkładu t-Studenta odczytujemy wartość prawdopodobieństwa α dla $t_\alpha = 0,128$, które wynosi 0,9. Ponieważ jest to wartość prawdopodobieństwa dla bezwzględnej wartości zmiennej losowej T, czyli jest to prawdopodobieństwo przyjęcia przez tę zmienną wartości $t_\alpha = 0,128$ lub $t_\alpha = -0,128$, stąd szukana wartość prawdopodobieństwa:

$$P(T > 0,128) = \frac{P(|T| > 0,128)}{2} = \frac{0,90}{2} = 0,45,$$

co wynika z symetrii funkcji gęstości rozkładu t-Studenta. Graficznie wartość prawdopodobieństwa $P(T > 0,128)$ zaznaczono na rysunku 3.2.5.3.



Rys. 3.2.5.2. Wykres $P(|T_{13}| > 0,394)$



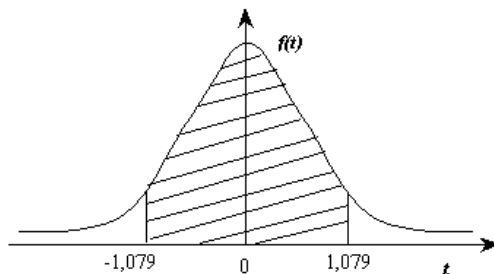
Rys. 3.2.5.3. Wykres $P(T_{13} > 0,128)$

Prawdopodobieństwo $P(|T| \leq 1,079)$ obliczamy korzystając z zależności:

$$P(|T| \leq 1,079) = 1 - P(|T| > 1,079) = 1 - P(|T_{13}| > 1,079) = 1 - 0,3 = 0,7.$$

Wartość prawdopodobieństwa $P(|T_{13}| > 1,079) = 0,3$ odczytujemy bezpośrednio z tablic rozkładu t-Studenta.

Wartość $P(|T_{13}| \leq 1,079) = 0,7$ zaznaczono na rysunku 3.2.5.4.



Rys. 3.2.5.4. Wykres $P(|T_{13}| \leq 1,079)$

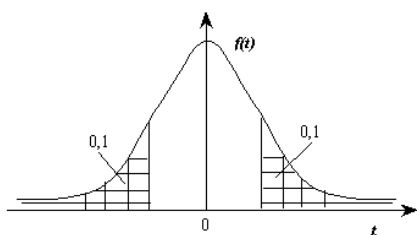
b) wartość t_α zmiennej losowej T spełniająca warunek: $P(|T| > t_\alpha) = 0,20$ odczytujemy z tablicy rozkładu t-Studenta na przecięciu wiersza dla $n = 13$ stopni swobody oraz kolumny dla $\alpha = 0,2$. Otrzymujemy $t_\alpha = 1,350$.

Wartość $P(|T_{13}| > t_\alpha) = 0,20$ zaznaczono na rysunku 3.2.5.5.

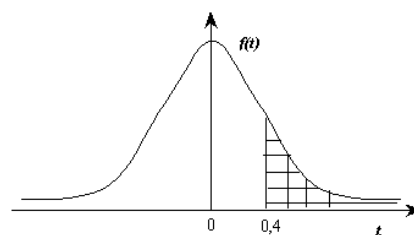
Aby wyznaczyć wartość t_α , dla której $P(T > t_\alpha) = P(T_{13} > t_\alpha) = 0,40$ wykorzystujemy symetryczność funkcji gęstości rozkładu t-Studenta. Wartość t_α odczytujemy z tablicy rozkładu t-Studenta dla $n = 13$ oraz prawdopodobieństwa $P = 0,80$.

Ponieważ $P(|T| > t_\alpha) = 2\alpha$ stąd $t_\alpha = 0,259$.

Wartość $P(T_{13} > t_\alpha) = 0,40$ zaznaczono na rysunku 3.2.5.6.



Rys. 3.2.5.5. Wykres
 $P(|T_{13}| > t_\alpha) = 0,2$



Rys. 3.2.5.6. Wykres
 $P(T_{13} > t_\alpha) = 0,4$

3.3. Nierówność Czebyszewa

Nierówność Czebyszewa pozwala oszacować prawdopodobieństwo zdarzenia polegającego na tym, że wartość zmiennej losowej o nieznanym rozkładzie prawdopodobieństwa różni się od jej wartości średniej o dowolną dodatnią wartość ε , jeżeli znana jest wartość odchylenia standardowego σ . Nierówność Czebyszewa można zapisać następująco.

Jeżeli X jest dowolną zmienną losową o wartości oczekiwanej m oraz odchyleniu standardowym σ dla każdego $\varepsilon > 0$ jest spełniona nierówność:

$$P\{|X - m| \leq \varepsilon\} \geq 1 - \frac{\sigma^2}{\varepsilon^2}$$

lub

$$P\{|X - m| > \varepsilon\} < \frac{\sigma^2}{\varepsilon^2}.$$

Z powyższych nierówności wynika, że im mniejsza jest wartość odchylenia standardowego przy tej samej wartości ε , to mniejsze jest prawdopodobieństwo, że wartość zmiennej losowej X różni się o wartość większą od ε wartości oczekiwanej tej zmiennej.

Przykład 3.3.1

Wysokość podatku płaconego przez podatników jest zmienną losową o odchyleniu standardowym $\sigma = 5$ tysięcy złotych. Jakie jest prawdopodobieństwo tego, że wielkość podatku zapłaconego przez losowo wybranego podatnika różni się od średniej wielkości płaconego podatku co do wartości bezwzględnej o nie więcej niż 10 tysięcy złotych?

Rozwiązanie

$$P\{|X - m| \leq 10\} \geq 1 - \frac{25}{10^2} = 0,75.$$

Przykład 3.3.2

W pewnej miejscowości nadmorskiej w miesiącach wakacyjnych było średnio 20 dni z temperaturą powyżej 25°C. Odchylenie standardowe wynosiło 4°C. Jakie prawdopodobieństwo ma urlopowicz, że wybierając się na urlop w lipcu i sierpniu będzie miał od 15 do 25 dni z temperaturą powyżej 25°C.

Rozwiązanie

Urlopowicz chciałby wiedzieć, jakie jest prawdopodobieństwo, że temperatura będzie się wahała ± 5 stopni wokół wartości średniej temperatury, czyli $\varepsilon = 5$.

Ponieważ $P(15 \leq X \leq 25) = P(|X - 20| \leq 5)$.

Z nierówności Czebyszewa $P\{|X - m| \leq \varepsilon\} \geq 1 - \frac{\sigma^2}{\varepsilon^2}$ wynika, że

$$P\{|X - 20| \leq 5\} \geq 1 - \frac{4^2}{5^2}.$$

Stąd prawdopodobieństwo, że urlopowicz będzie miał od 15 do 25 dni z temperaturą powyżej 25°C wynosi $P\{|X - 20| \leq 5\} \geq 0,36$.

3.4. Zmienna losowa dwuwymiarowa

Zmienną losową dwuwymiarową (X, Y) nazywa się parę funkcji, która każdemu zdarzeniu elementarnemu $e \in E$ przyporządkowuje dwie liczby rzeczywiste $(X(e), Y(e))$. Jeżeli zbiór zdarzeń elementarnych jest skończony lub policzalny, to każda para funkcji (X, Y) jest zmienną losową dwuwymiarową. W przypadku nieskończonego zbioru zdarzeń elementarnych funkcja ta musi spełniać pewne warunki.

Dwuwymiarowa zmienna losowa (X, Y) jest zmienną dyskretną (skokową), jeżeli zmienne X i Y mają skończony lub policzalny zbiór wartości.

Rozkładem łącznym dyskretnej zmiennej losowej dwuwymiarowej (X, Y) nazywa się zbiór prawdopodobieństw:

$$P(X = x_i, Y = y_j) \quad \text{dla } i = 1, 2, \dots, r; j = 1, 2, \dots, s.$$

Prawdopodobieństwa p_{ij} spełniają warunek $\sum_{i=1}^r \sum_{j=1}^s p_{ij} = 1$.

Dystrybuantę $F(x, y)$ zmiennej losowej dwuwymiarowej określa wzór:

$$F(x, y) = \sum_{x_i < x} \sum_{y_j < y} p_{ij}.$$

Rozkład dwuwymiarowej zmiennej losowej zapisywany jest w postaci tablicy korelacyjnej. W główce tablicy występują wartości zmiennej losowej Y , natomiast w boczku wartości zmiennej losowej X .

Tabela 3.4.1. Rozkład zmiennej losowej (X, Y)

Y X	y_1	y_2	y_3	...	y_s	
x_1	p_{11}	p_{12}	p_{13}	...	p_{1s}	$p_{1\cdot}$
x_2	p_{21}	p_{22}	p_{23}	...	p_{2s}	$p_{2\cdot}$
\vdots	\vdots	\vdots	\vdots	\vdots	\vdots	\vdots
x_r	p_{r1}	p_{r2}	p_{r3}	...	p_{rs}	$p_{r\cdot}$
	$p_{\cdot 1}$	$p_{\cdot 2}$	$p_{\cdot 3}$...	$p_{\cdot s}$	

Dwuwymiarowa zmienna losowa (X, Y) jest zmienną ciągłą, jeżeli zmienne losowe X i Y są ciągłe.

Funkcją gęstości $f(x, y)$ dwuwymiarowej zmiennej losowej (X, Y) łącznego rozkładu to dowolna funkcja spełniająca warunki:

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

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

Dystrybuantą zmienności losowej ciągłej (X, Y) nazywamy funkcję określoną wzorem:

$$F(x, y) = \int_{-\infty}^x \int_{-\infty}^y f(x, y) dx dy.$$

3.4.1. Rozkłady brzegowe i rozkłady warunkowe

Rozkładem brzegowym dyskretnej zmiennej losowej X nazywamy prawdopodobieństwo $P(X = x_i)$ określone następująco:

$$P(X = x_i) = \sum_{j=1}^s p_{ij} \quad \text{dla } i = 1, 2, \dots, r.$$

Zamiast $P(X = x_i)$ będzie stosowany zapis p_i .

Rozkładem brzegowym dyskretnej zmiennej losowej Y nazywamy prawdopodobieństwo $P(Y = y_j)$ określone następująco:

$$P(Y = y_j) = \sum_{i=1}^r p_{ij} \quad \text{dla } j = 1, 2, \dots, s.$$

Zamiast $P(Y = y_j)$ będzie stosowany zapis p_j .

Rozkład brzegowy zmiennej X przedstawiony jest w ostatniej kolumnie tabeli 3.4.1, a rozkład brzegowy zmiennej Y w ostatnim wierszu tej tabeli.

Rozkładami brzegowymi $f_1(x)$, $f_2(y)$ ciągłych zmiennych losowych X i Y nazywa się następujące funkcje:

$$f_1(x) = \int_{-\infty}^{\infty} f(x, y) dy,$$

$$f_2(y) = \int_{-\infty}^{\infty} f(x, y) dx.$$

Rozkładem warunkowym zmiennej losowej skokowej X pod warunkiem $Y = y_j$ nazywa się prawdopodobieństwa:

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

dla $i = 1, 2, \dots, r$, $j = 1, 2, \dots, s$ oraz $\sum_{i=1}^r p_{i/j} = 1$.

Jeżeli zmienna losowa Y przyjmuje s wartości, można otrzymać s rozkładów warunkowych zmiennej losowej X .

Rozkładem warunkowym zmiennej losowej skokowej Y pod warunkiem $X = x_i$ nazywa się prawdopodobieństwa:

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

$$\sum_{j=1}^2 p_{j/i} = 1.$$

Jeżeli zmienna losowa X przyjmuje r wartości, można otrzymać r rozkładów warunkowych zmiennej losowej Y .

Do opisu rozkładów zmiennych losowych ciągłych wykorzystuje się warunkowe funkcje gęstości.

Warunkowa funkcja gęstości zmiennej losowej X jest określona wzorem:

$$f(x/y) = \frac{f(x,y)}{f_2(y)},$$

a warunkowa funkcja gęstości zmiennej losowej Y jest określona wzorem:

$$f(y/x) = \frac{f(x,y)}{f_1(x)}.$$

3.4.2. Niezależność zmiennych losowych X i Y

Niech (X, Y) będzie dwuwymiarową zmienną losową skokową. Jeżeli dla każdej pary wartości (x_i, y_j) tej zmiennej jest spełniony warunek:

$$P(X = x_i, Y = y_j) = P(X = x_i) \cdot P(Y = y_j),$$

to zmienne X i Y są stochastycznie niezależne.

Warunek ten można zapisać krótko: $p_{ij} = p_i \cdot p_j$ dla każdej pary wskaźników $i = 1, 2, \dots, r, j = 1, 2, \dots, s$.

Jeżeli zmienne losowe X i Y są niezależne, wówczas:

$$P(X = x_i / Y = y_j) = \frac{P(X = x_i, Y = y_j)}{P(Y = y_j)} = \frac{P(X = x_i) \cdot P(Y = y_j)}{P(Y = y_j)} = P(X = x_i) = p_i. \quad \text{dla } i = 1, 2, \dots, r$$

Tak samo będzie dla pozostałych rozkładów warunkowych. Widać więc, że jeżeli zmienne losowe X i Y są niezależne, to wszystkie rozkłady warunkowe zmiennej X są takie same, ponadto są one identyczne z rozkładem brzegowym zmiennej losowej X . To samo można powiedzieć o rozkładach warunkowych zmiennej losowej Y , gdyż:

$$P(Y = y_j / X = x_i) = \frac{P(X = x_i, Y = y_j)}{P(X = x_i)} =$$
$$\frac{P(X = x_i) \cdot P(Y = y_j)}{P(X = x_i)} = P(Y = y_j) = p_{\cdot j} \quad \text{dla } j = 1, 2, \dots, s$$

Jeżeli (X, Y) jest zmienną losową ciągłą, to warunek niezależności zmiennych losowych X i Y może być zapisany następująco:

$$f(x, y) = f_1(x) \cdot f_2(y)$$

dla każdej pary liczb rzeczywistych x i y .

Korzystając ze wzoru na rozkład warunkowy zmiennej losowej X oraz przyjmując, że zmienne losowe X i Y są niezależne, otrzymujemy:

$$f(x/y) = \frac{f(x, y)}{f_2(y)} = \frac{f_1(x) \cdot f_2(y)}{f_2(y)} = f_1(x).$$

Oznacza to, że dla niezależnych zmiennych losowych X i Y rozkład warunkowy zmiennej losowej X jest identyczny z rozkładem brzegowym zmiennej losowej X .

Podobnie rozkład warunkowy zmiennej losowej Y jest identyczny z rozkładem brzegowym tej zmiennej, gdyż:

$$f(y/x) = \frac{f(x, y)}{f_1(x)} = \frac{f_1(x) \cdot f_2(y)}{f_1(x)} = f_2(y).$$

3.5. Powtórzenie w przykładach

Zadanie 1

Pracownia studencka wyposażona jest w pięć komputerów. Na podstawie dłuższych obserwacji stopnia ich wykorzystania określono prawdopodobieństwo $p = 0,1$ tego, że podczas zajęć jeden z komputerów jest wolny. Obliczyć prawdopodobieństwo, że w danej chwili wolne są:

- 2 komputery,
- przynajmniej dwa komputery.

Rozwiązanie

Zmienna losowa X ma rozkład Bernoulliego.

$$P(X = k) = \binom{n}{k} \cdot p^k \cdot q^{n-k} = \frac{n!}{k!(n-k)!} \cdot p^k \cdot q^{n-k} =$$
$$= \frac{5!}{2!(5-2)!} \cdot 0,1^2 \cdot 0,9^{5-2} = 0,0729.$$

$$\begin{aligned} P(X \geq 2) &= P(X = 2) + P(X = 3) + P(X = 4) + P(X = 5) = \\ &= 0,0729 + 0,0081 + 0,00045 + 0,00001 = 0,08146. \end{aligned}$$

Zadanie 2

Z bieżącej produkcji pobrano losowo 5 sztuk towaru. Niech zmienna losowa X oznacza liczbę pobranych sztuk wadliwych. Znaleźć rozkład zmiennej losowej X , jeśli wiadomo, że prawdopodobieństwo wystąpienia sztuki wadliwej wynosi 0,1.

Rozwiązanie

Zmienna losowa ma rozkład Bernoulliego $n = 5$; $k = 0, 1, 2, \dots, 5$; $p = 0,1$, $q = 0,9$.

$$P(X = k) = \binom{n}{k} \cdot p^k \cdot q^{n-k},$$

$$P(X = 0) = \binom{5}{0} \cdot 0,1^0 \cdot 0,9^5 = 1 \cdot 1 \cdot 0,9^5 = 0,59049,$$

$$P(X = 1) = \binom{5}{1} \cdot 0,1^1 \cdot 0,9^4 = 5 \cdot 0,1 \cdot 0,9^4 = 0,32805,$$

$$P(X = 2) = \binom{5}{2} \cdot 0,1^2 \cdot 0,9^3 = 10 \cdot 0,01 \cdot 0,9^3 = 0,07290,$$

$$P(X = 3) = \binom{5}{3} \cdot 0,1^3 \cdot 0,9^2 = 10 \cdot 0,001 \cdot 0,9^2 = 0,00810,$$

$$P(X = 4) = \binom{5}{4} \cdot 0,1^4 \cdot 0,9^1 = 5 \cdot 0,0001 \cdot 0,9^1 = 0,00045,$$

$$P(X = 5) = \binom{5}{5} \cdot 0,1^5 \cdot 0,9^0 = 1 \cdot 0,00001 \cdot 0,9^0 = 0,00001.$$

Zadanie 3

Prawdopodobieństwo wyprodukowania wadliwego towaru wynosi 0,03. Obliczyć prawdopodobieństwo, że wśród 200 sztuk znajdzie się jedna sztuka wadliwa.

Rozwiązanie

Oznaczamy przez X zmienną losową, która wyraża liczbę wadliwego towaru. Intensywność wyprodukowania wadliwego towaru wynosi:

$$\lambda = n \cdot p = 200 \cdot 0,03 = 6.$$

Prawdopodobieństwo, że będzie jedna sztuka wadliwa wynosi:

$$P(X = 1) = \frac{\lambda^k}{k!} \cdot e^{-\lambda} = \frac{6^1}{1!} \cdot e^{-6} = 0,0154.$$

Zadanie 4

Zmienna losowa X ma normalny rozkład prawdopodobieństwa o parametrach $m = 24$, $\sigma = 12$. Obliczyć prawdopodobieństwo, że zmienna ta przyjmie wartości pomiędzy 17,4 a 58,8.

Rozwiązanie

Obliczamy wartości zmiennej losowej standaryzowanej U dla zadanych wartości zmiennej losowej X :

$$u_1 = \frac{x_1 - m}{\sigma} = \frac{17,4 - 24}{12} = -0,55,$$
$$u_2 = \frac{x_2 - m}{\sigma} = \frac{58,8 - 24}{12} = 2,9.$$

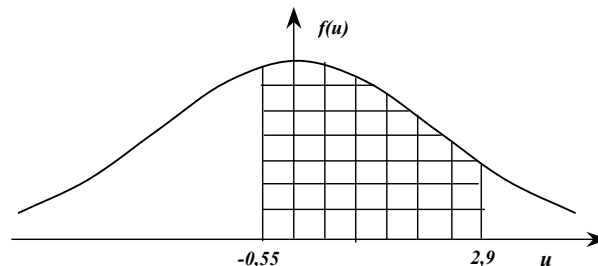
Korzystając z tablic dystrybucyj rozkładu normalnego wyznaczamy wartości skumulowanego prawdopodobieństwa zmiennej losowej U dla $u_1 = -0,55$ i $u_2 = 2,90$.

$$\Phi(u_1 = -0,55) = 0,2088,$$

$$\Phi(u_2 = 2,90) = 0,4981.$$

Wartość szukanego prawdopodobieństwa wynosi:

$$P(17,4 < X < 58,8) = P(-0,55 < U < 2,9) =$$
$$= P(U > -0,55) - P(U > 2,9) = 0,5 + \Phi(0,55) - (0,5 - \Phi(2,9)) =$$
$$= 0,2088 + 0,4981 = 0,7096.$$



Rys. 3.5.1. Obszar pod krzywą rozkładu normalnego dla $(17,4 < X < 58,8)$

Zadanie 5

Waga ciała populacji studentów pewnej uczelni ma rozkład normalny $N(74, 14)$. Obliczyć prawdopodobieństwo, że przypadkowo napotkana studentka waży od 60 do 67 kilogramów.

Rozwiązanie

Obliczamy wartości zmiennej losowej standaryzowanej U dla zadanych granic przedziału wagi studentki:

$$u_1 = \frac{x_1 - m}{\sigma} = \frac{60 - 74}{14} = -1,0,$$

$$u_2 = \frac{x_2 - m}{\sigma} = \frac{67 - 74}{14} = -0,5.$$

Korzystając z tablic dystrybucyj rozkładu normalnego wyznaczamy wartości skumulowanego prawdopodobieństwa zmiennej losowej U dla $u_1 = -1,0$ i $u_2 = -0,5$.

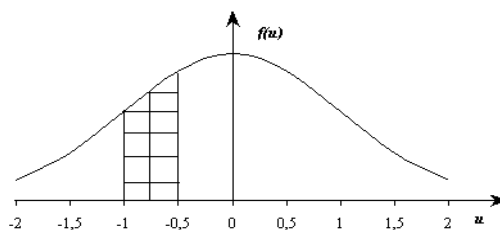
$$\Phi(u_1 = -1,0) = 0,2420,$$

$$\Phi(u_2 = -0,5) = 0,3085.$$

Wartość szukanego prawdopodobieństwa wynosi:

$$P(60 < X < 67) = P(-1,0 < U < -0,5) =$$

$$P(U > -1,0) - P(U > -0,5) = 0,5 + \Phi(1,0) - (0,5 + \Phi(0,5)) = 0,3413 - 0,1515 = 0,1898.$$



Rys. 3.5.2. Obszar pod krzywą rozkładu normalnego dla $P(60 < X < 67)$

Zadanie 6

Zmienna losowa X ma rozkład t-Studenta o 25 stopniach swobody. Wyznaczyć następujące prawdopodobieństwa:

a) $P(T > 1,71)$,

b) $P(|T| \leq 1,32)$.

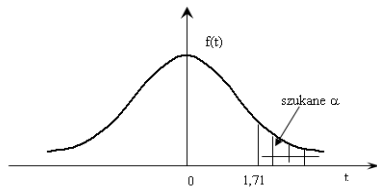
Rozwiązanie

a) z tablic rozkładu t-Studenta wyznaczamy takie α , aby spełniona była zależność $P(|T_n| \geq t_\alpha) = \alpha$ dla $n = 25$.

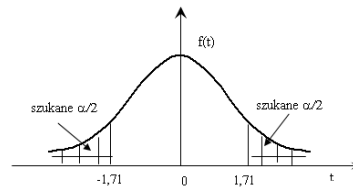
Szukamy w tablicy takiego prawdopodobieństwa, aby $P(|T_{25}| > 1,71) = 0,1$ (rys 3.5.3).

$$\text{Stąd } P(T_{25} > 1,71) = \frac{P(|T_{25}| > 1,71)}{2} = \frac{0,1}{2} = 0,05 \text{ (rys 3.5.4).}$$

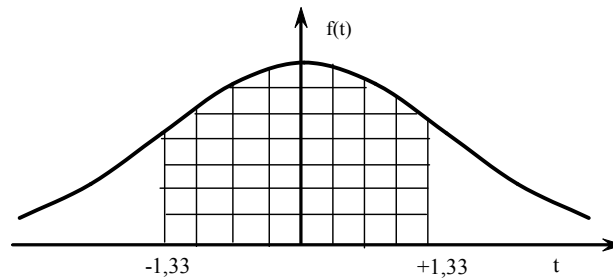
b) $P(|T| \leq 1,32) = 1 - P(|T| > 1,32) = 1 - 0,2 = 0,8$ (rys. 3.5.5).



Rys. 3.5.3. $P(|T_{25}| > 1,71)$



Rys 3.5.4. $P(|T_{25}| > 1,71)$



Rys. 3.5.5. $P(|T_{25}| \leq 1,32)$

Zadanie 7

Zmienna losowa X ma rozkład χ^2 o 20 stopniach swobody (χ^2_{20}). Wyznaczyć następujące prawdopodobieństwa:

- a) $P(10,85 < X < 28,40)$,
- b) $P(X < 22,76)$,
- c) $P(X > 10,85)$.

Rozwiązanie

- a) $P(10,85 < X < 28,40) = F(28,40) - F(10,85) = 0,9 - 0,05 = 0,85$.
- b) $P(X < 22,76) = F(22,76) = 0,70$.
- c) $P(X > 10,85) = 1 - F(10,85) = 0,95$.

Zadanie 8

W miejscowości Nowiny 50 rolników obsiało swoje pola zbożem. W lipcu 2020 roku, tuż przed żniwami spadł grad, który objął swym zasięgiem duży obszar. Likwidator szkód oszacował, że w sąsiedniej miejscowości wskutek

tego gradobicia średni procent szkody w uprawach zbożowych wynosi 15, zaś odchylenie standardowe 5. Obliczyć prawdopodobieństwo, że procent szkody w Nowinach wyniesie:

- a) od 10 do 20 b) mniej niż 13 c) przekroczy 20

Rozwiązanie

$$\begin{aligned} \text{a) a) } P(10 < X < 20) &= P\left(\frac{10-15}{5} < U < \frac{20-15}{5}\right) = P(-1 < U < 1) = \\ &= \Phi(1) - \Phi(-1) = \Phi(1) - [1 - \Phi(1)] = 2\Phi(1) - 1 = 0,6826; \end{aligned}$$

$$\text{b) } P(X < 13) = 1 - P\left(U < \frac{13-15}{5}\right) = P(U < -0,4) = 1 - \Phi(0,4) = 0,3446;$$

$$\begin{aligned} \text{c) } P(X > 20) &= 1 - P(X \leq 20) = 1 - P\left(U < \frac{20-15}{5}\right) = 1 - P(U < 1) = 1 - \Phi(1) = \\ &= 0,1587. \end{aligned}$$

ROZDZIAŁ IV TWIERDZENIA GRANICZNE

Twierdzenia graniczne dotyczą problemów ukazujących własności graniczne ciągów zmiennych losowych. Dzieli się na twierdzenia graniczne lokalne i twierdzenia graniczne integralne.

Twierdzeniami lokalnymi nazywamy te twierdzenia, które omawiają zbieżność ciągu funkcji prawdopodobieństwa w przypadku ciągu $\{X_n\}$ zmiennych losowych typu skokowego lub zbieżność ciągu gęstości prawdopodobieństwa w przypadku ciągu $\{X_n\}$ zmiennych losowych typu ciągłego.

Twierdzeniami integralnymi nazywamy te twierdzenia, które dotyczą zbieżności ciągu dystrybuant zmiennych losowych $\{X_n\}$ do pewnej dystrybuanty granicznej. Twierdzenia integralne, rozważające ciągi sum niezależnych zmiennych losowych o dystrybuancie granicznej, będącej dystrybuantą rozkładu normalnego nazywamy centralnymi.

4.1. Twierdzenia lokalne

Z twierdzeniem lokalnym mamy do czynienia przy omawianiu rozkładu Poissona. Niekiedy bywa ono nazywane lokalnym twierdzeniem Poissona.

Lokalne twierdzenie Poissona

Jeżeli $\{X_n\}$ jest ciągiem zmiennych losowych o rozkładzie dwumianowym (Bernoulliego) o parametrach n i p , czyli:

$$P(X_n = k) = \binom{n}{k} p^k q^{n-k}, \quad 0 < p < 1, \quad k = 0, 1, 2, \dots, n,$$

przy czym parametr p jest funkcją n postaci $p = \frac{\lambda}{n}$, gdzie $\lambda > 0$ jest ustaloną liczbą $n = 1, 2, \dots$, to

$$\lim_{n \rightarrow \infty} P(X_n = k) = P(X = k) = \frac{\lambda^k}{k!} e^{-\lambda} \quad \text{dla } k = 0, 1, 2, \dots, n.$$

Zatem, zasadniczą część twierdzenia Poissona możemy interpretować w ten sposób, że funkcja prawdopodobieństwa rozkładu dwumianowego jest zbieżna do funkcji prawdopodobieństwa rozkładu Poissona dla $n \rightarrow \infty$ (dużych n), jeśli tylko spełniony jest warunek $p = \frac{\lambda}{n}$. W miarę zwiększania się n parametr p staje się coraz mniejszy, a więc iloczyn $n \cdot p$ jest stały i równy liczbie λ z góry ustalonej. Równość $\lim_{n \rightarrow \infty} P(X_n = k) = P(X = k)$ pozwala sto-

sować wzór przybliżony $P(X_n = k) \approx P(X = k)$, jeśli tylko n jest dostatecznie duże i λ dostatecznie małe, co w praktyce oznacza $n \geq 100$ i $\lambda \leq 20$. Jeśli warunek na λ nie jest spełniony nie można korzystać ze wzoru Poissona. Stosujemy wówczas inny wzór przybliżony

$$P(X_n = k) = P_{n,k} \approx \frac{1}{\sqrt{2\pi\sigma}} e^{-\frac{1}{2}\left(\frac{k-m}{\sigma}\right)^2},$$

gdzie: $m = n \cdot p$, $\sigma = \sqrt{npq}$, $q = 1 - p$.

Błąd, który popełniamy stosując ten przybliżony wzór jest wystarczająco mały, jeśli wielkości n , k oraz $n - k$ są dostatecznie duże. Korzystanie z powyższego wzoru uzasadnione jest twierdzeniem Moivre'a-Laplace'a.

Przykład 4.1.1

Rzucamy 144 razy kostką do gry. Obliczyć prawdopodobieństwo, że 25 razy otrzymamy piątkę.

Rozwiązanie

Mamy: $p = \frac{1}{6}$, $q = \frac{5}{6}$, $k = 25$, $n = 144$, stąd $m = n \cdot p = 144 \cdot \frac{1}{6} = 24$,

$$\sigma = \sqrt{npq} = \sigma = \sqrt{144 \cdot \frac{1}{6} \cdot \frac{5}{6}} = \sqrt{20}.$$

$$P_{144,25} \approx \frac{1}{\sqrt{2\pi \cdot 20}} e^{-\frac{1}{2}\left(\frac{25-24}{\sqrt{20}}\right)^2} = \frac{1}{\sqrt{40\pi}} e^{-\frac{1}{40}} \approx 0,087.$$

Lokalne twierdzenie Moivre'a-Laplace'a

Jeżeli $\{Y_n\}$ jest ciągiem standaryzowanych zmiennych losowych $\{X_n\}$ o rozkładzie dwumianowym, to

$$\lim_{n \rightarrow \infty} \sqrt{npq} \cdot P(Y_n = y_n) = \frac{1}{\sqrt{2\pi}} e^{-\frac{1}{2}y^2},$$

gdzie: $y_n = \frac{k - np}{\sqrt{npq}}$ jest wartością, jaką przyjmuje zmienna losowa $\{Y_n\}$,

gdy zmienna losowa $\{X_n\}$ przyjmuje wartość k taką, że gdy $n \rightarrow \infty$ to również $k \rightarrow \infty$ w taki sposób, że $(n - k) \rightarrow \infty$ oraz $\lim_{n \rightarrow \infty} Y_n = Y$.

Wniosek z lokalnego twierdzenia Moivre'a-Laplace'a

Dla dużych n prawdopodobieństwa rozkładu dwumianowego mogą być obliczone za pomocą funkcji gęstości rozkładu normalnego:

$$\binom{n}{k} p^k q^{n-k} = \frac{1}{\sqrt{npq}} \cdot f\left(\frac{k - np}{\sqrt{npq}}\right);$$

gdzie f oznacza tu gęstość rozkładu $N(0,1)$.

Przykład 4.1.2

Prawdopodobieństwo, że w ciągu pewnego czasu T przepali się żarówka wynosi $\frac{1}{4}$. Obliczyć prawdopodobieństwo, że w ciągu tego czasu przepalą się 42 żarówki spośród 192.

Rozwiązanie

Niech zmienna losowa X oznacza liczbę przepalonych żarówek spośród 192. Należy obliczyć $P(X = 42)$. Zmienna losowa ma rozkład dwumianowy, a zatem

$$P(X = 42) = \binom{192}{42} \left(\frac{1}{4}\right)^{42} \left(\frac{3}{4}\right)^{150}.$$

Iloczyn $n \cdot p = 192 \cdot \frac{1}{4} = 48$ jest duży. Do obliczenia szukanego prawdopodobieństwa zastosujemy wniosek z lokalnego twierdzenia Moivre'a-Laplace'a.

$$\begin{aligned} P(X = 42) &= \binom{192}{42} \left(\frac{1}{4}\right)^{42} \left(\frac{3}{4}\right)^{150} = \frac{1}{6} f\left(\frac{42 - 48}{6}\right) = \\ &= \frac{1}{6} f(-1) = \frac{1}{6} f(1) = \frac{1}{6} \cdot 0,242 = 0,0403. \end{aligned}$$

4.2. Twierdzenia integralne

Centralne twierdzenie Lindeberga-Levy'ego

Niech $\{X_i\}$ będzie ciągiem zmiennych losowych, o których zakładamy, że:

- a) są niezależne,
- b) mają identyczne rozkłady,
- c) mają skończone wartości przeciętne $E(X_i) = m < \infty$,
- d) mają skończone i dodatnie wariancje $0 < D^2(X_i) < \sigma^2 < \infty$.

Wprowadzimy ciąg zmiennych $\{Y_n\}$, gdzie:

$$Y_n = \sum_{i=1}^n X_i.$$

Niech $\{U_n\}$ będzie ciągiem standaryzowanych zmiennych losowych Y_n , tzn.

$$U_n = \frac{Y_n - E(Y_n)}{D(Y_n)}.$$

Wówczas spełniona jest równość:

$$\lim_{n \rightarrow \infty} F_n(u) = \Phi(u),$$

gdzie u jest dowolną liczbą rzeczywistą, $\Phi(u)$ dystrybuantą standaryzowanego rozkładu normalnego, zaś $F_n(u)$ jest ciągiem dystrybuant odpowiadających zmiennym losowym U_n .

Wniosek z centralnego twierdzenia Lindeberga-Levy'ego

Dla dużych n zmienna losowa $Y_n = X_1 + X_2 + \dots + X_n$, gdzie X_1, \dots, X_n są niezależnymi zmiennymi losowymi o takim samym rozkładzie z wartością oczekiwaną m i wariancją σ^2 ma w przybliżeniu rozkład normalny $N(nm, \sigma\sqrt{n})$, stąd

$$P\left(\frac{Y_n - nm}{\sigma\sqrt{n}} < a\right) = \Phi(a),$$

$$P\left(a \leq \frac{Y_n - nm}{\sigma\sqrt{n}} \leq b\right) = \Phi(b) - \Phi(a),$$

$$P\left(\frac{Y_n - nm}{\sigma\sqrt{n}} \geq b\right) = 1 - \Phi(b).$$

Przykład 4.2.1

Wiadomo, że średnia waga dorosłego człowieka wynosi 75 kg, a odchylenie standardowe 3 kg. Samolot pasażerski zabiera 81 osób. Obliczyć prawdopodobieństwo, że łączna waga pasażerów przekroczy 6 ton.

Rozwiązanie

Niech X_i oznacza zmienną losową, którą jest waga i -tego człowieka. Centralne twierdzenie graniczne mówi, że jeśli n jest duże, to rozkład zmiennej losowej można przybliżyć rozkładem normalnym z wartością oczekiwaną $n \cdot m = 81 \cdot 75 \text{ kg} = 6075 \text{ kg}$ oraz odchyleniem standardowym $\sigma\sqrt{n} = 3\sqrt{81} \text{ kg}$.

Mamy rozkład $N(6075 \text{ kg}; 3\sqrt{81} \text{ kg})$, a zatem

$$\begin{aligned} P(U_{81} > 6000) &= P\left(U > \frac{6000 - 6075}{27}\right) = P\left(U > \frac{-75}{27}\right) = \\ &= P(U > -2,8) = \Phi(-2,8) = 0,9974. \end{aligned}$$

Integralne twierdzenie Moivre'a-Laplace'a

Jeżeli $F_n(u)$ jest ciągiem dystrybucji zmiennych losowych

$$U_n = \frac{Y_n - np}{\sqrt{npq}},$$

przy czym Y_n mają rozkład dwumianowy, to dla każdej wartości rzeczywistej n zachodzi wzór:

$$\lim_{n \rightarrow \infty} F_n(u) = \Phi(u).$$

Wniosek z integralnego twierdzenia Moivre'a-Laplace'a

Dla dużych n zmienna losowa X o rozkładzie dwumianowym z parametrami n i p ma rozkład w przybliżeniu normalny $N(nm, \sqrt{npq})$, czyli inaczej asymptotycznie normalny, stąd

$$P\left(\frac{Y_n - nm}{\sqrt{npq}} < a\right) = \Phi(a),$$

$$P\left(a \leq \frac{Y_n - nm}{\sqrt{npq}} \leq b\right) = \Phi(b) - \Phi(a),$$

$$P\left(\frac{Y_n - nm}{\sqrt{npq}} \geq b\right) = 1 - \Phi(b).$$

Przykład 4.2.2

Prawdopodobieństwo uzyskania połączenia telefonicznego w Radomiu w godzinach przedpołudniowych wynosi $p = 0,8$ (dane umowne). Zakładamy, że połączenia są realizowane niezależnie. Obliczyć prawdopodobieństwo, że spośród 1000 abonentów połączenie uzyskało:

- ponad 800 abonentów,
- mniej niż 900 abonentów,
- liczba abonentów, którzy uzyskali połączenie zawiera się pomiędzy 700 i 800.

Rozwiązanie

Korzystamy z wniosku z integralnego twierdzenia Moivre'a-Laplace'a. Z warunków zadania mamy: $n \cdot p = 800$, $\sqrt{npq} = 12,6$. Oznaczamy przez X liczbę abonentów, którzy uzyskali połączenie; $X \sim N(800; 12,6)$ i obliczamy żądane prawdopodobieństwa.

a) $P(X > 800) = P\left(U > \frac{800 - 800}{12,6}\right) = P(U > 0) = 0,5,$

$$\text{b) } P(X < 900) = P\left(U < \frac{900 - 800}{12,6}\right) = P(U < 7,94) = 1,$$

$$\begin{aligned} \text{c) } P(700 < X < 800) &= P\left(\frac{700 - 800}{12,6} < U < \frac{800 - 800}{12,6}\right) \\ &= P(-7,94 < U < 0) = 0,5. \end{aligned}$$

4.3. Prawa wielkich liczb

Prawo wielkich liczb można traktować jako szczególny przypadek twierdzeń integralnych (ale nie są to twierdzenia centralne). Prawa wielkich liczb są to twierdzenia graniczne, na podstawie których badamy zbieżność ciągów zmiennych losowych w sensie zbieżności według prawdopodobieństwa (słabe prawa) lub w sensie zbieżności z prawdopodobieństwem 1 (mocne prawa). Ograniczymy się do rozważania słabych praw wielkich liczb.

Niech dany będzie ciąg zmiennych losowych $\{Y_n\}$ określonych na zbiorze zdarzeń elementarnych Ω i niech każda ze zmiennych losowych $\{Y_n\}$ ma wartość oczekiwaną $E(Y_n) = m$ dla $n \in \mathbb{N}$. Powiadamy, że ciąg $\{Y_n\}$ jest zbieżny według prawdopodobieństwa do wartości oczekiwanej m , jeżeli dla dowolnej liczby dodatniej ε

$$\lim_{n \rightarrow \infty} P(|Y_n - m| < \varepsilon) = 1.$$

Dla ciągu $\{Y_n\}$ zachodzi prawo wielkich liczb. Oznacza to, że gdy n jest duże, to prawdopodobieństwo, iż zmienna losowa $\{Y_n\}$ przyjmie wartość z dowolnie małego (lecz ustalonego) otoczenia wartości oczekiwanej jest bliskie 1. Wynika stąd, że $\{Y_n\}$ ma rozkład silnie skupiony wokół wartości oczekiwanej m .

Prawo wielkich liczb Bernoulliego

Jeżeli $\{X_n\}$ jest ciągiem zmiennych losowych o rozkładach dwumianowych

$$P(X_n = k) = \binom{n}{k} p^k q^{n-k},$$

to ciąg $\left\{\frac{X_n}{n}\right\}$ jest zbieżny według prawdopodobieństwa do p , tzn. dla każdej dodatniej liczby ε

$$\lim_{n \rightarrow \infty} P\left(\left|\frac{X_n}{n} - p\right| < \varepsilon\right) = 1.$$

Ponieważ $m = E\left(\frac{X_n}{n}\right) = p$, więc powyższe twierdzenie jest prawem wielkich liczb.

W celu zrozumienia „o co chodzi” wprowadzimy następujące oznaczenia:

X_n – liczba sukcesów w n doświadczeniach Bernoulliego,

$Y_n = \frac{X_n}{n}$ – częstość sukcesu,

p – prawdopodobieństwo sukcesu w jednym doświadczeniu.

Jeśli liczba doświadczeń Bernoulliego jest duża, to z prawdopodobieństwem bliskim jedności częstość sukcesu przyjmuje wartości mało różniące się od prawdopodobieństwa sukcesu.

Z prawa wielkich liczb Bernoulliego wynika, że prawdopodobieństwo zdarzenia może być oceniane przez częstość tego zdarzenia w długim ciągu powtórzeń doświadczenia, w którym to doświadczenie występuje. A zatem, uprawniona jest interpretacja prawdopodobieństwa zdarzenia za pomocą częstości tego zdarzenia. Można wykazać, że dla dowolnego $\varepsilon > 0$ i dla dużych n zachodzi zależność:

$$P\left(\left|\frac{X_n}{n} - p\right| < \varepsilon\right) = 2\Phi\left(\varepsilon\sqrt{\frac{n}{pq}}\right) - 1 \quad (*)$$

Wynika to z twierdzenia integralnego Moivre’a-Laplace’a.

Przykład 4.3.1

Wadliwość partii towaru wynosi 0,2. Z partii tego towaru pobrano losowo ze zwracaniem próbę liczącą 400 sztuk. Obliczyć prawdopodobieństwo, że wadliwość w tej próbie będzie odchyłać się od wadliwości partii towaru o mniej niż 0,05.

Rozwiązanie

Niech X_{400} oznacza liczbę sztuk wadliwych w próbie liczącej 400 sztuk, a więc wadliwością tej próby jest $\frac{X_{400}}{400}$. Prawdopodobieństwo wadliwości tej partii wynosi $p = 0,2$. Należy zatem obliczyć:

$$P\left(\left|\frac{X_{400}}{400} - 0,2\right| < 0,05\right).$$

Na podstawie wzoru (*) podanego powyżej otrzymujemy:

$$P\left(\left|\frac{X_{400}}{400} - 0,2\right| < 0,05\right) = 2\Phi\left(0,05\sqrt{\frac{400}{0,2 \cdot 0,8}}\right) - 1 = 0,9876.$$

Prawo wielkich liczb Chinczyna

Niech dany będzie ciąg zmiennych losowych $\{X_n\}$ niezależnych, o jednakowym rozkładzie, o wartości oczekiwanej $m = E(X_n)$. Niech ponadto $\{\bar{X}_n\}$ będzie ciągiem średnich arytmetycznych

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

wówczas ciąg $\{\bar{X}_n\}$ będzie zbieżny według prawdopodobieństwa do m , czyli dla ciągu $\{\bar{X}_n\}$ zachodzi prawo wielkich liczb, bowiem

$$E(\bar{X}_n) = E\left(\frac{1}{n} \sum_{i=1}^n X_i\right) = \frac{1}{n} E\left(\sum_{i=1}^n X_i\right) = \frac{1}{n} \sum_{i=1}^n E(X_i) = \frac{1}{n} \cdot n \cdot m = m.$$

Średnia arytmetyczna dużej liczby zmiennych losowych niezależnych o jednakowym rozkładzie o wartości oczekiwanej m przyjmuje z prawdopodobieństwem bliskim jedności wartości mało różniące się od m . Ta interpretacja ma liczne zastosowania w teorii błędów przypadkowych.

W prawie wielkich liczb Chinczyna nie zakłada się w odróżnieniu od twierdzenia Lindeberga-Levy'ego, że zmienne losowe X_k mają wariancję. Gdyby jednak założyć, że zmienne te mają wariancję $\sigma^2 > 0$, to z twierdzenia Lindeberga-Levy'ego można wyprowadzić (dla dużych n) następującą zależność

$$P\left(|\bar{X}_n - m| < \varepsilon\right) = 2\Phi\left(\frac{\varepsilon\sqrt{n}}{\sigma}\right) - 1 \quad \text{dla } \varepsilon > 0 \quad (**)$$

Przykład 4.3.2

Dokonano 100 pomiarów pewnego detalu. Obliczyć prawdopodobieństwo, że średnia arytmetyczna tych pomiarów będzie odchyłać się od wielkości mierzonej o 0,05 cm, jeśli wiadomo, że odchylenie standardowe poszczególnego pomiaru wynosi 0,5 cm.

Rozwiązanie

Niech X_{100} oznacza średnią ze 100 pomiarów. Na podstawie wzoru (**) i danych z zadania obliczamy

$$P\left(|\bar{X}_{100} - m| < 0,05\right) = 2\Phi\left(\frac{0,05 \cdot \sqrt{100}}{0,5}\right) - 1 = 2\Phi(1) - 1 = 2 \cdot 0,84134 - 1 \approx 0,68.$$

ROZDZIAŁ V STATYSTYKI I ICH ROZKŁADY

5.1. Sposoby doboru próby

Powszechnie stosowanym badaniem częściowym jest tzw. badanie reprezentacyjne. Przy badaniu tym konieczne jest, aby część podlegająca badaniu, czyli próba została wyodrębniona z populacji generalnej poprzez właściwe losowanie. Zbiorowość próbna stanowi częśćkę pobraną z populacji generalnej w ściśle określony sposób. W praktyce doboru elementów do próby możemy posługiwać się różnymi sposobami losowania. Losowe pobieranie próby z całej populacji jest gwarancją, że unikniemy błędu obciążenia.

Przykład 1

Założmy, że poszukujemy oszacowania frakcji konsumentów, którzy byliby zainteresowani zakupem nowego produktu (np. gatunek wina), wchodzącego na polski rynek. Jak zebrać informacje dotyczące preferencji zakupowych do tych badań?

Rozwiązanie

Populacja nie została jasno określona, musimy ją zatem zdefiniować zgodnie z celem badania. Jakich konsumentów mamy na myśli: czy tylko dorosłych, czy z całymi rodzinami, czy tylko tych, którzy od czasu do czasu piją wino? Czy chodzi nam o konsumentów w całej Polsce? Musimy odpowiedzieć na te pytania, zanim zaczniemy zbierać dane. Interesujemy się wprowadzeniem nowego gatunku wina na rynek, czyli określeniem frakcji rynku napojów alkoholowych przypadających na ten gatunek wina po wprowadzeniu go na rynek. Jako interesującą nas populację przyjmiemy wszystkich tych ludzi, którzy przynajmniej od czasu do czasu są konsumentami napojów alkoholowych. Aby wyznaczyć próbę losową z całej populacji ludzi spożywających przynajmniej od czasu do czasu napoje alkoholowe musimy mieć *operat losowania*. Oznacza to, że potrzebujemy listy wszystkich tych ludzi, spośród których będziemy mogli losowo wybrać tyle osób do próby, ile potrzebujemy. W praktyce taka lista jest zazwyczaj niedostępna, a więc trzeba zdobyć próbę w inny sposób. Najczęściej stosowaną metodą jest wysłanie ankietów do miejsc, w których można znaleźć interesujących nas konsumentów. Są to zwykle różne centra handlowe, gdzie kupujący są wybierani losowo i za pomocą wstępnych pytań sprawdzani, czy należą do interesującej nas populacji, a przede wszystkim, czy przynajmniej od czasu do czasu spożywają napoje alkoholowe. Podajemy następnie tak wybranym osobom do spró-

bowania badany produkt (taste test) i prosimy o wypełnienie ankiety. Ta metoda doboru próby spełni swe zadanie pod warunkiem, że ankieterzy nie wybiorą do próby w sposób nielosowy np. tylko młodych osób, czy tylko elegancko ubranych. Musimy zastanowić się także, czy wybrani w danych centrach ludzie stanowią reprezentatywną frakcję potencjalnych konsumentów, bo jeśli nie, to pojawi się błąd obciążenia. Nie możemy faworyzować żadnej grupy społecznej ani też lokalizacji badanych centrów. Najlepiej badanie przeprowadzić w różnych miejscach.

Losowanie niezależne (z powtórzeniami) polega na zachowaniu kolejnych jednostek próby. Po wylosowaniu jednostki populacji i zanotowaniu wartości badanej cechy tej jednostki zwracamy ją do populacji. Ten sam element populacji może być wylosowany wielokrotnie, a zatem uzyskujemy próbę z powtórzeniami. W ten sposób prawdopodobieństwo wylosowania każdego elementu w każdym kolejnym losowaniu jest niezmiennie.

Losowanie zależne (bez powtórzeń) polega na tym, że każdy element populacji może być wylosowany tylko jeden raz, a po wylosowaniu nie zostaje zwrócony do populacji. Uzyskujemy wówczas próbę bez powtórzeń. Wynik z pierwszego losowania zmienia warunki następnego losowania, tj. zmienia prawdopodobieństwo wylosowania kolejnego elementu.

Jeśli populacja jest bardzo liczna, w wyniku losowania niezależnego i zależnego praktycznie dostajemy te same wyniki. Jeśli natomiast populacja nie jest liczna, lepiej jest przeprowadzić losowanie zależne, bowiem niesie ono większą gwarancję uzyskania reprezentatywnej próby. O ile w losowaniu niezależnym dokładność uzyskiwanych wyników zależy wyłącznie od wielkości próby, o tyle w losowaniu zależnym na wyniki wpływa, obok liczebności próby, również procentowy udział próby w ogólnej liczebności. W przypadku, gdy próba stanowi mały odsetek populacji generalnej, wyniki uzyskane w losowaniu zależnym będą podobne do wyników uzyskanych w losowaniu niezależnym. Pomimo to, stosować należy raczej losowanie zależne, gdyż jest ono technicznie prostsze, a zatem łatwiejsze do przeprowadzenia [9].

Losowanie jedno- i wielostopniowe. Losowanie jednostopniowe polega na losowaniu elementów do próby od razu (wprost, bezpośrednio). Przy losowaniu wielostopniowym może wystąpić kilka stopni losowania. Kolejny stopień losowania korzysta z wylosowanych poprzednio zestawów elementów populacji. Losowanie wieloetapowe nie gwarantuje uzyskania lepszych wyników w porównaniu z losowaniem bezpośrednim, ma jednak zalety

praktyczne: pozwala mniejszym nakładem pracy i w krótszym czasie pobrać próbę.

Losowanie indywidualne i zespołowe (grupowe). Losowanie indywidualne ma miejsce wtedy, gdy losujemy poszczególne elementy danej populacji. Gdy za jednostkę losowania przyjmujemy określony zestaw (grupę) badanych jednostek, które są elementami populacji, mamy wówczas do czynienia z losowaniem zespołowym (grupowym). Próba utworzona na podstawie takiego schematu losowania powstaje z elementów pochodzących ze wszystkich wylosowanych zestawów (grup).

Losowanie zespołowe możemy zastosować wtedy, gdy nie jest dany operat losowania i elementy pogrupowane są w zespoły, zawierające pewną liczbę elementów populacji. Taka sytuacja może wystąpić w przypadku bardzo dużej populacji, rozmieszczonej na dużym obszarze. Z mniejszych regionów próbę znacznie łatwiej pobrać, niż z dużych poprzez losowanie proste lub warstwowe z całości. Załóżmy, że populacja generalna podzielona została na M zespołów, a spośród wszystkich M zespołów wybieramy m zespołów. Możemy wyróżnić dwa przypadki. W pierwszym przypadku pobieramy do próby każdy element z każdego z m zespołów i mamy wówczas schemat losowania zespołowego jednostopniowego. W drugim przypadku losujemy m zespołów, a potem z każdego z nich losujemy n elementów. Ten schemat losowania nazywa się losowaniem zespołowym dwustopniowym.

Losowanie nieograniczone i ograniczone. Losowanie nieograniczone ma miejsce wtedy, gdy losujemy z całej populacji. W przypadku, kiedy losowania dokonujemy z poszczególnych części populacji, na które podzielono populację, to dokonujemy losowania ograniczonego. Należy pamiętać, że losujemy z każdej części oddzielnie. Częściami populacji mogą być tzw. warstwy, rozłączne i wewnętrznie jednorodne.

Losowanie warstwowe. Często zdarza się, że w populacji istnieją różne grupy, wewnątrz których elementy są pod pewnym względem do siebie podobne. W takich przypadkach możemy zwiększyć precyzję doboru próby (tj. zmniejszyć wariancję estymatorów), czy też zmniejszyć koszt przeprowadzenia badania przez oddzielne traktowanie różnych grup. Technika polegająca na losowaniu próby z każdej warstwy, a następnie uśrednianiu wyników nazywa się losowaniem warstwowym.

W losowaniu warstwowym zakładamy, że populacja N -elementowa może zostać podzielona na m N_i -elementowych grup, $i = 1, 2, \dots, m$. Wszystkie

m warstw to warstwy rozłączne; suma ich liczebności stanowi całą populację: $N_1 + N_2 + \dots + N_m = N$.

W praktyce stosuje się dwa podstawowe rodzaje losowania warstwowego: losowanie proporcjonalne i losowanie warstwowe optymalne. Nazwy te pochodzą od sposobu rozłożenia, czyli tzw. lokalizacji całej próby między poszczególne warstwy. Jeżeli z każdej warstwy wylosujemy do próby liczbę elementów proporcjonalną do wielkości warstwy, to losowanie warstwowe nosi nazwę losowania proporcjonalnego, jeśli zaś losujemy z poszczególnej warstwy liczbę elementów do próby proporcjonalną nie tylko do wielkości danej warstwy, lecz i odchylenia standardowego badanej cechy w warstwie, to losowanie warstwowe nosi nazwę losowania optymalnego bądź losowania Neymana.

Przy szacowaniu wartości średniej m cechy mierzalnej w populacji w oparciu o schemat losowania warstwowego próby stosujemy estymator, który jest ważoną średnią ze wszystkich uzyskanych w poszczególnych warstwach. Wariancja tego estymatora ma różną postać w zależności od tego, czy losowanie warstwowe było proporcjonalne, czy optymalne.

Losowanie systematyczne. Jeżeli elementy populacji są ponumerowane lub występują co pewien, ściśle określony czas, wówczas próba może być wylosowana w ten sposób, że w jej skład wchodzi jednostki oddzielone od siebie stałymi przedziałami liczbowymi lub czasowymi. Aby wybrać n -elementową próbę systematyczną spośród N elementów populacji, dzielimy te N elementów populacji na n grup po k elementów każda i stosujemy następującą regułę: z pierwszych k elementów populacji wybieramy losowo jeden element, w kolejnym kroku wybieramy każdy k -ty element po nim aż do uzyskania n -elementowej próby. Taki typ losowania nie daje jednakowych szans trafienia do próby wszystkim elementom. Np. do próby włączamy co 20-tą ($k=20$) jednostkę aż do uzyskania liczebności 40 potrzebnych elementów. Pierwszy element wybieramy losowo spośród liczb całkowitych od 1 do 20. Jeśli wylosowana liczba ma numer 7, to próba systematyczna będzie zawierać elementy: 7, $7+20 = 27$, $27+20 = 47$ i tak dalej, aż do uzyskania 40 potrzebnych elementów. W przypadku, gdy $k = N/n$ nie jest liczba całkowitą, wtedy przyjmujemy za k liczbę naturalną najbliższą tej wartości. Oprócz łatwości wybierania próby w sposób systematyczny, metoda ta ma inne zalety: jeśli $k = N/n$, to estymator z próby wartości średniej jest nieobciążony, losowanie systematyczne jest bardziej precyzyjne niż losowanie proste (dzieli populację na n k -elementowych warstw). Różnica między losowaniem prostym a systematycznym polega na tym, że losowanie systematyczne jest bardziej równomierne niż losowanie warstwowe, gdyż w losowaniu warstwowym próby z kolejnych warstw losuje się oddzielnie. Losowanie

to porównuje się czasem z losowaniem zespołowym, jako technikę wybierania jednego zespołu z populacji wszystkich k zespołów.

Odmowa odpowiedzi stanowi jeden z poważniejszych problemów pojawiających się w praktyce losowania próby, bowiem oznacza utratę informacji. Jeśli pobieramy próbę tylko z warstwy osób odpowiadających, wtedy pojawia się błąd obciążenia. Należy zatem pobrać próbę także z warstwy nieodpowiadających, co jest trudne do wykonania, bo niektórzy respondenci twardo nie chcą brać udziału w badaniu. Można w takiej sytuacji pobrać niewielką losową próbę spośród nieodpowiadających i zadać dwa pytania, z czego jedno krępujące, drugie obojętne, aby ocenić prawdopodobieństwo udzielenia odpowiedzi na krępujące pytanie. Zachowując prywatność respondentów możemy obliczyć zagregowaną odpowiedź na krępujące pytanie.

Przedstawiliśmy kilka zaawansowanych metod doboru próby, które pozwalają na uzyskanie dobrej precyzji badania. Nie wyczerpuje to obszernego tematu, dotyczącego pobierania próby losowej, który został szerzej rozwinęty także w innych opracowaniach.

Badamy cechę X populacji. Niech X_1, X_2, \dots, X_n będą zmiennymi losowymi niezależnymi o jednakowym rozkładzie, takim jak rozkład cechy X .

Próba losowa n -elementowa ze względu na cechę X (próba n -elementowa) jest to zmienna losowa n -wymiarowa X_1, X_2, \dots, X_n . Zmienna losowa X_1 jest modelem wartości cechy X pierwszego elementu wylosowanego z populacji do próby, X_2 modelem drugiego elementu itd. Losując ze zwracaniem powodujemy, że każdy element populacji ma taką samą szansę trafienia do próby, dlatego przyjmuje się, że zmienne losowe są niezależne [4]. Każdą wartość x_1, x_2, \dots, x_n próby nazywamy realizacją próby lub także próbą. Schematy i techniki losowania przedstawimy w aneksie.

Przykład 5.1

Rozważmy populację gospodarstw domowych w Radomiu. Populację tę badamy ze względu na cechę X -liczbę osób w gospodarstwie. Z populacji pobieramy próbę 5-elementową, a zatem losujemy ze zwracaniem 5 gospodarstw domowych. Przypuśćmy, że otrzymaliśmy wartość cechy X : 3, 2, 1, 4, 3. Wynika stąd, że zmienna losowa X_1 oznaczająca liczbę osób w wylosowanym pierwszym gospodarstwie domowym przyjęła wartość 3, zmienna losowa X_2 oznaczająca liczbę osób w wylosowanym drugim gospodarstwie domowym przyjęła wartość 2 itd.

Próba (X_1, X_2, \dots, X_n) przyjęła wartość $(3, 2, 1, 4, 3)$. Załóżmy, że badanie powtórzono, otrzymano następujące wartości cechy X : 3, 2, 1, 4, 3.

Otrzymaliśmy inną wartość próby, a mianowicie (3, 2, 1, 2, 4). Ciągi (3, 2, 1, 4, 3) i (3, 2, 1, 2, 4) są realizacjami próby $(X_1, X_2, X_3, X_4, X_5)$.

Aby można było przeprowadzić analizę statystyczną należy przekształcić próbę, czyli rozpatrywać funkcje próby. **Funkcje próby** nazywamy statystykami $U_n = g(X_1, X_2, \dots, X_n)$. Rozkład statystyki U_n będącej funkcją próby zależy od rozkładu w populacji i od liczebności próby.

Przykład 5.1 (cd.)

Jeżeli chcemy wyznaczyć średnią liczbę osób w gospodarstwach domowych wybranych do próby, to należy rozważyć zmienną losową. Zmienna ta jest funkcją próby X_1, X_2, X_3, X_4, X_5 – jest zatem statystyką. Wartościami (realizacjami) tej statystyki dla realizacji wybranej próby są liczby:

$$u_5 = \frac{3+2+1+4+3}{5} = 2,6, \quad u_5 = \frac{3+2+1+2+4}{5} = 2,4.$$

W tym przykładzie statystyką jest średnia z próby.

Jednym z najważniejszych problemów wnioskowania statystycznego jest umiejętność wyznaczania teoretycznego rozkładu prawdopodobieństwa danej statystyki U_n . Rozkład prawdopodobieństwa na ogół zależy od następujących czynników:

- liczebności próby,
- schematu (sposóbu) losowania elementów do próby,
- rozkładu populacji, z której próba pochodzi,
- postaci funkcyjnej g , określającej statystykę U_n .

Ze względu na liczebność próby wyróżniamy dwa podstawowe typy rozkładów statystyk: rozkłady dokładne i rozkłady graniczne.

Rozkładem dokładnym statystyki U_n nazywamy rozkład prawdopodobieństwa U_n wyznaczony dla każdej liczby naturalnej n , którą traktuje się jako parametr znany tego rozkładu i oznaczający liczebność próby losowej X . Rozkłady dokładne znajdują zastosowanie najczęściej w badaniach reprezentacyjnych, w których mamy do czynienia z małą liczbą obserwacji (n jest rzędu co najwyżej kilkudziesięciu).

Rozkładem granicznym statystyki U_n nazywamy rozkład prawdopodobieństwa U_n , który otrzymujemy przy dużej próbie, czyli dla $n \rightarrow \infty$. Jest to taki rozkład, którego dystrybuenta jest granicą ciągu dystrybuant rozkładów dokładnych statystyki U_n przy $n \rightarrow \infty$.

Określenie „duża próba” nie oznacza jednej liczby, za pomocą której wyrażamy liczebność próby. Ma to związek z szybkością zbieżności dokładnego rozkładu statystyki U_n do jej rozkładu granicznego. Dla niektórych statystyk przyjmujemy $n > 30$, dla innych postulat ten spełnia dopiero $n > 100$. Najczęściej obserwowanym typem rozkładu populacji, z której

losujemy próbę jest rozkład normalny, a najczęściej wykorzystywane we wnioskowaniu statystycznym statystyki to: miary skupienia, rozproszenia, korelacji, statystyki pozycyjne, momenty z próby. Ogromne znaczenie mają: średnia arytmetyczna, wariancja, odchylenie standardowe, mediana, współczynnik korelacji i współczynnik regresji. Pomijamy metody wyznaczania dokładnych statystyk ze względu na objętość opracowania.

5.2. Średnia z próby n-elementowej

Średnia z próby n-elementowej jest to statystyka

$$\bar{X}_n = \frac{1}{n} \sum_{i=1}^n X_i .$$

Jeżeli cecha X populacji ma wartość oczekiwaną m i wariancję σ^2 , to $E(\bar{X}_n) = m$, $D^2(\bar{X}_n) = \frac{\sigma^2}{n}$, $D(\bar{X}_n) = \frac{\sigma}{\sqrt{n}}$.

Jeżeli cecha X populacji ma rozkład normalny $N(m, \sigma)$, to średnia arytmetyczna \bar{X}_n ma rozkład normalny $N\left(m, \frac{\sigma}{\sqrt{n}}\right)$. Wynika to z własności rozkładu normalnego.

Przykład 5.2.1

Wiadomo, że cecha X populacji ma rozkład normalny $N(3,1)$. Obliczyć następujące prawdopodobieństwa:

a) $P(|X-3| < 0,1)$, b) $P(|\bar{X}_{16} - 3| < 0,1)$, c) $P(|\bar{X}_{144} - 3| < 0,1)$.

Rozwiązanie

a) $P(|X-3| < 0,1) = 2 \cdot \Phi(0,1) - 1 = 2 \cdot 0,5398 - 1 = 0,08$.

b) statystyka \bar{X}_{16} ma rozkład $N\left(3, \frac{1}{\sqrt{16}}\right)$, czyli $N(3, \frac{1}{4})$, stąd

$$\begin{aligned} P(|\bar{X}_{16} - 3| < 0,1) &= P\left(|\bar{X}_{16} - 3| : \frac{1}{4} < 0,1 : \frac{1}{4}\right) = 2 \cdot \Phi(0,4) - 1 \\ &= 2 \cdot 0,4556 - 1 = 0,30. \end{aligned}$$

c) statystyka \bar{X}_{144} ma rozkład $N\left(3, \frac{1}{\sqrt{144}}\right)$, czyli $N\left(3, \frac{1}{12}\right)$

$$\begin{aligned} P(|\bar{X}_{144} - 3| < 0,1) &= P\left(|\bar{X}_{144} - 3| : \frac{1}{12} < 0,1 : \frac{1}{12}\right) = 2 \cdot \Phi(1,25) - 1 \\ &= 2 \cdot 0,8944 - 1 = 1,7880 - 1 = 0,7880. \end{aligned}$$

Zauważamy, że prawdopodobieństwo dla zmiennej losowej X jest małe, dla wartości X_{16} umiarkowane, zaś dla wartości X_{144} duże. Średnia arytmetyczna dużej liczby zmiennych losowych ma rozkład skupiony przy wartości oczekiwanej.

Przykład 5.2.2

Z populacji o rozkładzie normalnym i odchyleniu standardowym σ pobrano próbę losową o liczebności n . Ile razy należy zwiększyć liczebność próby, aby odchylenie standardowe średniej arytmetycznej z próby zmalało trzykrotnie?

Rozwiązanie

Oznaczamy liczebność próby przez N . Z warunków zadania wynika, że zachodzi równość:

$$\frac{\sigma}{\sqrt{n}} = 3 \frac{\sigma}{\sqrt{N}}.$$

Po przekształceniach

$$\frac{\sigma^2}{n} = \frac{9\sigma^2}{N}, \quad \text{stąd } \frac{1}{n} = \frac{9}{N}, \quad \text{a zatem } N = 9n.$$

Aby zmniejszyć odchylenie standardowe średniej arytmetycznej z próby trzykrotnie, musimy zwiększyć liczebność próby dziewięciokrotnie.

5.3. Wariancja i odchylenie z próby n -elementowej

Wariancja z próby n -elementowej jest to statystyka

$$S_n^2 = \frac{1}{n} \sum_{i=1}^n (X_i - \bar{X}_n)^2.$$

Odchylenie z próby n -elementowej jest to statystyka

$$S_n = \sqrt{\frac{1}{n} \sum_{i=1}^n (X_i - \bar{X}_n)^2}.$$

Jeśli elementy próby mało różnią się od siebie, wówczas realizacja s_n^2 statystyki S_n^2 jest liczbą bliską zera. Jeśli natomiast elementy próby różnią się znacznie od siebie, to realizacja jest dużą liczbą. Podobnie dzieje się z odchyleniem standardowym. A zatem, statystyki S_n^2 i S_n są miarami zróżnicowania elementów próby względem średniej z próby.

Z wariancją z próby związane są statystyki

$$\widehat{S}_n^2 = \frac{1}{n-1} \sum_{i=1}^n (X_i - \bar{X}_n)^2 \text{ oraz } \check{S}_n^2 = \frac{1}{n} \sum_{i=1}^n (X_i - m)^2,$$

gdzie m oznacza wartość oczekiwaną cechy X w populacji. Między statystykami \widehat{S}_n^2 i \check{S}_n^2 występują związki:

$$\widehat{S}_n^2 = \frac{n}{n-1} S_n^2 \text{ oraz } n \cdot S_n^2 = (n-1) \widehat{S}_n^2 = \sum_{i=1}^n (X_i - \bar{X}_n)^2.$$

Jeżeli założymy, że cecha X populacji ma rozkład normalny $N(m, \sigma)$, to statystyka

$$\frac{n \cdot \widehat{S}_n^2}{\sigma^2} = \sum_{i=1}^n \left(\frac{X_i - m}{\sigma} \right)^2$$

jest sumą kwadratów n niezależnych zmiennych losowych o rozkładzie normalnym $N(0,1)$, a zatem ma rozkład χ^2 z n stopniami swobody.

Statystyka

$$\frac{n \cdot S_n^2}{\sigma^2} = \sum_{i=1}^n \left(\frac{X_i - \bar{X}_n}{\sigma} \right)^2$$

różni się tym od statystyki $\frac{n \cdot \widehat{S}_n^2}{\sigma^2}$, że zamiast różnicy $X_i - m$ występuje różnica $X_i - \bar{X}_n$. Dowodzi się, że ma ona także rozkład χ^2 , lecz z $n-1$ stopniami swobody.

Jeżeli cecha X populacji ma wariancję σ^2 , to ciągi $\{S_n^2\}, \{\widehat{S}_n^2\}, \{\check{S}_n^2\}$ są zbieżne według prawdopodobieństwa do wariancji σ^2 , zaś ciągi $\{S_n\}, \{\widehat{S}_n\}, \{\check{S}_n\}$ są zbieżne według prawdopodobieństwa do odchylenia standardowego σ . Gdy próba jest liczna wartości statystyk $S_n^2, \widehat{S}_n^2, \check{S}_n^2$ mogą służyć do oceny wariancji σ^2 , zaś wartości statystyki $S_n, \widehat{S}_n, \check{S}_n$ do oceny odchylenia standardowego σ . Wartości oczekiwane statystyk $S_n^2, \widehat{S}_n^2, \check{S}_n^2$ są następujące:

$$E(S_n^2) = \frac{n-1}{n} \sigma^2, \quad E(\widehat{S}_n^2) = \sigma^2, \quad E(\check{S}_n^2) = \sigma^2.$$

Przykład 5.3.1

Z populacji o rozkładzie normalnym $N(m, 4)$ wylosowano 10-elementową próbę. Jakie jest prawdopodobieństwo, że wariancja z próby S^2 przekroczy liczbę 20.

Rozwiązanie

Wykorzystamy fakt, że statystyka $\frac{n}{\sigma^2} S^2$ ma rozkład χ^2 o $n-1$ stopniach swobody, a zatem:

$$\frac{n}{\sigma^2} S^2 \leq \frac{n}{\sigma^2} \cdot 20$$

$$\begin{aligned} P(S_{10}^2 \leq 20) &= P\left(\frac{n}{\sigma^2} S^2 \leq \frac{n}{\sigma^2} \cdot 20\right) = P\left(\frac{n}{\sigma^2} S^2 \leq \frac{n}{16} \cdot 20\right) = \\ &= P\left(\frac{n}{\sigma^2} S^2 \leq 12,5\right) = 0,815. \end{aligned}$$

Z warunków zadania wynika, że

$$P(S_{10}^2 > 20) = 1 - P(S_{10}^2 \leq 20) = 1 - 0,815 = 0,185.$$

Korzystając z tablic rozkładu χ^2 dla 9 stopni swobody oraz z prawdopodobieństwa przeciwnego obliczyliśmy szukane prawdopodobieństwo.

Przykład 5.3.2

Z populacji o rozkładzie normalnym $N(m, \sqrt{2})$ wylosowano próbę 9-elementową. Jakie jest prawdopodobieństwo, że wariancja z próby S^2 nie przekroczy liczby 4.

Rozwiązanie

Ponieważ statystyka $\frac{n}{\sigma^2} S^2$ ma rozkład χ^2 o $n-1$ stopniach swobody, stąd

$$\frac{n}{\sigma^2} S^2 \leq \frac{n}{\sigma^2} \cdot 4$$

$$\begin{aligned} P(S_9^2 \leq 4) &= P\left(\frac{n}{\sigma^2} S^2 \leq \frac{n}{\sigma^2} \cdot 4\right) = P\left(\frac{n}{\sigma^2} S^2 \leq \frac{9}{2} \cdot 4\right) = \\ &= P\left(\frac{n}{\sigma^2} S^2 \leq 18\right) \approx 0,978. \end{aligned}$$

Korzystając z tablic rozkładu χ^2 znajdujemy dla 8 stopni swobody szukane prawdopodobieństwo.

5.4. Statystyka Studenta z próby

Definicja *rozkładu t-Studenta* (Gosseta) z n stopniami swobody jako rozkładu zmiennej losowej jest następująca:

$$T = \frac{X}{\sqrt{\frac{Y_n}{n}}},$$

gdzie X jest zmienną losową o rozkładzie normalnym $N(0, 1)$, Y_n jest zmienną losową o rozkładzie χ^2 z n stopniami swobody. Zakładamy ponadto, że X , Y są zmiennymi losowymi niezależnymi (porównaj definicję z rozdz. III). Jeśli cecha X populacji ma rozkład normalny $N(m, \sigma)$, to statystyka

$$U_n = \frac{\bar{X}_n - m}{S_n} \cdot \sqrt{n-1}$$

ma rozkład t-Studenta z $n-1$ stopniami swobody. Do rozkładu t-Studenta sięgamy między innymi wtedy, gdy nie jest znana wartość odchylenia standardowego σ populacji, która posiada rozkład normalny. Szczególnie w przypadku małej próby, gdy nie można zastosować statystyki \bar{X} przydatna okazuje się statystyka t-Studenta.

Przykład 5.4.1

Z populacji o rozkładzie normalnym $N(12, \sigma)$ i nieznanym odchyleniu standardowym pobrano próbę losową.

- Jakie jest prawdopodobieństwo, że średnia arytmetyczna w próbie 10-elementowej jest większa od 11,5? Odchylenie w tej próbie wynosi 1,5.
- Jakie jest prawdopodobieństwo, że średnia arytmetyczna w próbie 50-elementowej jest większa od 11,5?

Rozwiązanie

- a) Zmienna losowa $U_n = \frac{\bar{X}_n - m}{S_n} \cdot \sqrt{n-1}$ ma rozkład t-Studenta z 9 stopniami swobody.

$$P(\bar{X} > 11,5) = P\left(\frac{\bar{X} - 12}{1,5} \cdot \sqrt{10-1} > \frac{11,5 - 12}{1,5} \cdot \sqrt{10-1}\right) =$$

$$P(U_9 > -1) = 1 - P(U_9 < -1) \approx 1 - \frac{0,30}{2} = 1 - 0,15 = 0,85.$$

Prawdopodobieństwo, że średnia z próby 10-elementowej jest większa od 11,5 wynosi 0,85.

b) Korzystamy z faktu, że dla dużych n rozkład statystyki

$$U_n = \frac{\bar{X}_n - m}{S_n} \cdot \sqrt{n-1} \text{ jest zbieżny do rozkładu } N(0,1), \text{ stąd}$$

$$P(\bar{X} > 11,5) = P\left(\frac{\bar{X} - 12}{1,5} \cdot \sqrt{50-1} > \frac{11,5 - 12}{1,5} \cdot \sqrt{50-1}\right) =$$

$$P\left(U > \frac{-0,5}{1,5} \cdot 7\right) = P(U < -2,33) = 0,5 + \Phi(2,33) \approx 0,5 + 0,4893 = 0,9893.$$

Prawdopodobieństwo, że średnia z próby jest większa od 12,5 wynosi 0,9893.

5.5. Rozkład frakcji z próby

Rozważając *rozkład wskaźnika struktury* (frakcji z próby) wprowadzimy oznaczenia:

Y_n – zmienna losowa oznaczająca liczbę sukcesów w n doświadczeniach Bernoulliego,

p – prawdopodobieństwo sukcesu w jednym doświadczeniu ($q = 1-p$),

$\frac{Y_n}{n}$ – wskaźnik struktury (częstość sukcesu, frakcja).

Ponieważ zmienna losowa Y_n ma rozkład dwumianowy z parametrami n i p , to na mocy integralnego twierdzenia Moivre'a-Laplace'a ma ona rozkład w przybliżeniu normalny $N(np, \sqrt{npq})$, a zatem częstość sukcesu $\frac{Y_n}{n}$ ma

rozkład w przybliżeniu normalny $N(np, \sqrt{\frac{pq}{n}})$. Tak więc częstość sukcesu

$\frac{Y_n}{n}$ ma rozkład w przybliżeniu normalny (asymptotycznie normalny)

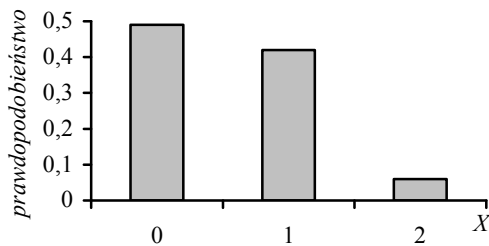
$N(np, \sqrt{\frac{pq}{n}})$, gdzie p jest prawdopodobieństwem sukcesu w jednym doświadczeniu.

Rozkład frakcji z próby \hat{p} jest rozkładem dwumianowym z parametrami n i p , gdzie n jest liczebnością próby, a p frakcją elementów danej kategorii w populacji. Gdy liczebność próby n wzrasta, wówczas znajduje zastosowanie centralne twierdzenie graniczne. Gdy $p = 0,5$, tzn. w przypadku rozkładu symetrycznego, rozkład dwumianowy jest bliski normalnemu przy wszystkich n .

Symetria rozkładu powoduje, że zbieżność rozkładu do normalnego jest szybka. Rozważmy przypadek, gdy rozkład dwumianowy nie jest syme-

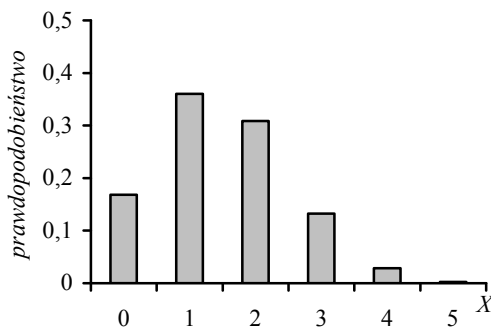
tryczny zakładając że $p = 0,25$, zaś $n = 2$, $n = 5$, $n = 10$, $n = 15$. Z rysunków 5.5.1, 5.5.2, 5.5.3, 5.5.4, wynika, że przy małych n rozkład jest prawostronnie asymetryczny. Gdy liczebność próby wzrasta, to rozkład frakcji z próby zbliża się do rozkładu normalnego.

Dla $n = 2$ $p = 0,3$ – sukces, $q = 1 - p = 0,7$ – porażka.



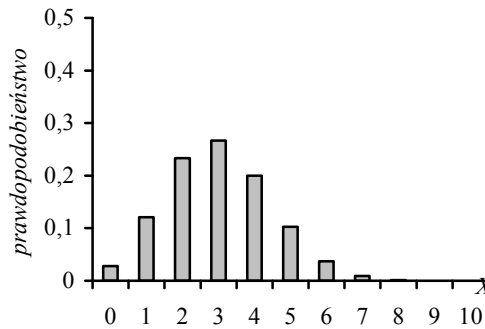
Rys. 5.5.1. Rozkład frakcji z próby dla $n = 2$

Dla $n = 5$ $p = 0,3$; $q = 0,7$



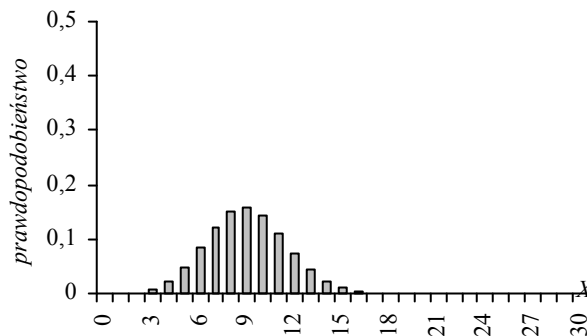
Rys. 5.5.2. Rozkład frakcji z próby dla $n = 5$

Dla $n = 10$ $p = 0,3$; $q = 0,7$



Rys. 5.5.3. Rozkład frakcji z próby dla $n = 10$

Dla $n = 30$ $p = 0,3$; $q = 0,7$



Rys. 5.5.4. Rozkład frakcji z próby dla $n = 30$

Przykład 5.5.1

Zakładamy, że wśród badanej populacji o odpowiednich dochodach i reprezentujących odpowiedni styl życia 25% jest zainteresowanych zakupem samochodu dobrej marki. Wylosowano 100 osób z tej grupy. Jakie jest prawdopodobieństwo, że co najmniej 20% osób wylosowanych do próby zakupi samochód?

Rozwiązanie

Stosowane w praktyce reguły zalecają, że posługiwanie się normalnymi przybliżeniami rozkładu frakcji jest dopuszczalne wtedy, gdy zarówno $n \cdot p$, jak też $n \cdot q$ są większe od 5. Z warunków zadania wynika, że $n \cdot p = 100 \cdot 0,25 = 25$ oraz $n \cdot q = 100 \cdot 0,75 = 75$. Obie liczby są większe od 5, a więc możemy wykorzystać normalne przybliżenie rozkładu frakcji z próby. Średnią jest $p = 0,25$, zaś odchyleniem standardowym jest $\sqrt{\frac{pq}{n}} = 0,0433$. A zatem

$$P(\hat{p} \geq 0,20) = P\left(U \geq \frac{0,20-0,25}{0,0433}\right) = P(U \geq -1,15) = 0,8749 \text{ (87,49\%)}.$$

Prawdopodobieństwo, że co najmniej 20% osób wylosowanych do próby zakupi samochód dobrej marki wynosi 87,49%.

ROZDZIAŁ VI ESTYMACJA PARAMETRÓW

Estymacja polega na szacowaniu wartości parametrów lub postaci rozkładu zmiennej losowej w populacji generalnej na podstawie rozkładu empirycznego uzyskanego z próby. Wychodzimy od wyników próby i na ich podstawie formułujemy wnioski o populacji generalnej. W teorii estymacji wyróżnia się: estymację parametryczną (dotyczy szacowania parametrów rozkładu) i estymację nieparametryczną (dotyczy szacowania postaci funkcyjnej rozkładu, np. w postaci dystrybuanty). Do oceny parametru możemy zastosować estymację punktową lub estymację przedziałową.

6.1. Estymacja parametrów populacji generalnej

W teorii estymacji podstawową rolę odgrywa pojęcie *estymatora*. Estymator jest to wielkość (statystyka, charakterystyka) wyznaczana na podstawie próby losowej, służąca do oceny wartości nieznanymi parametrów populacji generalnej.

Estymatorem parametru Θ rozkładu zmiennej losowej X nazywamy taką statystykę: $Z_n = f(X_1, X_2, \dots, X_n)$, której rozkład zależy od szacowanego parametru Θ .

Na przykład dla populacji generalnej o rozkładzie $N(m, \sigma)$ z nieznaną średnią m i wariancją σ^2 statystyki \bar{X} oraz S^2 , tj. średnia i wariancja z próby są estymatorami, gdyż ich rozkłady zależą od odpowiednich parametrów populacji.

Estymator jako statystyka z próby jest zmienną losową i jego rozkład jest determinowany przez rozkład zmiennej losowej X w populacji generalnej.

Estymacji pewnego parametru za pomocą estymatora Z_n dokonuje się na podstawie wyników próby losowej, dlatego istnieje możliwość popełnienia błędu.

Ocena parametru Θ to wartość liczbowa $z_n = f(x_1, x_2, \dots, x_n)$, jaką przyjmuje estymator Z_n parametru Θ dla realizacji próby (x_1, x_2, \dots, x_n) . Ocena z_n jest zatem realizacją zmiennej losowej.

Błąd szacunku (estymacji) to różnica pomiędzy uzyskaną liczbową oceną z_n parametru Θ i jego prawdziwą wartością, tj. $z_n - \Theta$. Na błąd szacunku, który jest również zmienną losową, mają wpływ zarówno prawidłowy dobór próby losowej, szczególnie jej liczebność, jak i wybór najlepszego estymatora Z_n .

Własności dobrego estymatora

Nieobciążoność. Estymator Z parametru Θ nazywamy nieobciążonym, jeżeli jego wartość oczekiwana jest równa szacowanemu parametrowi. Gwarantuje on, że za jego pomocą możliwe jest uzyskiwanie ocen wolnych od błędu systematycznego, tj.

$$\bigwedge_{n \in \mathbb{N}} E(Z) = \Theta.$$

Zgodność. Estymator Z parametru Θ nazywamy zgodnym, jeśli jest stochastycznie zbieżny (zbieżny według prawdopodobieństwa) do szacowanego parametru Θ , tj.

$$\lim_{n \rightarrow \infty} P\{|Z_n - \Theta| < \varepsilon\} = 1.$$

Estymator jest zgodny, gdy podlega prawu wielkich liczb. Jeżeli liczebność próby rośnie, to prawdopodobieństwo, że wartość estymatora Z różni się dowolnie mało od wartości parametru Θ zbliża się do 1. Przy konkretnej (małej) próbie kryterium zgodności nie znajduje zastosowania.

Efektywność. Nieobciążony estymator Z parametru Θ nazywamy efektywnym, gdy ma najmniejszą wariancję spośród nieobciążonych estymatorów tego parametru. Poziom wariancji jest tu miarą efektywności estymatorów. Estymatory efektywne nie zawsze istnieją.

Dostateczność (wystarczalność). Estymator Z parametru Θ nazywamy dostatecznym, gdy zawiera wszystkie informacje, jakie na temat parametru Θ występują w próbie i żaden inny estymator nie może dać dodatkowych informacji o szacowanym parametrze.

Dla zrozumienia wielu ważnych rozkładów stosowanych w praktyce kluczowe znaczenie ma pojęcie **stopni swobody**. Pozwala ono wyjaśnić, dlaczego wariancja próby S^2 jest nieobciążonym estymatorem σ^2 , a więc pozwala odpowiedzieć na pytanie, dlaczego licznik wzoru na wariancję S^2 dzielimy przez $n - 1$, a nie przez n .

Przykład 6.1.1

Rozważmy wyniki obserwacji czterech zmiennych X_1, X_2, X_3, X_4 , których wartości z próby są następujące: $x_1 = 4, x_2 = 6, x_3 = 10, x_4 = 12$. Obliczamy:

$$\bar{x} = \frac{1}{4}(x_1 + x_2 + x_3 + x_4) = \frac{1}{4}(4 + 6 + 10 + 12) = 8.$$

Interesuje nas, ile z czterech możliwych wyników obserwacji zmiennych X_1, X_2, X_3, X_4 , mogłoby się swobodnie zmieniać? Ilu wyników obserwacji moż-

na przypisać dowolne wartości, jeśli ich średnia jest znana? Załóżmy, że nie znamy wartości zmiennej X_4 w próbie, mamy więc: $\frac{1}{4}(4+6+10+x_4) = 8$. Rozwiązując to równanie z jedną niewiadomą znajdujemy nieznaną wartość $x_4 = 12$.

Gdy w zbiorze n wyników nie znamy jednego z nich, ale znamy średnią wszystkich n wyników, wówczas ten wynik jest w istocie znany, bo łatwo wyznaczony przez $n - 1$ wyników obserwacji oraz wartość średnią. Czwarty możliwy wynik nie może się swobodnie zmieniać.

Gdy w zbiorze n wyników obserwacji nie znamy więcej niż jednego z nich, np. dwóch zmiennych: X_3 i X_4 , ale znamy średnią z próby, wtedy możemy napisać równanie:

$$\frac{1}{4}(4+6+x_3+x_4) = 8, \quad \text{a stąd } 4+6+x_3+x_4 = 32.$$

Otrzymaliśmy równanie z dwiema niewiadomymi, którego nie możemy jednoznacznie rozwiązać. Istnieje pewien stopień swobody; te wartości mogą się swobodnie „poruszać”. Jeżeli przypiszemy jednemu z tych nieznanych wyników określoną wartość, to wartość tego drugiego może być ustalona na podstawie trzech wyników obserwacji i znanej średniej. Mamy jeden stopień swobody.

Gdy rozpatrujemy n możliwych wyników obserwacji i znamy ich średnią, to średnia działa jako ograniczenie na wyniki obserwacji, pozostawiając nam $n - 1$ stopni swobody. Średnia jest ograniczeniem w tym sensie, że suma odchyłeń od średniej musi być równa zero. W naszym przykładzie odchylenia przyjmują wartości:

$$d_1 = x_1 - \bar{x} = 4 - 8 = -4,$$

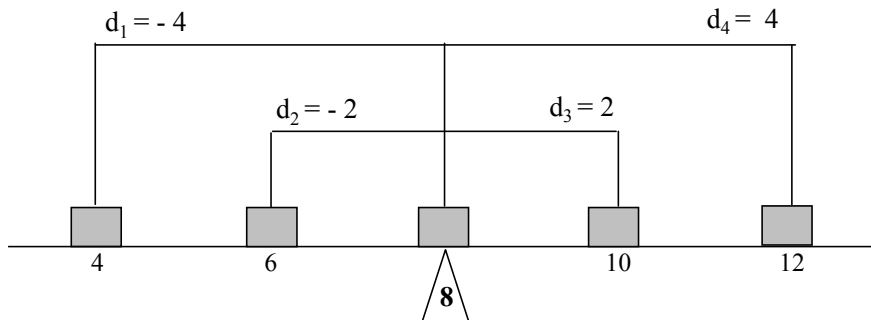
$$d_2 = x_2 - \bar{x} = 6 - 8 = -2,$$

$$d_3 = x_3 - \bar{x} = 10 - 8 = 2,$$

$$d_4 = x_4 - \bar{x} = 12 - 8 = 4.$$

Zauważamy, że suma odchyłeń $d_1 + d_2 + d_3 + d_4 = (-4) + (-2) + 2 + 4 = 0$. Dowolne trzy odchylenia od znanej średnie mogą zmieniać się swobodnie bez naruszenia ograniczenia, jakim jest $\bar{x} = 8$, o ile tylko czwarte odchylenie zostanie tak dobrane, aby suma wszystkich odchyłeń sprowadzała się do zera. Ponownie zauważamy, że gdy średnia jest znana, to tylko $n - 1$ wyników obserwacji może się swobodnie „poruszać”, a więc mamy $n - 1$ stopni swobody.

Jeśli wartość któregośkolwiek wyniku obserwacji jest nieznaną, to możemy obliczyć jej odchylenie od średniej wiedząc, że suma odchyłeń musi być równa zero. Wynika to z własności średniej arytmetycznej.



Rys. 6.1.1. Cztery odchylenia wyników obserwacji od ich średniej

Skoncentrujemy się na wariancji z próby S^2 , która jest sumą kwadratów odchyłeń wyników obserwacji od średniej podzieloną przez liczebność. Gdy mamy do czynienia z danymi opisującymi całą populację, wówczas wariancję definiujemy następująco:

$$\sigma^2 = \frac{1}{n} \sum_{i=1}^N (x_i - m)^2 .$$

W przypadku zbioru danych, które są próbkami z populacji, a nie całymi populacjami, należy posługiwać się definicją wariancji:

$$s^2 = \frac{1}{n-1} \sum_{i=1}^N (x_i - \bar{x})^2 .$$

Ponieważ obliczenie wariancji zakłada znajomość \bar{x} i n wyników obserwacji, to jak wiadomo z wcześniejszych rozważań tylko $n - 1$ wyników obserwacji może być dowolnie ustalonych. Dlatego też, jeśli chcemy posługiwać się średnim odchyleniem kwadratowym jako miarą zmienności w próbie, to przeciętna powinna opierać się tylko na $n - 1$ swobodnych wynikach obserwacji.

Wyjaśniony został powód, dla którego obliczając S^2 , nieobciążony estymator wariancji w populacji σ^2 , sumę kwadratów odchyłeń od średniej dzielimy przez $n - 1$. Intuicyjne zrozumienie pojęcia stopni swobody jest bardzo ważne w wielu zastosowaniach. Liczba stopni swobody jest równa liczbie wszystkich pomiarów (która nie musi być równa liczbie wyników obserwacji) pomniejszonej o liczbę wszystkich ograniczeń narzuconych na te pomiary. Ograniczeniem jest każda wielkość, która zostaje obliczona na podstawie tych samych pomiarów.

Przykład 6.1.2

Kierownik pewnej firmy usługowej dysponuje sumą 5 000 zł, która ma być rozdzielona w całości na doskonalenie pięciu stanowisk pracy. Ile stopni swobody ma kierownik?

Rozwiązanie

Założenie, że na doskonalenie pięciu stanowisk wydana będzie suma 5 000 zł oznacza, że średnio biorąc na każde stanowisko przeznaczony się 1 000 zł. Kierownik może swobodnie wybierać (w granicach 5 000 zł) sumy przeznaczone na doskonalenie czterech stanowisk. Gdy te sumy przydzielą, nie będzie już mógł rozważać, ile przydzielić na rozwój ostatniego, piątego stanowiska, bo to musi być suma dokładnie równa różnicy między całą sumą a sumą przydzieloną czterem stanowiskom. Kierownik ma zatem 1 stopień swobody.

6.1.1. Estymacja punktowa

Estymacja punktowa polega na ocenie nieznanego wartości parametru Θ w populacji za pomocą jednej konkretnej liczby, otrzymanej z wyników badań próby losowej.

Estymatorem średniej arytmetycznej z populacji generalnej $E(X) = m$ jest średnia arytmetyczna z próby losowej \bar{x} o rozkładzie normalnym w przypadku dużych prób oraz o rozkładzie t-Studenta dla małych prób.

Estymatorem częstości względnej z populacji generalnej $E(W) = p_i$ jest częstość względna (wskaźnik struktury) z próby losowej w_i o rozkładzie normalnym w przypadku dużych prób oraz o rozkładzie Bernoulliego dla małych prób.

Estymatorem wariancji z populacji generalnej $E(S^2(X)) = V(X)$ jest wariancja z próby S^2 o rozkładzie normalnym w przypadku dużych prób oraz o rozkładzie χ^2 (chi kwadrat) dla małych prób.

Przy ocenie punktowej prawdopodobieństwo tego, że estymator przyjmie wartość równą wartości szacowanego parametru jest bliskie zeru ponieważ oceny tego samego parametru uzyskane z różnych prób w oparciu o ten sam estymator różnią się prawie zawsze między sobą. Oznacza to, że z prawdopodobieństwem bliskim jedności popełniamy błąd, szacując punktowo parametr Θ w populacji. Dlatego lepsze (wiarygodniejsze) oszacowanie parametru Θ uzyskuje się stosując estymację przedziałową.

Przykład 6.1.1.1

W pewnym mieście wylosowano 10 sklepów w określonym dniu i zbadano cenę butów (w zł) dostarczanych przez producentów A i B. Rezultaty obserwacji przedstawiono w tabeli (dane umowne).

Sklep	Cena butów (w zł)	Producent
1	120	A
2	117	A
3	116	A
4	118	A
5	112	B
6	119	A
7	115	B
8	116	A
9	114	B
10	118	A

- Oszacować przeciętną cenę butów w określonym dniu w całej zbiorowości (w całym mieście),
- Oszacować odchylenie standardowe ceny butów w określonym dniu w badanym mieście,
- Oszacować odsetek (frakcję) butów pochodzących od producenta A wśród ogółu butów,
- Wyznaczyć błąd standardowy frakcji butów pochodzących od producenta A.

Rozwiązanie

Tablica obliczeniowa

Sklep	Cena butów (w zł)	Producent	$x_i - \bar{x}$	$(x_i - \bar{x})^2$
1	120	A	3,3	10,89
2	117	A	0,3	0,09
3	116	A	-0,7	0,49
4	118	A	1,3	1,69
5	112	B	-4,7	22,09
6	119	A	2,3	5,29
7	115	B	-1,7	2,89
8	116	A	-0,7	0,49
9	114	B	-2,7	7,29
10	118	A	1,3	1,69
Razem	1167	x	x	52,90

- wiedząc, że estymacja punktowa sprowadza się do znalezienia jednej wartości służącej do oszacowania nieznannej średniej ceny butów w całym

mieście, znajdujemy średnią arytmetyczną (\bar{x}) na podstawie próby. Korzystając z tablicy obliczeniowej mamy

$$\bar{x} = \frac{\sum_{i=1}^{10} x_i}{n} = \frac{1167}{10} = 116,7.$$

Przeciętna cena butów w tym mieście, pochodzących od producentów A i B wynosi 116,7 zł.

$$b) s = \sqrt{\frac{\sum (x_i - \bar{x})^2}{n-1}} = \sqrt{\frac{52,90}{10-1}} = 2,42.$$

Odchylenie standardowe ceny butów dostarczanych przez producentów A i B w badanym mieście wynosiło 2,42 zł. Oznacza to, że cena butów w poszczególnych sklepach różniła się od średniego poziomu o $\pm 2,42$ zł, przeciętnie biorąc.

$$c) \hat{p} = \frac{7}{10} = 0,7 \text{ (70\%)}$$

Producent A dostarcza 70% ogółu wszystkich butów do badanych sklepów.

d) błąd standardowy frakcji z próby

$$\sigma_{\hat{p}} = \sqrt{\frac{p(1-p)}{n}} = \sqrt{\frac{0,70 \cdot 0,30}{10}} = 0,14 \text{ (14\%)}$$

Błąd standardowy frakcji butów dostarczanych przez producenta A w sklepach w badanym mieście wynosi 14%.

6.1.2. Estymacja przedziałowa

Estymacja przedziałowa polega na budowaniu przedziału liczbowego (tzw. przedziału ufności), w którym z określonym z góry prawdopodobieństwem bliskim jedności będzie zawierała się wartość szacowanego parametru Θ .

Przedział ufności budujemy na podstawie próby, gdzie obserwujemy wartości realizacji zmiennych losowych X_1, X_2, \dots, X_n , których rozkład zależy od Θ .

Niech $Y_1 = g_1(X_1, X_2, \dots, X_n)$ oraz $Y_2 = g_2(X_1, X_2, \dots, X_n)$ będą statystykami (nie zależą od Θ). Jeżeli dla z góry ustalonego $\alpha \in (0,1)$ spełniony jest warunek:

$$P\{g_1(Z_n) < \Theta < g_2(Z_n)\} = 1 - \alpha,$$

to przedział (Y_1, Y_2) nazywamy **przedziałem ufności** dla parametru Θ , a liczbę $1 - \alpha$ poziomem ufności lub **współczynnikiem ufności**. Oznacza to,

że z prawdopodobieństwem $1 - \alpha$ przedział ufności obejmuje szacowany parametr Θ .

Przedział ufności jest przedziałem losowym i jego położenie na osi liczbowej ulega zmianom. Przedziały ufności są na ogół różne dla różnych prób. Szacowany, nieznaną parametr Θ jest stały, nie jest zmienną losową.

Poziom ufności jest to prawdopodobieństwo równe $1 - \alpha$, z jakim przedział ufności pokrywa nieznaną wartość parametru Θ . Najczęściej przyjmuje się poziom ufności równy 0,90, 0,95 lub 0,99.

Długość przedziału ufności jest to różnica pomiędzy górną $g_2(Z_n)$ i dolną $g_1(Z_n)$ granicą przedziału lub końcami przedziału, które są zmiennymi losowymi.

Miara precyzji estymacji przedziałowej jest to długość przedziału ufności. Dokładność (precyzja) estymacji przedziałowej zwiększa się wraz ze zmniejszaniem się długości przedziału ufności. Nieuzasadnione przyjmowanie wysokiego współczynnika ufności nie jest korzystne, ponieważ przy ustalonej liczności próby rozszerza się przedział ufności, czyli maleje precyzja szacunku.

Poziom ufności $1 - \alpha$ ma interpretację częstościową. Dla każdej serii niezależnych prób o stałej wielkości n w około $(1 - \alpha) \cdot 100\%$ przypadków przedziały ufności pokrywają nieznaną wartość parametru Θ , natomiast w około $\alpha \cdot 100\%$ przypadkach prawdziwa wartość Θ nie będzie pokryta przez przedziały ufności.

Przedział ufności dla wartości przeciętnej

Jeżeli cecha X w zbiorowości generalnej ma rozkład $N(m, \sigma)$, to przedział ufności dla parametru m ma postać:

- dla dużych prób $n > 30$:

$$\bar{X} - t_\alpha \frac{\sigma}{\sqrt{n}} < m < \bar{X} + t_\alpha \frac{\sigma}{\sqrt{n}},$$

gdzie: \bar{X} – średnia arytmetyczna z obserwacji w próbie,

t_α – parametr odczytywany z tablic rozkładu normalnego dla relacji

$$\Phi(t_\alpha) = \frac{1 - \alpha}{2};$$

- dla małych prób $n \leq 30$:

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

lub

$$\bar{X} - t_{\alpha, n-1} \frac{S_1}{\sqrt{n}} < m < \bar{X} + t_{\alpha, n-1} \frac{S_1}{\sqrt{n}},$$

gdzie: $t_{\alpha, n-1}$ – parametr odczytywany z tablic rozkładu t-Studenta dla $n - 1$ stopni swobody,

$$P\{|T_{n-1}| > t_{\alpha, n-1}\} = \alpha,$$

$$S = \sqrt{\frac{\sum_{i=1}^n (x_i - \bar{x})^2}{n}},$$

$$S_1 = \sqrt{\frac{\sum_{i=1}^n (x_i - \bar{x})^2}{n-1}}.$$

Maksymalny błąd szacunku $d_{\bar{x}}$ wartości przeciętnej m jest równy połowie długości przedziału ufności, przy czym dla:

– dużych prób

$$d_{\bar{x}} = t_{\alpha} \frac{\sigma}{\sqrt{n}},$$

– małych prób

$$d_{\bar{x}} = t_{\alpha, n-1} \frac{S}{\sqrt{n-1}}.$$

Względny stopień precyzji szacunku parametru m :

$$V_m = \frac{t_{\alpha} \cdot \sigma}{\bar{X} \cdot \sqrt{n}} \cdot 100\%.$$

Minimalna liczność próby. Jeżeli cecha X ma rozkład normalny, to minimalną liczebność próby, niezbędną do oszacowania wartości przeciętnej m na poziomie ufności $1 - \alpha$, z maksymalnym błędem szacunku nieprzekraczającym $d_{\bar{x}}$ obliczamy ze wzoru:

$$n = \frac{t_{\alpha}^2 \cdot \sigma^2}{d_{\bar{x}}^2}.$$

Uwagi

1. Długość przedziału ufności wynosi $2t_{\alpha} \frac{\sigma}{\sqrt{n}}$ lub $2t_{\alpha, n-1} \frac{S_1}{\sqrt{n}}$.

2. Przy zadanym poziomie ufności im większa jest liczebność, tym krótszy przedział ufności.
3. Przy ustalonej liczebności próby, wraz ze wzrostem poziomu ufności rośnie rozpiętość (długość) przedziału ufności.
4. Im krótszy przedział, tym mniejszy błąd szacunku, co oznacza większą precyzję (dokładność) szacunku.

Przykład 6.1.2.1

Trwałość pracy pewnego urzędnika jest zmienną losową X o rozkładzie $N(m, 100 \text{ h})$. Z wyprodukowanej partii pobrano próbę 16 urzędników i otrzymano średnią trwałość $\bar{x} = 2600 \text{ h}$. Oszacować średnią trwałość urzędnika z tej partii na poziomie ufności 0,95. Obliczyć względny błąd oszacowania.

Rozwiązanie

Zastosujemy przedział ufności dla średniej populacji generalnej, kiedy znane jest odchylenie standardowe

$$P(\bar{x} - t_{\alpha} \frac{\sigma}{\sqrt{n}} < m < \bar{x} + t_{\alpha} \frac{\sigma}{\sqrt{n}}) = 1 - \alpha .$$

Mamy $\Phi(t_{\alpha}) = \frac{1 - \alpha}{2} = 0,475 \Rightarrow t_{\alpha} = 1,96$.

A zatem błąd bezwzględny, czyli połowa długości przedziału

$$d_{\bar{x}} = t_{\alpha} \frac{\sigma}{\sqrt{n}} = \frac{1,96 \cdot 100}{\sqrt{16}} = 49 \text{ h} .$$

W naszym przykładzie przedział ufności będzie więc następujący:

$$2600 - \frac{1,96 \cdot 100}{\sqrt{16}} < m < 2600 + \frac{1,96 \cdot 100}{\sqrt{16}} ,$$
$$2551 < m < 2649 .$$

Błąd względny $\frac{d_{\bar{x}}}{\bar{x}} = \frac{49}{2600} = 1,9\%$.

W długiej serii niezależnych prób w 95 przypadkach na 100 oszacowany przedział będzie pokrywał nieznaną wartość średnią trwałości urzędnika wyprodukowanej partii. Błąd względny tego oszacowania wynosi 1,9%.

Przykład 6.1.2.2

Zakładając, że kwartalne wydatki na reklamę można uznać za cechę o rozkładzie $N(m, \sigma)$, wylosowano do próby 100 zakładów usługowych i otrzymano następujący rozkład wydatków na reklamę:

Kwartalne wydatki w tys. zł	0 – 5	5 – 10	10 – 15	15 – 20
Liczba zakładów	10	20	40	30

Wyznaczyć na poziomie ufności $1 - \alpha = 0,96$ przedział ufności dla przeciętnych kwartalnych wydatków na reklamę. Jaka będzie dokładność oszacowania, gdy poziom ufności będzie równy 0,9.

Rozwiązanie

Obliczamy średnią arytmetyczną ważoną:

$$\bar{x} = \frac{1}{n} \sum_{i=1}^k \dot{x}_i \cdot n_i = \frac{2,5 \cdot 10 + 7,5 \cdot 20 + 12,5 \cdot 40 + 17,5 \cdot 30}{100} = 12.$$

Ponieważ odchylenie standardowe w populacji nie jest znane, przyjmujemy $\sigma \approx S$ i obliczamy odchylenie standardowe z próby:

$$S = \sqrt{\frac{\sum_{i=1}^k (\dot{x}_i - \bar{x})^2 \cdot n_i}{n}} = \sqrt{\frac{(2,5 - 12)^2 \cdot 10 + (7,5 - 12)^2 \cdot 20 + (12,5 - 12)^2 \cdot 40 + (17,5 - 12)^2 \cdot 30}{100}} = 4,72.$$

Wartość parametru $t_\alpha = 2,05$ odczytujemy z tablicy rozkładu normalnego standaryzowanego dla wartości $\Phi(t_\alpha)$ obliczonej z zależności:

$$\Phi(t_\alpha) = \frac{1 - \alpha}{2} = \frac{1 - 0,04}{2} = 0,48.$$

Przedział ufności dla przeciętnych kwartalnych wydatków na reklamę:

$$\begin{aligned} \bar{X} - t_\alpha \frac{\sigma}{\sqrt{n}} < m < \bar{X} + t_\alpha \frac{\sigma}{\sqrt{n}}, \\ 12 - t_\alpha \frac{4,72}{\sqrt{100}} < m < 12 + t_\alpha \frac{4,72}{\sqrt{100}}, \\ 12 - 2,05 \cdot 0,472 < m < 12 + 2,05 \cdot 0,472, \\ 11,032 < m < 12,968. \end{aligned}$$

Możemy stwierdzić z prawdopodobieństwem 0,96, że zakłady przeznaczają kwartalnie na reklamę nie mniej niż 11032 zł i nie więcej niż 12968 zł.

Maksymalny błąd szacunku wynosi:

$$d_{\bar{x}} = t_\alpha \frac{\sigma}{\sqrt{n}} = 2,05 \cdot \frac{4,72}{\sqrt{100}} = 0,968.$$

Wyniki obliczeń przedziału ufności oraz maksymalnego błędu szacunku dla poziomu ufności $1 - \alpha = 0,9$:

$$11,221 < m < 12,779 ,$$

$$d_{\bar{x}} = 0,779 .$$

Dla poziomu ufności 0,9 otrzymano krótszy przedział ufności i mniejszy błąd szacunku, tj. większą dokładność oszacowania. Oznacza to, że im jesteśmy bardziej ufni, tym mniej precyzyjni.

Przykład 6.1.2.3

Poddano badaniom wydatki na odzież w rodzinach pięcioosobowych. Z populacji tych rodzin wylosowano próbę 289-elementową. Na podstawie przeprowadzonych badań ustalono przeciętną skalę wydatków na odzież na poziomie $\bar{X} = 100$ zł. Badania z lat ubiegłych wykazały, że rozkład wydatków na odzież jest rozkładem normalnym o stałej wariancji $\sigma^2 = 576$. Wyznaczyć:

- przedział ufności wydatków na odzież w rodzinach pięcioosobowych, przyjmując poziom ufności $1 - \alpha = 0,98$,
- maksymalny błąd szacunku,
- względny stopień precyzji szacunku parametru m ,
- minimalną licznosc próby dla poziomu ufności 0,98 i 0,95 oraz porównać wyniki.

Rozwiązanie

a) obliczamy przedział ufności dla wartości średniej mając dane: $\bar{X} = 100$, $\sigma = 24$:

$$t_{\alpha} = 2,35 \quad \text{dla } \Phi(t_{\alpha}) = \frac{1 - \alpha}{2} = \frac{0,98}{2} = 0,49 ,$$

$$\bar{X} - t_{\alpha} \frac{\sigma}{\sqrt{n}} < m < \bar{X} + t_{\alpha} \frac{\sigma}{\sqrt{n}} ,$$

$$96,68 < m < 103,32 .$$

Możemy stwierdzić z prawdopodobieństwem 0,98, że pięcioosobowe rodziny przeznaczają nie mniej niż 96,68 zł i nie więcej niż 103,32 zł na zakup odzieży.

b) maksymalny błąd szacunku wynosi:

$$d_{\bar{x}} = t_{\alpha} \frac{\sigma}{\sqrt{n}} = 2,35 \cdot \frac{24}{\sqrt{289}} = 3,32 .$$

Błąd szacunku wartości średniej wynosi 3,32 zł.

c) stopień precyzji szacunku wynosi:

$$V_m = \frac{t_\alpha \cdot \sigma}{\bar{X} \cdot \sqrt{n}} \cdot 100\% = \frac{2,35 \cdot 24}{100 \cdot \sqrt{289}} \cdot 100\% = 3,32\% .$$

d) minimalna liczność próby dla poziomu ufności 0,98 wynosi:

$$n = \frac{t_\alpha^2 \cdot \sigma^2}{d_{\bar{x}}^2} = \frac{2,35^2 \cdot 576}{3,32^2} = 288,6 \approx 289 .$$

Minimalna liczność próby dla poziomu ufności 0,95 wynosi:

$$n = \frac{t_\alpha^2 \cdot \sigma^2}{d_{\bar{x}}^2} = \frac{1,95^2 \cdot 576}{3,32^2} = 198,7 \approx 199 .$$

Przy większym poziomie ufności otrzymujemy większą minimalną liczebność próby.

Przykład 6.1.2.4

Przeprowadzono badanie czasu trwania pewnej reakcji chemicznej. W tym celu wykonano 10 niezależnych doświadczeń i otrzymano wyniki 8, 14, 10, 12, 7, 13, 11, 12 10, 9 (w sekundach). Zakładając, że w określonych warunkach badany czas jest zmienną losową o rozkładzie normalnym

- oszacować przedziałowo średni czas trwania badanej reakcji, przyjmując współczynnik ufności na poziomie 0,95,
- ustalić, jak zmieni się precyzja oszacowania średniej, jeśli wielkość próby zwiększymy czterokrotnie?

Rozwiązanie

$$\bar{x} = \frac{1}{n} \sum_{i=1}^n x_i = \frac{1}{10} \cdot 106 = 10,6 \text{ sek},$$

$$S^2 = \frac{1}{n} \sum_{i=1}^k (x_i - \bar{x})^2 = \frac{1}{10} \cdot 44,4 = 4,44 ,$$

$$S = \sqrt{S^2} = 2,12 \text{ sek}.$$

Tablica obliczeniowa

x_i	$x_i - \bar{x}$	$(x_i - \bar{x})^2$
8	2,6	6,76
14	3,4	11,56
10	-0,6	0,36
12	1,4	1,96
7	-3,6	12,96
13	2,4	5,76
11	0,4	0,16
12	1,4	1,96
10	-0,6	0,36
9	-1,6	2,56
106	Σ	44,40

a) ponieważ próba jest mała $n = 10$ przedział ufności obliczamy ze wzoru:

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

gdzie: $t_{\alpha, n-1}$ – wartość zmiennej losowej t-Studenta, którą odczytuje się z tablicy tego rozkładu dla $n-1$ stopni swobody i dla danego z góry prawdopodobieństwa $1 - \alpha$, tak aby spełniona była relacja:

$$P\{-t_{\alpha, n-1} < t < t_{\alpha, n-1}\} = 1 - \alpha.$$

Dla $1 - \alpha = 0,95$ oraz $n - 1 = 9$ stopni swobody statystyka t_{α} jest równa 2,262. Korzystając z obliczeń, tzn. przyjmując wartość średnią i odchylenie standardowe z próby możemy wyznaczyć przedział ufności dla średniego czasu trwania badanej reakcji.

$$10,6 - 2,262 \cdot \frac{2,12}{\sqrt{10}} < m < 10,6 + 2,262 \cdot \frac{2,12}{\sqrt{10}},$$

$$9,1 < m < 12,1.$$

Średni czas trwania reakcji jest nie mniejszy niż 9,1 sekundy i nie większy niż 12,1 sekundy, co możemy zapisać przedziałowo: [9,1 s; 12,1 s].

b) aby ustalić precyzję oszacowania średniego czasu reakcji zastosujemy wzór:

$$V_m = \frac{t_{\alpha} \cdot \sigma}{\bar{X} \cdot \sqrt{n}} \cdot 100\%,$$

przyjmując $s \approx \sigma$, $\bar{x} = \bar{X}$.

Dla próby 10-elementowej

$$V_m = \frac{2,262 \cdot 2,12}{10,6 \cdot 3,16} = 0,143 \text{ (14,3\%)}$$

Dla próby czterokrotnie większej (40-elementowej)

$$V_m = \frac{2,262 \cdot 2,12}{10,6 \cdot 6,32} = 0,072 \text{ (7,2\%)}$$

Gdy zwiększymy próbę czterokrotnie, to precyzja oszacowania wzrośnie dwukrotnie.

Przedział ufności dla wariancji i odchylenia standardowego

Przedział ufności dla wariancji możemy wyznaczyć wtedy, gdy cecha X charakteryzująca populację ma rozkład $N(m, \sigma)$ lub zbliżony do rozkładu normalnego (gdy liczebność próby jest duża). W zależności od tego, z jak liczną próbą mamy do czynienia oraz, jaki estymator wariancji z próby zastosujemy do budowy przedziału ufności, przedział ten ma postać:

$$\frac{nS^2}{\chi_2^2} < \sigma^2 < \frac{nS^2}{\chi_1^2},$$

gdzie: χ_1^2 oraz χ_2^2 odczytujemy z tablic rozkładu chi-kwadrat dla n stopni swobody (lub dla $n - 1$ stopni swobody, por. rozdz. V) w ten sposób, że dla danego poziomu ufności $1 - \alpha$ zachodzą równości:

$$P(\chi^2 > \chi_2^2) = \frac{\alpha}{2}$$

oraz

$$P(\chi^2 > \chi_1^2) = 1 - \frac{\alpha}{2}.$$

Przykład 6.1.2.6

Czas pracy żarówek produkowanych w firmie POLAM ma rozkład $N(750, \sigma)$. Na podstawie 16-elementowej próby losowej obliczono $S^2 = 2500$. Wyznaczyć 98-procentowy przedział ufności dla wariancji i odchylenia standardowego.

Rozwiązanie

Korzystamy z zależności na przedział ufności dla wariancji, który ma zastosowania dla małych prób.

Z tablic rozkładu χ^2 dla 16 stopni swobody odczytujemy wartości $\chi_1^2 = 5,812$ oraz $\chi_2^2 = 32,000$ spełniające warunki:

$$P(\chi^2 > \chi_2^2) = \frac{\alpha}{2} = \frac{0,02}{2} = 0,01,$$

$$P(\chi^2 > \chi_1^2) = 1 - \frac{\alpha}{2} = 1 - \frac{0,02}{2} = 0,99.$$

Przedział ufności dla wariancji wynosi:

$$\frac{nS^2}{\chi_2^2} < \sigma^2 < \frac{nS^2}{\chi_1^2},$$

$$\frac{16 \cdot 2500}{32} < \sigma^2 < \frac{16 \cdot 2500}{5,812},$$

$$1250 < \sigma^2 < 6882,3.$$

Stąd przedział ufności dla odchylenia standardowego dla $n < 30$ wynosi:

$$35,36 < \sigma < 82,96.$$

Można przyjąć z prawdopodobieństwem 0,98, że odchylenie standardowe nie będzie mniejsze od 35,36 oraz nie będzie większe od 82,96 jednostek czasu pracy.

Przykład 6.1.2.7

W zakładzie zatrudniającym 1000 robotników wylosowano 20 osób, dla których określono odchylenie standardowe $S = 6,4\%$ dla wykonania normy przez tych pracowników. Zbudować przedział ufności pokrywający z prawdopodobieństwem 0,95 nieznaną wartość wariancji dla wszystkich robotników.

Rozwiązanie

Z tablic rozkładu χ^2 dla dziewiętnastu stopni swobody odczytujemy wartości $\chi_1^2 = 8,907$ oraz $\chi_2^2 = 32,852$ spełniające warunki:

$$P(\chi^2 > \chi_2^2) = \frac{\alpha}{2} = \frac{0,05}{2} = 0,025,$$

$$P(\chi^2 > \chi_1^2) = 1 - \frac{\alpha}{2} = 1 - \frac{0,05}{2} = 0,975.$$

Przedział ufności dla wariancji wynosi:

$$\frac{(n-1)S^2}{\chi_2^2} < \sigma^2 < \frac{(n-1)S^2}{\chi_1^2},$$
$$\frac{(20-1) \cdot 6,4}{32,852} < \sigma^2 < \frac{(20-1) \cdot 6,4}{8,907},$$
$$3,70 < \sigma^2 < 13,65.$$

Jeżeli cecha X w zbiorowości generalnej ma rozkład normalny $N(m, \sigma)$ lub jest zbliżony do normalnego i liczność próby $n > 30$, to przedział ufności dla odchylenia standardowego σ może mieć postać:

$$\frac{S}{1 + \frac{t_\alpha}{\sqrt{2n}}} < \sigma < \frac{S}{1 - \frac{t_\alpha}{\sqrt{2n}}}.$$

Przykład 6.1.2.8

Pewna firma dysponuje w Polsce siecią 300 sklepów. W badaniach utargu ze sprzedaży tej firmy pobrano próbę losową złożoną z 40 sklepów. Na podstawie próby wyznaczono wariancję dziennego utargu $S^2 = 90\,000$ zł. Ustalić przedział ufności dla odchylenia standardowego utargu w sklepach tej firmy przy współczynniku ufności 0,99.

Rozwiązanie

Z tablic rozkładu normalnego odczytujemy wartość $t_\alpha = 2,60$ dla $\alpha = 0,01$ oraz

$$\Phi(t_\alpha) = \frac{1 - \alpha}{2} = \frac{0,99}{2} = 0,495.$$

Wyznaczamy przedział ufności dla odchylenia standardowego:

$$\frac{S}{1 + \frac{t_\alpha}{\sqrt{2n}}} < \sigma < \frac{S}{1 - \frac{t_\alpha}{\sqrt{2n}}},$$
$$\frac{\sqrt{90000}}{1 + \frac{2,60}{\sqrt{2 \cdot 40}}} < \sigma < \frac{\sqrt{90000}}{1 - \frac{2,60}{\sqrt{2 \cdot 40}}},$$
$$232,4 < \sigma < 422,9.$$

Można przyjąć z prawdopodobieństwem 0,99, że odchylenie standardowe nie będzie mniejsze od 232,4 oraz nie będzie większe od 422,9 złotych.

Przedział ufności dla wskaźnika struktury

W badaniach statystycznych pojawia się problem oszacowania prawdopodobieństwa wystąpienia wariantu cechy zwanego sukcesem albo oznaczenia, jaki procent zbiorowości posiada wyróżnioną cechę. Jest to szczególnie ważne, gdy cecha opisująca zbiorowość jest cechą niemierzalną i podstawową charakterystyką populacji jest frakcja (procent) wyróżnionych elementów, zwane też wskaźnikiem struktury w populacji. Podstawą konstrukcji przedziału ufności dla prawdopodobieństwa sukcesu p jest częstość występowania tego sukcesu, czyli $\frac{X}{n}$. Przedział ten ma zastosowanie dla dużych prób $n > 100$.

$$\frac{X}{n} - t_{\alpha} \sqrt{\frac{\frac{X}{n} \left(1 - \frac{X}{n}\right)}{n}} < p < \frac{X}{n} + t_{\alpha} \sqrt{\frac{\frac{X}{n} \left(1 - \frac{X}{n}\right)}{n}} .$$

Przykład 6.1.2.9

Badamy niezawodność pewnego radaru. Na 400 obrotów anteny radarowej obiekt znajdujący się w obszarze obserwacji radaru został wykryty 350 razy. Niech p oznacza prawdopodobieństwo wykrycia obiektu przy jednym obrocie anteny (niezawodność radaru) Wyznaczyć przedział ufności dla prawdopodobieństwa p , przyjmując współczynnik ufności na poziomie $1 - \alpha = 0,95$.

Rozwiązanie

Niech X oznacza zmienną losową przyjmującą wartość 1, gdy w jednym obrocie anteny obiekt został wykryty, zaś wartość 0, gdy nie został wykryty. Zmienna losowa X ma rozkład zero-jedynkowy z parametrem p . Prawdopodobieństwo p oszacujemy przedziałem ufności dla wskaźnika struktury, który jest wskaźnikiem struktury w próbie. Z danych zadania wynika, że $X/n = 350/400 = 0,875$ (87,5%).

$\Phi(t_{\alpha}) = 1 - \alpha/2 = 0,975$ stąd $t_{\alpha} = 1,96$.

$$0,875 - 1,96 \sqrt{\frac{0,875 \cdot 0,125}{400}} < p < 0,875 + 1,96 \sqrt{\frac{0,875 \cdot 0,125}{400}} .$$

A zatem

$$0,875 - 0,032 < p < 0,875 + 0,032.$$

Ostatecznie otrzymujemy przedział ufności dla prawdopodobieństwa p

$$0,843 \text{ (84,3\%)} < p < 0,907 \text{ (90,7\%)}.$$

Niezawodność radaru z ufnością 0,95 zawiera się w przedziale [84,3%; 90,7%].

Przykład 6.1.2.10

Na pewnym przejściu granicznym przebadano wybraną losowo próbę osób, które wyjeżdżały za granicę i zapytano o cel wyjazdu. Otrzymano następujące wyniki:

Cel wyjazdu	Liczba osób
Owiedziny	40
Wycieczka	60
Zakupy	80
Praca	20

Oszacować metodą przedziałową na poziomie ufności 0,90 procent osób, które udają się po zakupy.

Rozwiązanie

Prawdopodobieństwo, że wśród osób przekraczających granicę znajdują się osoby udające się po zakupy wynosi:

$$p = \frac{X}{n} = \frac{80}{200} = 0,4,$$

gdzie: X – liczba osób udających się po zakupy,
 n – licznosc próby badanych osób.

Wartość parametru $t_{\alpha} = 1,64$ odczytujemy z tablic rozkładu normalnego dla

$$\Phi(t_{\alpha}) = \frac{1 - \alpha}{2} = \frac{0,90}{2} = 0,45.$$

Przedział ufności dla wskaźnika struktury:

$$\frac{X}{n} - t_{\alpha} \sqrt{\frac{\frac{X}{n}(1 - \frac{X}{n})}{n}} < p < \frac{X}{n} + t_{\alpha} \sqrt{\frac{\frac{X}{n}(1 - \frac{X}{n})}{n}},$$
$$\frac{80}{200} - 1,64 \sqrt{\frac{\frac{80}{200}(1 - \frac{80}{200})}{200}} < p < \frac{80}{200} + 1,64 \sqrt{\frac{\frac{80}{200}(1 - \frac{80}{200})}{200}},$$
$$0,3432 < p < 0,4568,$$
$$34,32\% < p < 45,68\%.$$

Możemy stwierdzić z prawdopodobieństwem 0,90, że odsetek osób przekraczających granicę w celu dokonania zakupów nie jest mniejszy niż 34,32% i nie większy niż 45,68% (czyli od 34,32% do 45,68% osób udaje się po zakupy).

Przykład 6.1.2.11

Przedsiębiorstwo międzynarodowe, prowadzące w Polsce działalność postanowiło określić frakcję pracowników, którzy podwyższają swe kwalifikacje. W tym celu należało ustalić minimalną liczebność próby. Z poprzedniego badania wiadomo, że frakcja ta wynosiła 20%, a średni błąd szacunku 6,7%. Wykonać obliczenia przyjmując poziom ufności 0,90.

Rozwiązanie

Populacja ma rozkład dwupunktowy z parametrem p . Wykorzystując wzór na błąd

$$d_{\bar{x}} = t_{\alpha} \sqrt{\frac{p(1-p)}{n}},$$

po wykonaniu odpowiednich przekształceń otrzymujemy:

$$n = \frac{t_{\alpha}^2 \cdot pq}{d_{\bar{x}}^2}.$$

$$\text{A zatem } n = \frac{(1,64)^2 \cdot 0,20 \cdot 0,80}{(0,067)^2} = 95,62 \approx 96.$$

6.2. Powtórzenie w przykładach

Zadanie 1

Wyznaczyć minimalną liczebność próby dla oszacowania średniego wzrostu chłopców w wieku 12 lat, przy współczynniku ufności 0,95 oraz dopuszczalnym błędzie szacunku 5 cm, jeżeli w próbie wstępnej liczącej 10 osób otrzymano następujące wyniki (w cm): 125, 150, 145, 130, 155, 140, 160, 125, 155, 135.

Rozwiązanie

Obliczamy wartość średnią oraz wariancję

$$\bar{x} = \frac{1}{n} \sum_{i=1}^{10} x_i = \frac{125 + 150 + 145 + 130 + 155 + 140 + 160 + 125 + 155 + 135}{10} = 142,$$

$$\begin{aligned}\sigma^2 &= \frac{1}{n} \sum_{i=1}^n (x_i - \bar{x})^2 = \\ &= \frac{(125-142)^2 + (150-142)^2 + \dots + (155-142)^2 + (135-142)^2}{10} = 151.\end{aligned}$$

Z tablic rozkładu t-Studenta dla $\alpha = 0,05$ i $n - 1 = 9$ stopni swobody odczytujemy wartość parametru $t_{\alpha, n-1} = 2,262$.

Wartość błędu szacunku jest dana i wynosi $d_{\bar{x}} = 5$ cm.

$$n = \frac{t_{\alpha, n-1}^2 \cdot \sigma^2}{d_{\bar{x}}^2} = \frac{2,262^2 \cdot 151}{5^2} = 30,9 \approx 31.$$

Minimalna liczność próby dla oszacowania średniego wzrostu chłopców przy współczynniku ufności wynoszącym 0,95 oraz błędzie szacunku wynoszącym 5 cm powinna wynosić 31 osób.

Zadanie 2

W przedsiębiorstwie budowlanym wylosowano 16 osób w celu ustalenia średniej płacy robotników o stażu pracy do jednego roku. Średnia płaca wynosiła 2140 zł, a odchylenie standardowe S było równe 220 zł. Zakładając poziom ufności równy 0,95 zbudować przedział ufności pokrywający nieznaną średnią wartość płacy całego przedsiębiorstwa.

Rozwiązanie

Ponieważ próba jest mała $n = 16$, przedział ufności obliczamy ze wzoru:

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

gdzie: $t_{\alpha, n-1}$ – wartość zmiennej losowej t-Studenta, którą odczytuje się z tablicy tego rozkładu dla $n-1$ stopni swobody i dla danego z góry prawdopodobieństwa $1 - \alpha$, tak aby spełniona była relacja:

$$P\{-t_{\alpha, n-1} < t < t_{\alpha, n-1}\} = 1 - \alpha.$$

Dla $1 - \alpha = 0,95$ oraz $n-1 = 15$ stopni swobody statystyka t_{α} jest równa 2,131. Podstawiając dane otrzymujemy szukany przedział ufności dla wartości średniej płacy:

$$2140 - 2,131 \cdot \frac{220}{\sqrt{16}} < m < 2140 + 2,131 \cdot \frac{220}{\sqrt{16}},$$

$$2023 < m < 2257.$$

Średnia płaca w przedsiębiorstwie jest nie mniejsza niż 2023 zł i nie większa niż 2257 zł.

Zadanie 3

W pewnym przedsiębiorstwie zbadano staż pracy pracowników umysłowych. W tym celu wylosowano niezależnie 196-elementową próbę i otrzymano średni staż pracy 6,9 lat oraz odchylenie standardowe 2,8 lat. Oszacować za pomocą przedziału ufności na poziomie $1 - \alpha = 0,90$ średni staż pracy tej grupy pracowników. Co należy uczynić, aby zwiększyć precyzję oszacowania?

Rozwiązanie

Z tablic dystrybuanty rozkładu normalnego $N(0,1)$ odczytujemy, że $t_\alpha = 1,64$ (odczyt dla $1 - \alpha/2 = 0,995$). Podstawiając do odpowiedniego wzoru otrzymujemy przedział ufności

$$6,9 - 1,64 < m < 6,9 + 1,64,$$

czyli

$$6,57 < m < 7,23.$$

Przedział liczbowy (6,57; 7,23) z prawdopodobieństwem 0,90 pokrywa prawdziwy, średni staż pracy pracowników w tym przedsiębiorstwie. Tylko w 10 przypadkach na 100 przedział ten nie pokryje szacowanego parametru. Aby zwiększyć precyzję oszacowania należy zwiększyć liczebność próby albo zwiększyć poziom ufności. Przyjmując np. $1 - \alpha = 0,95$ przy tej samej liczebności próby $n = 196$ otrzymamy następujący przedział ufności

$$6,9 - 1,96 < m < 6,9 + 1,96.$$

A zatem

$$6,51 < m < 7,29.$$

Porównując zbudowane przedziały ufności dla $1 - \alpha = 0,90$ oraz $1 - \alpha = 0,95$ zauważamy, że w miarę wzrostu wartości współczynnika ufności (tzn. im bliższy jest on jedności) długość przedziału ufności wzrasta, a tym samym zwiększa się dokładność (pewność) szacunku, tzn. prawdopodobieństwo pokrycia parametru danym przedziałem.

Zadanie 4

Spośród populacji studentów Radomia wylosowano 400 osób, którym zadano pytanie czy palą papierosy? Stwierdzono, że 160 osób spośród badanych sporadycznie bądź stale pali papierosy. Przyjmując współczynnik ufności na poziomie 0,95 zbudować przedział ufności dla nieznannej frakcji palących.

Rozwiązanie

Prawdopodobieństwo, że wśród wylosowanych studentów trafimy na osobę palącą wynosi:

$$p = \frac{X}{n} = \frac{160}{400} = 0,4,$$

gdzie: X – liczba osób niepalących,
 n – licznosc próby wylosowanych studentów.

Wartość parametru t_α wynosi 1,95 dla

$$\Phi(t_\alpha) = \frac{1 - \alpha}{2} = \frac{0,95}{2} = 0,475.$$

Przedział ufności dla wskaźnika struktury:

$$\frac{X}{n} - t_\alpha \sqrt{\frac{\frac{X}{n}(1 - \frac{X}{n})}{n}} < p < \frac{X}{n} + t_\alpha \sqrt{\frac{\frac{X}{n}(1 - \frac{X}{n})}{n}},$$
$$0,352 < p < 0,448.$$

Możemy stwierdzić z prawdopodobieństwem 0,95, że odsetek palących studentów w Radomiu nie jest mniejszy niż 35,2% i nie większy niż 44,8% (czyli od 35,2% do 44,8% studentów pali papierosy).

Zadanie 5

Magazyn „Fortune” corocznie publikuje listę 500 największych firm usługowych w USA, należących do sześciu wielkich grup branżowych - *Fortune Service 500*. Grupy te oraz liczby firm należących do każdej z nich podaje tabela 1 [za: 1].

Tabela 1. Lista Fortune Service 500

Grupa	Liczba firm
Usługi zróżnicowane	100
Bankowość komercyjna	100
Usługi finansowe (włączając lokaty i ubezpieczenia)	150
Handel detaliczny	50
Transport	50
Energetyka	50
Razem	500

Zgromadzić próbę losową firm, obliczyć średnią wartość dochodu netto (w mln \$) dla firm w próbie i na tej podstawie estymować wartość średnią dochodu netto dla całej populacji 500 firm. Wyznaczyć przedział ufności dla wartości średniej dochodu netto wszystkich firm, przyjmując poziom 95% pewności.

Rozwiązanie

Z analizy wstępnej wynika, że firmy należące do tej samej grupy branżowej mają podobne dochody netto, dlatego można potraktować te sześć grup jako warstwy i losować próbę z każdej grupy. Na podstawie danych z tabeli 1 można obliczyć wagi dla każdej z warstw, stosując definicję: $W_i = N_i/N$. Otrzymane wyniki są następujące:

$$W_1 = N_1/N = 100/500 = 0,2,$$

$$W_2 = N_2/N = 100/500 = 0,2,$$

$$W_3 = N_3/N = 150/500 = 0,3,$$

$$W_4 = N_4/N = 50/500 = 0,1,$$

$$W_5 = N_5/N = 50/500 = 0,1,$$

$$W_6 = N_6/N = 50/500 = 0,1.$$

Decydujemy się na wylosowanie próby 100 firm spośród 500 wybierając proporcjonalną alokację. W przypadku proporcjonalnej alokacji cała 100-elementowa próba musi być rozdzielona między warstwy zgodnie z proporcją określoną przez wagi poszczególnych warstw. Dla każdej warstwy (dla każdego $i = 1, 2, \dots, 6$) obliczamy wartość $n_i = n \cdot W_i$, wyznaczając w ten sposób następujące rozmiary prób:

$$n_1 = 20, \quad n_2 = 20, \quad n_3 = 30, \quad n_4 = 10, \quad n_5 = 10, \quad n_6 = 10.$$

Zakładamy, że wartości dochodu netto w każdej warstwie mają w przybliżeniu rozkład normalny oraz estymowana wariancja w każdej warstwie (która będzie estymowana z danych) jest równa rzeczywistej wartości wariancji w warstwie tak, że może zostać zastosowany rozkład normalny. Dokonujemy losowania prób, a następnie obliczamy wartości średnie i wariancje oraz zestawiamy dotychczas uzyskane wyniki w tabeli 2.

Tabela 2. Rezultat pobrania próby

Warstwa	Wartość średnia \bar{x}_i	Wariancja s_i^2	n_i	W_i
1	52,7	97 650	20	0,2
2	112,6	64 300	20	0,2
3	85,6	76 990	30	0,3
4	12,6	18 320	10	0,1
5	8,9	9 037	10	0,1
6	52,3	83 500	10	0,1

Estymator wartości średniej populacji w losowaniu warstwowym obliczymy z następującego wzoru:

$$\bar{x}_{st} = \sum_{i=1}^6 W_i \bar{x}_i = 0,2 \cdot 52,7 + 0,2 \cdot 112,6 + 0,3 \cdot 85,6 + 0,1 \cdot 12,6 + 0,1 \cdot 8,9 + 0,1 \cdot 52,3 \\ = 66,12,$$

gdzie: \bar{x}_{st} = warstwowa wartość średnia w próbie.

Natomiast estymowane odchylenie warstwowe wartości średniej w próbie wyznaczamy ze wzoru:

$$s(\bar{X}_{st}) = \sqrt{\frac{1-f}{n} \sum_{i=1}^6 W_i s_i^2} = \sqrt{\frac{0,8}{100} \cdot (0,2 \cdot 97\,650 + 0,2 \cdot 64\,300 + 0,3 \cdot 76\,990 + \\ 0,1 \cdot 18\,320 + 0,1 \cdot 9037 + 0,1 \cdot 83\,500)} = 23,8,$$

gdzie: frakcja z próby $f = 100/500 = 0,2$.

Wartość nieobciążonego estymatora średniego dochodu netto dla wszystkich firm w wykazie *Fortune Service 500* wynosi 66,12 mln \$.

Przedział ufności dla wartości średniej w losowaniu warstwowym obliczamy według wzoru $\bar{x}_{st} \pm t_{\alpha} s(\bar{X}_{st})$.

Przedział ufności dla wartości średniej dochodu netto wszystkich firm w wykazie wyznaczamy stosując powyższy wzór następująco:

$$66,12 - 1,96 \cdot 23,08 < \bar{x}_{st} < 66,12 + 1,96 \cdot 23,08.$$

Średni dochód netto w mln \$ firm w wykazie *Fortune Service 500* mieści się w przedziale [20,88 mln \$, 111,36 mln \$] z pewnością na poziomie 95%.

Zadanie 6

W pewnej małej miejscowości o liczbie ludności 10 000 mieszkańców należy oszacować procent mieszkańców chorych na astmę. Wylosowano zależnie próbę 800 osób i poddano badaniom lekarskim. Okazało się, że 56 osób choruje na astmę. Przyjmując współczynnik ufności 0,99 zbudować przedział ufności dla szacowanego procentu chorych na astmę wśród mieszkańców tej miejscowości.

Rozwiązanie

Ponieważ populacja generalna jest duża, a losowanie próby nieograniczone zależne, to do obliczenia przedziału ufności dla wskaźnika struktury zastosujemy następujący wzór:

$$\frac{m}{n} - t_{\alpha} \sqrt{\left(\frac{1}{n} - \frac{1}{N}\right) \cdot \frac{m}{n} \cdot \left(1 - \frac{m}{n}\right)} < p < \frac{m}{n} + t_{\alpha} \sqrt{\left(\frac{1}{n} - \frac{1}{N}\right) \cdot \frac{m}{n} \cdot \left(1 - \frac{m}{n}\right)},$$

gdzie: $\frac{m}{n}$ – częstość występowania elementów wyróżnionych w próbie,
N – liczebność populacji,

n – liczebność próby,

t_α – wartość odczytana z tablic rozkładu normalnego $N(0,1)$ dla założonego z góry współczynnika ufności $1 - \alpha$.

Z warunków zadania mamy:

$$\frac{m}{n} = \frac{56}{800} = 0,07 \quad 1 - \frac{m}{n} = 0,93 \quad \frac{m}{n} \left(1 - \frac{m}{n}\right) = 0,0651,$$

$$\frac{1}{n} - \frac{1}{N} = \frac{1}{800} - \frac{1}{10000} = 0,00115 \quad t_\alpha = 2,58.$$

Otrzymujemy zatem przedział ufności dla frakcji p

$$0,07 - 2,58\sqrt{0,00115 \cdot 0,0651} < p < 0,07 + 2,58\sqrt{0,00115 \cdot 0,0651},$$

stąd

$$0,048 < p < 0,092.$$

Przedział ufności dla szacowanego parametru chorych na astmę ma postać:
 $4,8\% < p\% < 9,2\%$.

ROZDZIAŁ VII

WERYFIKACJA HIPOTEZ STATYSTYCZNYCH

Weryfikacja (testowanie) hipotez statystycznych polega na sprawdzaniu określonych przypuszczeń (założeń) wysuniętych wobec parametrów lub rozkładów populacji generalnej na podstawie wyników z próby.

Hipotezę statystyczną nazywamy każdy osąd (przypuszczenie, założenie) dotyczący populacji generalnej wydany na podstawie wyników badania jej części (próby).

Hipoteza może dotyczyć:

- wartości parametrów rozkładu (hipotezy parametryczne), np. „wariancje dwóch populacji o rozkładzie normalnym są sobie równe”,
- postaci funkcyjnej rozkładu populacji (hipotezy nieparametryczne), np. „populacja generalna ma rozkład Poissona”.

Jeżeli powiemy, że „w roku 2030 będzie powódź”, to nie jest to hipoteza statystyczna, ponieważ nie ma tu mowy o postaci rozkładu ani o jego parametrach.

Jeżeli wiedza, jaką mamy o populacji generalnej ogranicza zbiór możliwych (przypuszczeń) hipotez co do rozkładów tej populacji, to zbiór taki nazywa się *zbiorem hipotez dopuszczalnych*. Na przykład, jeżeli wiemy, że badana zmienna losowa X w populacji generalnej podlega rozkładowi normalnemu, to zbiór hipotez dopuszczalnych zawiera wszystkie rozkłady normalne różniące się między sobą wartościami parametrów. Sprawdzanie w tej sytuacji, że populacja ma inny rozkład, np. dwumianowy jest bezpodstawne.

Jeżeli hipoteza jednoznacznie specyfikuje rozkład populacji generalnej, to nazywamy ją *hipotezą prostą*. Każda hipoteza, która nie jest prostą, nazywa się *hipotezą złożoną*.

Hipotezy ze zbioru dopuszczalnego dzielą się na proste i złożone. Na przykład hipoteza: parametr λ w rozkładzie Poissona jest równy 2 jest hipotezą parametryczną prostą; hipoteza: wariancja w rozkładzie normalnym jest większa od 3 jest hipotezą parametryczną złożoną (nie ma mowy o wartości średniej m).

Hipotezę zerową H_0 nazywamy hipotezę sprawdzaną (testowaną, weryfikowaną).

Hipotezę alternatywną H_1 nazywamy hipotezę, którą jesteśmy skłonni przyjąć, gdy odrzucamy hipotezę H_0 .

Test statystyczny jest to reguła postępowania, która przyporządkowuje wynikom próby losowej decyzję przyjęcia lub odrzucenia hipotezy H_0 .

Błąd I rodzaju polega na odrzuceniu hipotezy H_0 mimo, że jest ona prawdziwa.

Poziomem istotności α nazywamy prawdopodobieństwo popełnienia błędu I rodzaju. Wartości α są bliskie zera i na ogół są równe 0,01, 0,02, 0,05, 0,1.

Błąd II-go rodzaju polega na przyjęciu hipotezy H_0 , gdy jest ona fałszywa. Prawdopodobieństwo popełnienia błędu II rodzaju oznaczamy literą β . Dobry test statystyczny powinien mieć tę własność, że β jest bliskie zera.

W statystycznej kontroli jakości α jest określane jako *ryzyko producenta*, β zaś jako ryzyko odbiorcy. Wartości α i β są wzajemnie powiązane i zmniejszenie jednej z nich powoduje zwiększenie drugiej.

Test istotności to test, w którym bierzemy pod uwagę jedynie prawdopodobieństwo popełnienia błędu I rodzaju. W testach istotności pomija się kwestię prawdopodobieństwa popełnienia błędu II rodzaju. W teście istotności nie podejmuje się decyzji o przyjęciu sprawdzanej hipotezy H_0 , gdyż wtedy narażamy się na błąd II rodzaju, a przecież jego prawdopodobieństwo w tym teście nie jest brane pod uwagę. Test sprawdzanej hipotezy H_0 pozwala jedynie na ewentualne odrzucenie sprawdzanej hipotezy H_0 na poziomie istotności α , bądź też na stwierdzenie, że nie ma podstaw do jej odrzucenia.

Sprawdzaniem hipotezy nazywamy taką statystykę T , o znanym rozkładzie, której wartość obliczana na podstawie próby losowej pozwala na podjęcie decyzji, czy odrzucić hipotezę H_0 .

Dla hipotez parametrycznych sprawdzianami są *estymatory* odpowiednich parametrów, natomiast dla hipotez nieparametrycznych rolę sprawdzianów pełnią *mierniki rozbieżności* między rozkładem empirycznym a teoretycznym sformułowanym w hipotezie H_0 .

Zbiorem krytycznym Z (obszarem krytycznym) nazywamy zbiór tych wartości sprawdzianu hipotezy (statystyki T), które przemawiają za odrzuceniem hipotezy H_0 .

Zbiór krytyczny Z może być, w zależności od postaci hipotezy alternatywnej, zbiorem jednostronnym (prawo- lub lewostronnym) albo zbiorem dwustronnym. Rozkład sprawdzianu hipotezy określa, z jakich tablic należy korzystać przy odczytywaniu wartości krytycznych wyznaczających zbiór Z .

7.1. Hipotezy parametryczne

Hipotezy parametryczne są formułowane w odniesieniu do parametrów rozkładu przy założeniu, że postać rozkładu jest znana.

Etapy budowy testów istotności dla weryfikacji hipotezy H_0 .

Sformułowanie hipotezy H_0 i alternatywnej H_1 .

W testach istotności hipoteza H_0 jest hipotezą „o równości”

$$H_0 : m = m_0 .$$

Hipoteza alternatywna H_1 , w zależności od celu badania, może stanowić zaprzeczenie hipotezy H_0

$$H_1 : m \neq m_0 ,$$

może być hipotezą „o większości”

$$H_1 : m > m_0$$

lub hipotezą „o mniejszości”

$$H_1 : m < m_0 ,$$

Na przykład formułując hipotezę H_0 , że wartość średnia m_0 obliczona na podstawie wyników badań reprezentatywnej próby jest równa wartości średniej m w populacji można postawić następujące trzy hipotezy alternatywne H_1 :

- wartość średnia z próby i średnia w populacji nie są równe,
- wartość średnia z próby jest większa od wartości średniej w populacji,
- wartość średnia z próby jest mniejsza od wartości średniej w populacji.

Przyjęcie poziomu istotności α .

Poziom istotności α przyjmowany jest przez badacza lub dany w zadaniu. Im wyższy jest poziom istotności α , tym większa jest szansa odrzucenia hipotezy H_0 .

Wylosowanie n -elementowej próby prostej i wyznaczenie z niej takiej statystyki Z_n , której rozkład (dokładny lub graniczny) jest znany, przy założeniu prawdziwości sprawdzanej hipotezy H_0 .

Próba prosta złożona z n elementów i wylosowana ze skończonej lub nieskończonej populacji jest to próba losowa, której wyniki są niezależnymi zmiennymi losowymi, posiadającymi jednakowe rozkłady, takie same jak rozkład populacji.

Wybór sprawdzianu hipotezy.

Zakładamy, że rozkład cechy w zbiorowości generalnej jest $N(m, \sigma)$.

Wybór sprawdzianu hipotezy zależy od liczebności próby n oraz tego, czy odchylenie standardowe σ w zbiorowości generalnej jest znane.

Mamy następujące przypadki:

- a) σ jest znane oraz $n \leq 30$,
- b) σ jest znane oraz $n > 30$,
- c) σ jest nieznanne oraz $n > 30$, ale wówczas przyjmujemy $\sigma \approx S$ i sprawdzianem hipotezy $H_0 : m = m_0$ jest statystyka:

$$T = \frac{\bar{x} - m}{\sigma} \cdot \sqrt{n}$$

o rozkładzie $N(0,1)$.

Ze statystyki $T = \frac{\bar{x} - m}{\sigma} \cdot \sqrt{n}$ można korzystać również wtedy, gdy rozkład w zbiorowości jest dowolny, ale mamy do dyspozycji próbę o liczebności $n > 30$.

Jeżeli σ jest nieznane oraz $n \leq 30$, to sprawdzianem hipotezy $H_0 : m = m_0$ jest statystyka:

$$T = \frac{\bar{x} - m}{\sigma} \cdot \sqrt{n - 1}$$

o rozkładzie t-Studenta z $n - 1$ stopniami swobody.

Wybór obszaru krytycznego.

Obszar krytyczny zależy od postaci hipotezy alternatywnej H_1 . Dla hipotezy alternatywnej w postaci $H_1 : m \neq m_0$ obszar krytyczny jest dwustronny:

$$P\{|T| > t_\alpha\} = \alpha.$$

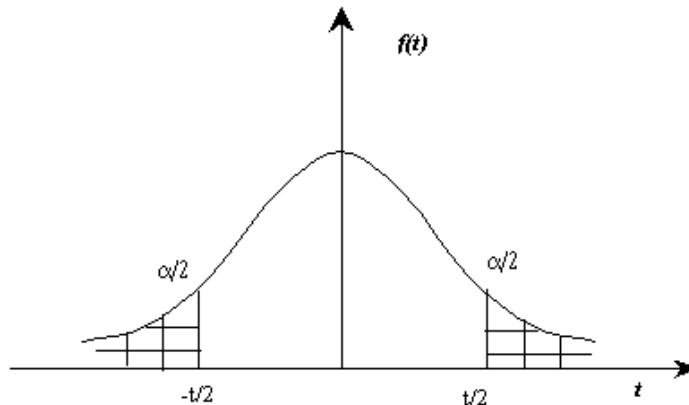
Jeżeli hipoteza alternatywna ma postać $H_1 : m < m_0$ obszar krytyczny jest lewostronny:

$$P\{T < -t_\alpha\} = \alpha.$$

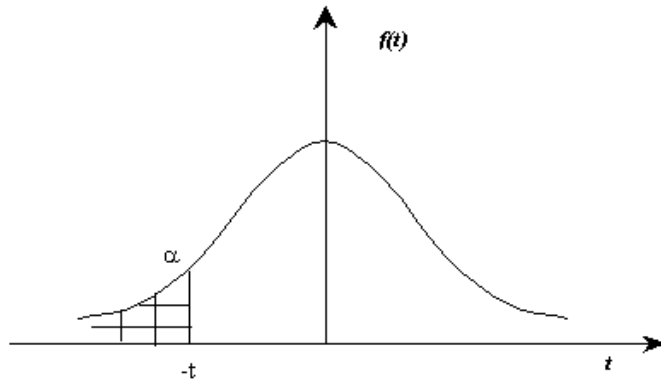
Jeżeli hipoteza alternatywna ma postać $H_1 : m > m_0$ obszar krytyczny jest prawostronny:

$$P\{T > t_\alpha\} = \alpha.$$

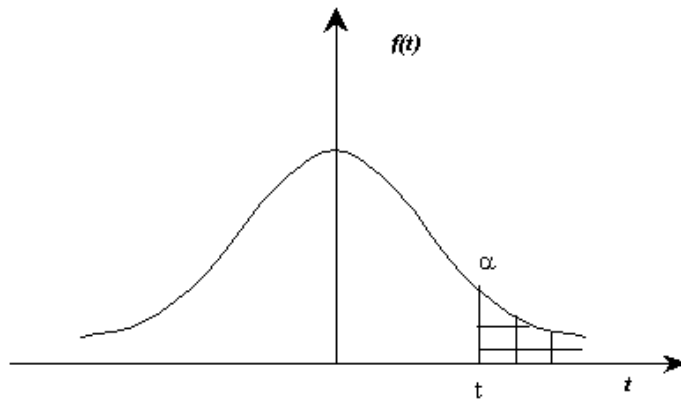
Relacje wyznaczające zbiór krytyczny Z oraz szkice zbiorów krytycznych pokazują poniższe rysunki (7.1.1–7.1.3).



Rys. 7.1.1. Zbiór krytyczny dwustronny $H_1 : m \neq m_0$ $P\{|T| > t_\alpha\} = \alpha$



Rys. 7.1.2. Zbiór krytyczny lewostronny $H_1 : m < m_0$ $P\{T < -t_\alpha\} = \alpha$



Rys. 7.1.3. Zbiór krytyczny prawostronny $H_1 : m > m_0$ $P\{T > t_\alpha\} = \alpha$

Wyznaczanie wartości krytycznej.
Gdy sprawdzianem hipotezy H_0 jest statystyka:

$$T = \frac{\bar{x} - m}{\sigma} \cdot \sqrt{n},$$

która przy założeniu prawdziwości hipotezy H_0 ma rozkład $N(0, 1)$, wówczas wartość krytyczną t_α odczytujemy z tablic wartości dystrybuanty $\Phi(t_\alpha)$ standaryzowanego rozkładu normalnego w ten sposób, że przy zbiorze dwustronnym:

$$\Phi(t_\alpha) = \frac{1 - \alpha}{2},$$

a przy zbiorze jednostronnym

$$\Phi(t_\alpha) = \frac{1}{2} - \alpha.$$

Dla przypadku, gdy sprawdzianem hipotezy H_0 jest statystyka:

$$T = \frac{\bar{x} - m}{\sigma} \cdot \sqrt{n-1},$$

która ma rozkład t-Studenta, wówczas wartość krytyczną t_α odczytujemy z tablic rozkładu t-Studenta dla $n-1$ stopni swobody w ten sposób, że przy zbiorze dwustronnym:

$$P(t_\alpha) = \alpha,$$

a przy zbiorach jednostronnych:

$$P(t_\alpha) = 2\alpha.$$

Sprawdzanie hipotezy H_0 .

Na podstawie wyników n -elementowej próby prostej oblicza się empiryczną wartość t_e sprawdzianu (statystyki T):

$$t_e = \frac{\bar{x} - m}{\sigma} \cdot \sqrt{n}$$

lub

$$t_e = \frac{\bar{x} - m}{\sigma} \cdot \sqrt{n-1}.$$

Jeśli relacja wyznaczająca zbiór krytyczny jest spełniona, czyli wartość empiryczna t_e statystyki T wpada w zbiór krytyczny, to hipotezę H_0 należy odrzucić. W przeciwnym przypadku mówimy, że nie ma podstaw do odrzucenia hipotezy H_0 .

Testem o dużym znaczeniu w zastosowaniach jest test istotności dla wskaźnika struktury p populacji generalnej, przy czym warunek konieczny stanowi dysponowanie dużą próbą. Hipotezy możemy sformułować następująco:

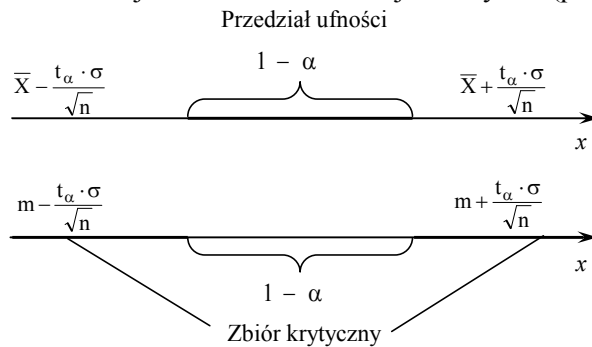
$$H_0: p = p_0,$$

$$H_1: p \neq p_0 \text{ lub } p > p_0 \text{ lub } p < p_0.$$

Statystyka T dla potrzeb weryfikacji hipotezy o wskaźniku struktury populacji generalnej p wyraża się wzorem:

$$T = \frac{\frac{x}{n} - p}{\sqrt{\frac{p(1-p)}{n}}},$$

Należy zwrócić uwagę na analogię przedziału ufności budowanego w ramach estymacji parametrycznej oraz obszaru krytycznego, określanego przy budowie testu parametrycznego do weryfikacji hipotez o parametrach rozkładu. Pokażemy to na przykładzie cechy X , która ma rozkład $N(m, \sigma)$ przy założeniu nieznannej wartości oczekiwanej i znanym σ (p. rys. 7.1.4).



Rys. 7.1.4. Związek pomiędzy przedziałem ufności a obszarem krytycznym

Przykład 7.1.1

Zakłada się, że trwałość opon samochodowych ma rozkład $N(m, \sigma)$. Producent twierdzi, że trwałość ta wynosi 50 tys. kilometrów przebiegu. Na podstawie badań próby 100 losowo wybranych opon otrzymano następujące wyniki: $\bar{x} = 45$ tys. oraz $s = 8$ tys. kilometrów. Czy na poziomie istotności $\alpha = 0,05$ można uważać, że producent ma rację?

Rozwiązanie

Hipoteza H_0 ma postać:

$$H_0 : m = 50,$$

hipotezą alternatywną H_1 będzie:

$$H_1 : m \neq 50.$$

Gdyby producent twierdził, że średnia trwałość opon samochodowych jest dłuższa od 50 tys. km, to hipoteza H_1 będzie miała postać:

$$H_1 : m > 50,$$

gdyby natomiast twierdził on, że trwałość będzie krótsza, to hipoteza H_1 będzie o mniejszości:

$$H_1 : m < 50.$$

Poziom istotności jest zadany i wynosi $\alpha = 0,05$.

Przy zadanej liczności próby $n = 100$ oraz znanym odchyleniu standardowym w próbie $s = 8$, sprawdzianem hipotezy H_0 będzie statystyka:

$$T = \frac{\bar{x} - m}{\sigma} \cdot \sqrt{n}.$$

Wartość empiryczna t_e statystyki T wynosi:

$$t_e = \frac{\bar{x} - m}{\sigma} \cdot \sqrt{n} = \frac{45 - 50}{8} \cdot \sqrt{100} = -6,25.$$

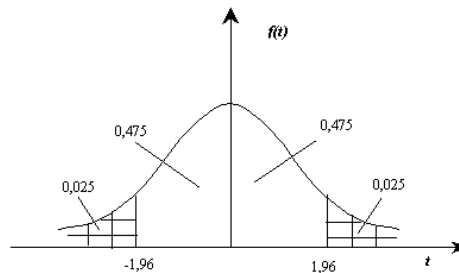
Sprawdzamy, czy wartość tej statystyki wpada do obszaru krytycznego, czy też nie.

Wartość krytyczna $t_\alpha = 1,96$ wyznaczana jest dla zbioru dwustronnego dla:

$$\Phi(t_\alpha) = \frac{1 - \alpha}{2} = \frac{1 - 0,05}{2} = 0,475,$$

ponieważ hipoteza alternatywna jest „o nierówności”.

Obszar krytyczny (p. rys. 7.1.5) przedstawia się następująco:



Rys. 7.1.5. Obszar krytyczny dwustronny dla $t_\alpha = 1,96$ i $\Phi(t_\alpha) = 0,475$

Ponieważ wartość empiryczna sprawdzianu należy do obszaru krytycznego $t_e < t_\alpha$ ($-6,25 < -1,96$) należy hipotezę H_0 odrzucić na korzyść hipotezy alternatywnej, tzn., że trwałość opon nie jest równa 50000 km.

Wartość krytyczna $t_\alpha = 1,65$ dla zbiorów lewo- i prawostronnego wynosi:

$$\Phi(t_\alpha) = \frac{1}{2} - \alpha = 0,5 - 0,05 = 0,45.$$

Gdyby producent twierdził, że trwałość opon jest mniejsza od 50000 km czyli hipoteza alternatywna byłaby „o mniejszości”, to nie byłoby podstaw do odrzucenia hipotezy H_0 , czyli można byłoby uważać, że trwałość opon jest niższa niż 50 000 km.

Przykład 7.1.2

Przeprowadzono kontrolę płatności zobowiązań podatkowych w pewnej branży usługowej w związku z opinią, że 20% firm płaci podatki nieregularnie. W celu zbadania, czy ta opinia jest słuszna wylosowano niezależnie z każdego województwa Polski 100 firm i okazało się, że $\bar{x} = 15\%$, a $s = 4\%$. Zakładając poziom istotności 0,05 zweryfikować hipotezę dotyczącą słuszności opinii, że 20% firm tej branży płaci nieregularnie podatki.

Rozwiązanie

Hipoteza H_0 ma postać:

$$H_0: m = 0,2$$

wobec hipotezy alternatywnej

$$H_1: m \neq 0,2.$$

W naszym przykładzie próba jest duża i wynosi $n > 30$ ($n = 100$), a odchylenie standardowe znane ($s = 0,04$), a zatem sprawdzianem hipotezy H_0 będzie statystyka:

$$T = \frac{\bar{x} - m}{\sigma} \cdot \sqrt{n}.$$

Wartość empiryczna t_e statystyki T wynosi:

$$t_e = \frac{\bar{x} - m}{\sigma} \cdot \sqrt{n} = \frac{45 - 50}{8} \cdot \sqrt{100} = -6,25.$$

Wartość krytyczna $t_\alpha = 1,96$ wyznaczana jest dla zbioru dwustronnego dla:

$$\Phi(t_\alpha) = \frac{1 - \alpha}{2} = \frac{1 - 0,05}{2} = 0,475,$$

ponieważ hipoteza alternatywna jest „o nierówności”.

Wartość empiryczna sprawdzianu należy do obszaru krytycznego $t_e < t_\alpha$ ($-6,25 < -1,96$).

Hipotezę H_0 należy odrzucić na korzyść hipotezy alternatywnej H_1 , a zatem opinia, że 20% firm badanej branży nieregularnie płaci podatki nie jest słuszna.

Przykład 7.1.3

Przedsiębiorstwo produkcyjne P zastanawia się nad podpisaniem umowy z BRE Bankiem S.A. na realizowanie zleceń płatniczych za pomocą nowego systemu rozliczeń. Rachunek kosztów wykazał, że system ten jest rentowny przy dużej skali obrotu, tzn. przy składaniu powyżej 368 zleceń miesięcznie. Zbadano ilość realizowanych zleceń w firmie w losowo wybranych 24 miesiącach i okazało się, że przeciętna ilość zleceń wynosi 375, a odchylenie

standardowe 15 zleceń. Zweryfikować odpowiednią hipotezę na poziomie istotności $\alpha = 0,05$.

Rozwiązanie

$$H_0: m = 368, \quad H_1: m > 368.$$

Ponieważ próba $n < 30$, a zatem sprawdzianem hipotezy będzie statystyka

$$T = \frac{\bar{x} - m}{\sigma} \cdot \sqrt{n-1}.$$

$$t_e = \frac{\bar{x} - m}{s} \cdot \sqrt{n-1} = \frac{375 - 368}{15} \cdot \sqrt{24-1} = 2,23,$$

$$P(t_\alpha) = 2\alpha = 2 \cdot 0,05 = 0,1 \Rightarrow t_\alpha = 1,714.$$

Ponieważ t_e wpada w zbiór krytyczny, bo $t_e = 2,23 > t_\alpha = 1,714$, zatem hipotezę H_0 należy odrzucić na korzyść hipotezy H_1 . Można uznać, że skala obrotów w badanej firmie jest wyższa od 368 zleceń, toteż korzystanie z nowego systemu będzie opłacalne.

Przykład 7.1.4

W latach 1990-1999 zameldowania polskich gości w pewnym kraju europejskim (w tys.) wynosiły: 20,7; 19,5; 19,8; 57,5; 82,5; 77,9; 105,7; 111,5; 97,6; 109,6. Zakładamy, że zameldowania te można uznać za cechę o rozkładzie normalnym. Przyjmując poziom istotności $\alpha = 0,005$ zweryfikować hipotezę, że liczba zameldowań w ciągu roku polskich gości w badanym kraju jest mniejsza od 71 tys. osób.

Rozwiązanie

$$H_0: m = 71, \quad H_1: m < 71.$$

Wykonując obliczenia otrzymujemy: $\bar{x} = 70,2$; $s = 36,3$. Próba jest mała, stąd

$$t_e = \frac{\bar{x} - m}{s} \cdot \sqrt{n-1} = \frac{70,2 - 71}{36,3} \cdot \sqrt{10-1} = -0,07,$$

$$P(t_\alpha) = 2\alpha = 2 \cdot 0,005 = 0,01 \Rightarrow t_\alpha = -3,25.$$

Nie ma podstaw do odrzucenia hipotezy H_0 , ponieważ $t_e = -0,07 > t_\alpha = -3,25$ nie wpada w zbiór krytyczny. Przypuszczenie, że liczba zameldowań jest mniejsza od 71 tys. osób jest nieuzasadnione.

Przykład 7.1.5

Pojawiła się opinia, że ponad 70% firm usługowych korzysta z reklamy radiowej. W celu zweryfikowania tej opinii przeprowadzono ankietę w 150 firmach usługowych i zapytano o rodzaj stosowanej reklamy. W grupie an-

kietowanej znalazły się 83 firmy korzystające z reklamy radiowej. Sprawdzić wysuniętą hipotezę na poziomie istotności $\alpha = 0,02$.

Rozwiązanie

$$H_0: p = 0,7, \quad H_1: p > 0,7.$$

Sprawdzianem hipotezy będzie statystyka:

$$T = \frac{\frac{\bar{x} - p}{\sqrt{\frac{p(1-p)}{n}}}$$

$$\Phi(t_\alpha) = \frac{1}{2} - \alpha = \frac{1}{2} - 0,02 = 0,48 \Rightarrow t_\alpha = 2,05.$$

Z obliczeń wynika, że $t_e = -3,74 < t_\alpha = 2,05$, a więc t_e nie wpada w zbiór krytyczny. Hipotezę H_0 odrzucamy na korzyść hipotezy alternatywnej H_1 , można zatem uznać, że 70% firm usługowych korzysta z reklamy radiowej.

Przykład 7.1.6

Przeprowadzono badanie wieku wśród studentów pewnego wydziału uczelni akademickiej i pobrano próbę losową $n = 16$ o średniej $\bar{x} = 28$ lat i odchyleniu standardowym $s = 3$ lata. Czy można przypuszczać, że są powody do odrzucenia hipotezy na poziomie $\alpha = 0,05$, że średni wiek w populacji wynosi $m = 30$ lat?

Rozwiązanie

Nie znamy wartości odchylenia standardowego w populacji, dlatego zastosujemy test t-Studenta. Mamy tu $m = 30$, $\bar{x} = 28$, $s = 3$, $n = 16$, a zatem:

$$t_e = \frac{28 - 30}{3} \sqrt{15} = -2,58.$$

Dla 15 stopni swobody oraz $\alpha = 0,05$ wartość krytyczna odczytana z tablic t-Studenta $t_{0,05}$ wynosi 2,131, a więc hipotezę należy odrzucić ($t_e < t_\alpha$).

7.2. Hipotezy nieparametryczne

Hipotezy nieparametryczne są formułowane w odniesieniu do postaci rozkładu lub dystrybuanty danej cechy w populacji generalnej, np.:

$$H_0: \text{cecha } X \text{ ma rozkład określony dystrybuantą } F(x) = F_0(x),$$

$$H_0: \text{cecha } X \text{ ma rozkład normalny } N(m, \sigma),$$

$$H_0: \text{cecha } X \text{ ma rozkład Poissona z parametrem } \lambda = 2.$$

Do weryfikacji tych hipotez H_0 służą *nieparametryczne testy istotności*, które dzielą się na trzy grupy:

- a) testy zgodności,
- b) testy niezależności,
- c) testy losowości próby.

Testy te w przeciwieństwie do testów parametrycznych nie wymagają założeń co do postaci rozkładu cechy w populacji generalnej.

Test zgodności chi-kwadrat

Za pomocą testu zgodności χ^2 mogą być weryfikowane hipotezy dotyczące rozkładów typu ciągłego i skokowego.

Sprawdzianem hipotezy H_0 jest statystyka:

$$\chi^2 = \sum_{i=1}^k \frac{(n_i - n \cdot p_i)^2}{n \cdot p_i},$$

gdzie: k – liczba przedziałów klasowych,

s – liczba parametrów wyznaczanych na podstawie wyników z próby,

p_i – prawdopodobieństwo, że cecha X przyjmie wartość należącą do i -tego przedziału klasowego,

n_i – liczebność przedziału,

n – liczebność próby.

Statystyka ta, przy założeniu prawdziwości hipotezy H_0 , ma rozkład χ^2 o $r = (k - s - 1)$ stopniach swobody.

Test zgodności χ^2 można stosować dla danych pochodzących z dużych prób prostych. Dane należy przedstawić w postaci szeregu rozdzielczego o minimum pięciu przedziałach klasowych ($k \geq 5$). Liczebność każdej klasy nie powinna być mniejsza od ośmiu ($n_i \geq 8$).

Relacja wyznaczająca obszar krytyczny ma postać:

$$P(\chi^2 > \chi_\alpha^2) = \alpha,$$

gdzie: χ_α^2 – wartość krytyczna odczytywana z tablic rozkładu χ^2 dla r stopni swobody oraz $P = \alpha$.

Przykład 7.2.1

Z analizy liczby ujawnionych kradzieży w pewnym supermarkecie otrzymano następujące dane:

Dzienna liczba kradzieży	0	1	2	3	4
Liczba dni	22	30	22	16	10

Czy na poziomie istotności $\alpha = 0,05$ można sądzić, że rozkład liczby kradzieży jest rozkładem Poissona?

Rozwiązanie

Zmienną losową X jest liczba kradzieży w ciągu dnia. Hipoteza H_0 ma postać:

$$H_0 : X - \text{ma rozkład Poissona,}$$

wobec hipotezy alternatywnej H_1 :

$$H_1 : X - \text{nie ma rozkładu Poissona.}$$

Aby wyznaczyć teoretyczne liczebności realizacji zmiennej losowej X należy obliczyć prawdopodobieństwa p_i tego, że zmienna ta przyjmuje wartość należącą do i -tego przedziału. Prawdopodobieństwa te wyznacza się z tablic rozkładu Poissona dla parametru tak, że:

$$p_i = P(X = x_i) \text{ przy czym } p_4 = 1 - \sum_{i=1}^3 p_i ,$$

zgodnie z zasadą, że $\sum_{i=1}^4 p_i = 1$.

Na podstawie danych oblicza się wartość parametru $\lambda = m = E(X)$, którego estymatorem jest wartość przeciętna \bar{x} oraz wartość statystyki:

$$\chi_e^2 = \sum_{i=1}^k \frac{(n_i - n \cdot p_i)^2}{n \cdot p_i} .$$

x_i	n_i	p_i	$n \cdot p_i$	$(n_i - n \cdot p_i)^2$	$\frac{(n_i - n \cdot p_i)^2}{n \cdot p_i}$
0	22	0,202	20,2	3,24	0,16
1	30	0,323	32,3	5,29	0,16
2	22	0,258	25,8	14,44	0,56
3	16	0,138	13,8	4,84	0,35
4	10	0,079	7,9	4,41	0,558
Razem	100	1,000	x	x	$\chi_e^2 = 1,788$

Z tablic rozkładu χ^2 dla $r = k - s - 1 = 3$ stopni swobody ($s = 1$, ponieważ na podstawie próby szacowano jeden parametr) i $P = 0,05$ odczytano wartość krytyczną $\chi_\alpha^2 = 7,815$.

Ponieważ $\chi_e^2 < \chi_\alpha^2$ nie ma podstaw, aby odrzucić hipotezę H_0 , że rozkład liczby kradzieży jest rozkładem Poissona.

Test zgodności Lambda-Kołmogorowa

Za pomocą testu zgodności Lambda-Kołmogorowa mogą być weryfikowane hipotezy, że określona cecha X ma w populacji generalnej rozkład typu ciągłego. W tekście tym porównuje się dystrybuantę teoretyczną z empiryczną. Warunki dotyczące danych z próby są identyczne jak dla testu χ^2 . Sprawdzianem hipotezy H_0 jest statystyka:

$$\lambda = D_n \cdot \sqrt{n},$$

gdzie: $D_n = \sup_{-\infty < x < +\infty} |F(x) - F^*(x)|,$

przy czym: $F(x)$ – dystrybuanta teoretyczna (hipotetyczna),

$F^*(x)$ – dystrybuanta empiryczna.

Wartość dystrybuanty teoretycznej weryfikowanego rozkładu odczytuje się z tablic statystycznych, korzystając z następującej zależności:

$$F(x) = \begin{cases} 0,5 - \Phi(x) & \text{dla } x < 0 \\ 0,5 & \text{dla } x = 0 \\ 0,5 + \Phi(x) & \text{dla } x > 0 \end{cases}.$$

Wartość dystrybuanty empirycznej $F^*(x)$ oblicza się z zależności:

$$F^*(x) = \frac{n_{sk}}{n},$$

gdzie: n_{sk} – skumulowana liczebność odpowiadająca wartościom cechy nie większym od x ,
 n – liczebność próby.

Zbiór krytyczny λ jest zbiorem prawostronnym określonym wzorem:

$$P(\lambda \geq \lambda_\alpha) = \alpha.$$

Wartość λ_α odczytuje się z tablic rozkładu Lambda-Kołmogorowa według zależności $Q(\lambda_\alpha) = 1 - \alpha$.

Hipotezę H_0 należy odrzucić, jeżeli obliczona empiryczna wartość λ_e należy do zbioru krytycznego λ_α , czyli $\lambda_e \geq \lambda_\alpha$, a w przeciwnym przypadku nie ma podstaw do jej odrzucenia.

Przykład 7.2.2

Wykonano badania 87 osób zatrudnionych przy produkcji pewnego towaru. Otrzymano następujące informacje o wydajności pracy:

Wydajność w szt./godz. $x_i^d - x_i^g$	Liczba osób n_i
0 - 2	3
2 - 4	5
4 - 6	12
6 - 8	20
8 - 10	19
10 - 12	16
12 - 14	8
14 - 16	4

Na poziomie istotności $\alpha = 0,05$ zweryfikować za pomocą testu λ -Kolmogorowa hipotezę, że rozkład wydajności pracy jest normalny.

Rozwiązanie

Oznaczając przez X wydajność pracy formułujemy hipotezę H_0 , że cecha X , która jest zmienną losową ma w populacji rozkład normalny:

$$H_0 : X - \text{ma rozkład normalny,}$$

wobec hipotezy alternatywnej H_1 :

$$H_1 : X - \text{nie ma rozkładu normalnego.}$$

Obliczamy wartość standaryzowanej zmiennej losowej Z w poszczególnych przedziałach klasowych:

$$z_i = \frac{x_i^g - \bar{x}}{s}; \quad i = 1, \dots, k,$$

gdzie: \bar{x} - wartość średnia ważona z próby,
 s - odchylenie standardowe z próby

$$\bar{x} = \frac{1}{n} \sum_{i=1}^k \dot{x}_i \cdot n_i = \frac{1 \cdot 3 + 3 \cdot 5 + 5 \cdot 12 + 7 \cdot 20 + 9 \cdot 19 + 11 \cdot 16 + 13 \cdot 8 + 15 \cdot 4}{87} = 8,38,$$

$$s = \sqrt{\frac{1}{n} \sum_{i=1}^k (\dot{x}_i - \bar{x})^2 \cdot n_i} = 3,29,$$

x_i^g	n_i	n_{sk}	$F^*(x)$	z_i	$\Phi(z_i) = F(x)$	$ F(x) - F^*(x) $
2	3	3	0,0345	-1,94	0,0262	0,0083
4	5	8	0,0920	-1,33	0,0918	0,0002
6	12	20	0,2299	-0,72	0,2358	0,0059
8	20	40	0,4552	-0,12	0,4562	0,0010
10	19	59	0,6782	0,49	0,6876	0,0094
12	16	75	0,8621	1,10	0,8438	$D_n = 0,0183$
14	8	83	0,9540	1,71	0,9554	0,0014
16	4	87	1,0000	2,32	0,9896	0,0104

Z tabeli odczytujemy wartość $D_n = 0,0183$ i obliczamy wartość statystyki $\lambda_e = D_n \cdot \sqrt{n} = 0,0183 \cdot \sqrt{87} = 0,1697$, którą porównujemy z odczytaną z tablic rozkładu λ -Kolmogorowa wartością krytyczną $\lambda_\alpha = 1,36$; dla $\alpha = 0,05$.

Ponieważ $\lambda_e < \lambda_\alpha$, to na poziomie istotności $\alpha = 0,05$ nie ma podstaw do odrzucenia hipotezy H_0 , że badana cecha (zmienna losowa) ma rozkład normalny.

7.3. Powtórzenie w przykładach

Zadanie 1

Z partii detali pobrano próbę liczącą 400 pudełek. Z próby tej otrzymano informację, że średnio w pudełku jest 47,1 detali przy odchyleniu standardowym 2,5. Na poziomie istotności $\alpha = 0,02$ zweryfikować hipotezę zerową, że w całej partii średnia liczba detali w pudełku wynosi 48 wobec hipotezy alternatywnej, że jest mniejsza od 48.

Rozwiązanie

$$H_0: m = 48, \quad H_1: m < 48.$$

Wartość empiryczna statystyki T wynosi:

$$t_e = \frac{\bar{x} - m}{\sigma} \cdot \sqrt{n} = \frac{47,1 - 48}{2,5} \cdot \sqrt{400} = -7,2.$$

Wartość krytyczna $t_\alpha = -2,05$ wyznaczona dla zbioru lewostronnego $\Phi(t_\alpha) = \frac{1}{2} - \alpha$, ponieważ hipoteza alternatywna dotyczy „mniejszości”. Ponieważ $t_e = -7,2 < t_\alpha = -2,05$, a zatem hipotezę H_0 należy odrzucić, bo wartość sprawdzianu hipotezy należy do obszaru krytycznego. Oznacza to, że średnia liczba detali w pudełku jest mniejsza niż 48.

Zadanie 2

Pewna grupa młodzieży wybrała się na wycieczkę statkiem. W czasie złej pogody statek mocno kołysał, a niektórzy uczestnicy doświadczali choroby morskiej. Dane dotyczące liczby zachorowań podczas podróży wskazuje poniższa tabela.

	Chorowało	Nie chorowało	Łącznie
Chłopcy	24	30	54
Dziewczęta	18	26	34
Łącznie	32	56	88

Rozwiązanie

Można by przypuszczać, że dane wskazują, iż dziewczęta lepiej znoszą kołysanie statkiem podczas złej pogody, a procent chorujących dziewcząt jest mniejszy niż chłopców. Aby to sprawdzić postawimy hipotezę, że częstość zachorowań nie zależy od płci oraz zastosujemy test χ^2 . W tym przypadku χ^2 będzie wynosiło 3,67 przy 1 stopniu swobody. Dla poziomu istotności $\alpha = 0,05$ wartość krytyczna χ^2 przy 1 stopniu swobody wyniesie 3,84. Relacja wyznaczająca obszar krytyczny ma postać: $P(\chi^2 > \chi_{\alpha}^2) = \alpha$. W przykładzie relacja ta jest przeciwna, bo $3,67 < 3,84$. Nie ma podstaw do odrzucenia hipotezy, że częstość zachorowań nie zależy od płci.

Zadanie 3

Popyt na wybrany towar modelujemy zmienną losową o rozkładzie normalnym. W próbie 10-elementowej otrzymaliśmy średnią 1250 kg i odchyleniu 50 kg. Na poziomie istotności $\alpha = 0,05$ zweryfikować hipotezę $H_0: \sigma = 45$ wobec hipotezy alternatywnej $H_1: \sigma > 45$.

Rozwiązanie

Hipoteza $H_0: \sigma = \sigma_0$ jest równoważna hipotezie $H_0: \sigma^2 = \sigma_0^2$, a zatem hipoteza o odchyleniu standardowym jest równoważna hipotezie o wariancji [Cieciński]. Sprawdzianem hipotezy w naszym przykładzie jest statystyka

$T = \frac{nS_n^2}{\sigma_0^2}$ o rozkładzie χ^2 z $n - 1$ stopniami swobody. Wykonujemy oblicze-

nia dla wartości empirycznej statystyki T i otrzymujemy $t_e = \frac{10 \cdot 50^2}{45} = 11,1$.

Z tablic rozkładu χ^2 dla 9 stopni swobody odczytujemy $\chi_{0,05}^2 = 16,919$, stąd $t_{\alpha} = 16,919 > t_e = 11,1$ a zatem nie ma podstaw do odrzucenia H_0 . Relacja $\chi_e^2 > \chi_{\alpha}^2$ wyznacza tu prawostronny zbiór krytyczny, gdzie χ_{α}^2 jest wartością krytyczną odczytaną z tablic rozkładu χ^2 dla $n - 1$ stopni swobody. Jeśli dla

danej próby losowej relacja wyznaczająca zbiór krytyczny jest spełniona, to H_0 odrzucamy na korzyść H_1 , w przeciwnym przypadku nie ma podstaw do odrzucenia H_0 .

Zadanie 4

W próbie losowej liczącej 10 wycieczek 7-dniowych zaobserwowano następujące koszty (w tys. zł): 7,5 14,0 8,5 15,0 12,0 6,0 11,5 13,5 9,0 10,0. Zweryfikować hipotezę, że przeciętne koszty wycieczek 7-dniowych zorganizowanych przez TURPOL w sierpniu 2008 roku były mniejsze od 13 tys. zł.

Rozwiązanie

$$H_0: m = 13, \quad H_1: m < 13.$$

Ponieważ próba jest mała ($n < 30$), to do weryfikacji hipotezy H_0 stosujemy statystykę:

$$T = \frac{\bar{x} - m}{s} \cdot \sqrt{n-1}.$$

Statystyka ta, przy założeniu prawdziwości H_0 ma rozkład t-Studenta o $n - 1$ stopniach swobody. Obliczamy średnią i odchylenie standardowe z próby, a zatem $\bar{x} = 10,7$ oraz $s = 2,8$. Stąd

$$t_e = \frac{10,7 - 13}{2,8} \cdot \sqrt{10-1} = -2,464.$$

Z tablic rozkładu t-Studenta odczytujemy wartość krytyczną t_α tak, aby dla danego α i $n - 1$ stopni swobody spełniona była relacja

$$P\{t_e \leq t_\alpha\} = \alpha.$$

Nierówność $t_e < t_\alpha$ wyznacza lewostronny obszar krytyczny testu. Dla lewostronnego obszaru krytycznego t odczytujemy dla 2α i wynikowi nadajemy znak „-”. A zatem $t_e = -2,464 < t_{2\alpha} = -1,383$.

Oznacza to, że hipotezę H_0 należy odrzucić na korzyść hipotezy alternatywnej H_1 . W rzeczywistości średnie koszty wycieczek 7-dniowych są niższe niż 13 tys. zł.

Zadanie 5

Przy wycenie lasu dębowego wylosowano 100 dębów i stosując odpowiednie urządzenia pomiarowe określono ich wysokość. W zastosowanym urządzeniu pomiarowym błąd pomiaru jest wielkością losową o rozkładzie $N(0; 1,25)$. Sprawdzić na poziomie istotności $\alpha = 0,05$, czy opis drzewostanu dokonany wcześniej przez biegłego (oparty na wyznaczonej przez niego średniej wysokości drzewa $m = 25$ m) jest do przyjęcia, jeśli z uzyskanej próby 100-elementowej otrzymano przeciętną wysokość dębu $\bar{x} = 27$ m.

Rozwiązanie

Liczebność pobranej próby jest duża ($n > 30$) oraz znana jest wartość parametru $\sigma = 1,45$, wobec tego:

$$H_0: m = 25, \quad H_1: m > 25.$$

Korzystamy ze statystyki T danej wzorem:

$$T = \frac{\bar{X} - m}{\sigma} \cdot \sqrt{n}.$$

Empiryczna wartość statystyki t_e wynosi:

$$t_e = \frac{27 - 25}{1,25} \cdot \sqrt{100} = 16.$$

Odczytując z tablic rozkładu $N(0, 1)$ wartość krytyczną otrzymujemy $t_\alpha = 1,64$ (prawostronny obszar krytyczny). Ponieważ $t_e = 16 > t_\alpha = 1,64$ hipotezę zerową należy odrzucić na korzyść hipotezy alternatywnej. A zatem opis drzewostanu dokonany wcześniej przez biegłego należy skorygować.

Zadanie 6

Właściciel salonu samochodowego twierdzi, że klienci w jednakowym stopniu preferują 5 typowych kolorów samochodów. Wylosowano niezależną próbę 200 klientów odwiedzających badany salon i otrzymano następujące dane dotyczące częstości preferowania kolorów samochodu:

Kolor samochodu	srebrny	czarny	czerwony	żółty	zielony
Liczba klientów	46	45	32	43	34

Zweryfikować hipotezę o słuszności opinii właściciela salonu, przyjmując poziom istotności $\alpha = 0,05$.

Rozwiązanie

Prawdopodobieństwo wyboru każdego z podanych kolorów $p_i = 1/5$.

$H_0: p = p_i$ (prawdopodobieństwo wyboru każdego ze wskazanych kolorów jest jednakowe),

$H_1: p \neq p_i$ (prawdopodobieństwo wyboru każdego ze wskazanych kolorów jest różne).

Tablica obliczeniowa

Kolory samochodów	n_i	$n \cdot p_i$	$(n_i - np_i)^2$	$\frac{(n_i - np_i)^2}{np_i}$
srebrny	46	40	36	0,900
czarny	45	40	25	0,625
czerwony	32	40	64	1,600
żółty	43	40	9	0,225
zielony	34	40	36	0,900
Razem	200	200	x	$\chi^2 = 4,250$

Z tablicy rozkładu χ^2 dla $5 - 1 = 4$ stopni swobody oraz $\alpha = 0,05$ odczytujemy wartość $\chi_{0,05}^2 = 9,488$.

Ponieważ $\chi^2 = 4,250 < \chi_{0,05}^2 = 9,488$ nie ma podstaw do odrzucenia hipotezy H_0 . Właściciel salonu miał rację. Nie ma podstaw, aby sądzić, że klienci preferują w szczególności sposób którykolwiek z kolorów.

ROZDZIAŁ VIII

WYBRANE POJĘCIA STATYSTYKI MATEMATYCZNEJ

Statystyka matematyczna – dział rachunku prawdopodobieństwa, poświęcony metodom zbierania danych i wyciągania z nich wniosków, gdy wiedza a priori o badanym problemie jest sformułowana w postaci pewnych modeli probabilistycznych, odpowiednich dla opisu zjawisk losowych.

Geneza – obserwowanie pojedynczych nierzadko przypadkowych i wyrwanych z kontekstu zdarzeń nie pozwala na formułowanie żadnych wniosków odnośnie ich istoty. Dopiero analiza dużej liczby takich zdarzeń pozwala na zaobserwowanie prawidłowości. Narzędziem, które pomaga w dostrzeżeniu i zrozumieniu tych prawidłowości są metody statystyczne. Służą do opisu przebiegu zjawisk i procesów oraz oceny związków między nimi. Stosowane są wszędzie tam, gdzie przeprowadza się pomiary, rejestracje i ewidencje zjawisk przyrodniczych, technicznych i ekonomicznych. Statystyka ma szerokie zastosowanie zarówno w zarządzaniu przedsiębiorstwem, gospodarką narodową, jak i gospodarstwem domowym. Metody statystyczne są niezbędne w takich dziedzinach, jak ocena jakości produkcji, analiza rynku, czy działalność firm ubezpieczeniowych.

Wnioskowanie statystyczne – uogólnianie wyników badania części zbiorowości zwanej próbą losową na całą zbiorowość (populację), z której ta część pochodzi, z równoczesnym szacowaniem wielkości popełnianego błędu tego uogólnienia. Uogólnianie wyników badania części zbiorowości na populację możemy przeprowadzić metodami estymacji lub weryfikacji.

Zbiorowość statystyczna (populacja generalna, populacja) – zbiór dowolnych elementów, nieidentycznych, stanowiących jedną logiczną całość. Zbiorowość statystyczna może być skończona $(1, 2, \dots, n)$ lub nieskończona, ale przeliczalna $(1, 2, \dots)$. Aby zbiór jednostek był zbiorowością statystyczną powinien być jednocześnie identycznym ze względu na cechę stałą oraz różnić się przynajmniej ze względu na jedną cechę zmienną.

Próba (próbka) – część, tj. podzbiór populacji, skończony lub nieskończony, podlegający bezpośrednio badaniu ze względu na ustaloną cechę w celu wyciągnięcia wniosków o kształtowaniu się wartości tej cechy w populacji.

Próba prosta n-elementowa, wylosowana ze skończonej lub nieskończonej populacji – próba losowa, której wyniki są niezależnymi zmiennymi losowymi posiadającymi jednakowe rozkłady, identyczne jak w populacji. Aby uzyskać próbę prostą należy zastosować schemat losowania indywidualnego, nieograniczonego, niezależnego.

Próba złożona – skonstruowana ze skończonej populacji przy zastosowaniu bardziej skomplikowanych niż przy próbach prostych schematach losowania, np. losowanie zależne warstwowe.

Liczebność próby – liczba jednostek, elementów populacji generalnej wybranych do próby. Liczebność próby zwykle oznaczamy przez n . Z uwagi na liczebność próby dzielimy na małe i duże. Gdy $n < 30$ mamy do czynienia z małą próbą.

Próba losowa – próba wybrana z populacji w sposób losowy, tzn. tak, że tylko przypadek decyduje o zaliczeniu elementu populacji generalnej do próby.

Próba reprezentacyjna – próba losowa, której struktura (budowa) pod względem badanej cechy nie różni się istotnie od struktury zbiorowości generalnej ze względu na tę cechę. Próby reprezentacyjne uzyskujemy za pomocą odpowiednich schematów losowania.

Schemat losowania próby – praktyczny sposób (algorytm) losowania elementów populacji generalnej do próby, uwzględniający możliwości techniczne, koszty i efektywność uzyskiwania wyników.

Losowanie niezależne – schemat losowania próby ze zwracaniem każdego wylosowanego elementu w trakcie losowania tak, że jeden i ten sam element może być wylosowany do próby więcej niż jeden raz.

Losowanie zależne – schemat losowania próby bez zwracania każdego wylosowanego elementu populacji generalnej tak, że jeden i ten sam element populacji może zostać wylosowany do próby tylko jeden raz.

Losowanie nieograniczone – losowanie elementów do próby od razu z całej populacji, w odróżnieniu np. od losowania warstwowego, które jest ograniczone.

Losowanie warstwowe – losowanie próby oddzielnie z każdej części populacji generalnej (tzw. warstwy populacji generalnej), na które została podzielona przed losowaniem.

Losowanie indywidualne – losowanie oddzielnie poszczególnych elementów populacji generalnej do próby w odróżnieniu np. od *losowania zespolonego*, w którym losuje się do próby pewne naturalne zespoły elementów populacji generalnej.

Wyniki próby – zaobserwowane wartości badanej cechy u tych elementów populacji generalnej, które zostały dobrane do próby. Wyniki próby losowej o liczebności n stanowią wartości n -wymiarowej zmiennej losowej. Wyniki uzyskane z obserwacji dużych prób (zawierających powyżej 30 elementów) grupuje się najczęściej w klasy, tworząc tzw. rozdzielcze szeregi statystyczne.

Szereg rozdzielczy – zbiorowość statystyczna podzielona na części (klasy) według określonych wariantów cechy z podaniem liczebności wyodrębnionych klas.

Przestrzeń próby – zbiór wszystkich możliwych wyników próby o liczebności n .

Rozkład populacji – rozkład wartości badanej cechy statystycznej w całej zbiorowości.

Parametry populacji – parametry rozkładu badanej cechy w populacji, charakteryzujące ten rozkład. Przy badaniu pełnym mówimy o parametrach populacji, przy badaniu częściowym o parametrach z próby, zwanych miernikami statystycznymi.

Statystyka z próby – zmienna losowa, będąca dowolną funkcją wyników próby losowej, np. średnia arytmetyczna czy mediana.

Rozkład statystyki – teoretyczny rozkład prawdopodobieństwa zmiennej losowej będącej statystyką. Rozkład ten zależy zwykle od rozkładu populacji i schematu losowania n -elementowej próby.

Standaryzacja rozkładu normalnego – zamiana rozkładu normalnego $N(\mu, \sigma)$ na rozkład normalny standaryzowany $N(0,1)$.

Estymacja – szacowanie wartości parametrów lub postaci rozkładu zmiennej losowej w populacji generalnej na podstawie rozkładu empirycznego uzyskanego z próby.

Estymator – dowolna statystyka Z służąca do oszacowania nieznannej wartości parametru Θ populacji generalnej.

Rozkład estymatora – rozkład prawdopodobieństwa statystyki będącej estymatorem parametru Θ .

Parametry rozkładu estymatora – najważniejsze to wartość oczekiwana $E(Z)$ oraz wariancja $S^2(Z)$ w rozkładzie statystyki Z będącej estymatorem jakiegoś parametru Θ populacji.

Błąd przeciętny szacunku – pierwiastek z wariancji, tzn. odchylenie standardowe $S(Z)$ w rozkładzie estymatora Z , za pomocą którego szacuje się parametr Θ populacji generalnej.

Estymacja punktowa – metoda szacunku nieznannej wartości parametru Θ populacji za pomocą jednej liczby otrzymanej z wyników próby losowej, polegająca na tym, że jako wartość parametru Θ przyjmuje się wartość estymatora Z tego parametru, otrzymaną z danej, n -elementowej próby losowej.

Estymacja przedziałowa – estymacja parametru Θ polegająca na budowaniu przedziału liczbowego, tzw. przedziału ufności dla tego parametru, w którym z określonym z góry prawdopodobieństwem bliskim jedności zawierać się będzie wartość szacowanego parametru.

Przedział ufności – losowy przedział wyznaczony za pomocą rozkładu estymatora, mający tę własność, że z dużym, z góry zadany, prawdopodobieństwem pokrywa wartość szacowanego parametru Θ . Przedział ufności można zapisać w postaci: $P\{a < \Theta < b\} = 1 - \alpha$, gdzie a , b stanowią odpowiednio dolną i górną granicę (końce) przedziału ufności, a prawdopodobieństwo $1 - \alpha$ jest dane z góry.

Współczynnik ufności – prawdopodobieństwo $1 - \alpha$ występujące z prawej strony wzoru na przedział ufności, a oznaczające prawdopodobieństwo, z jakim parametr Θ jest pokryty tym przedziałem.

Hipoteza statystyczna – każdy sąd (przypuszczenie, założenie) dotyczące populacji generalnej wydany na podstawie jej części (próby).

Weryfikacja (testowanie) hipotez statystycznych – sprawdzenie określonych przypuszczeń (założeń) dotyczących parametrów lub rozkładów populacji generalnej na podstawie wyników z próby.

Hipoteza parametryczna – hipoteza statystyczna precyzująca wartość parametru w rozkładzie populacji generalnej znanego typu.

Hipoteza nieparametryczna – hipoteza statystyczna precyzująca typ rozkładu populacji generalnej.

Hipoteza zerowa – podstawowa hipoteza statystyczna sprawdzana danym testem, oznaczana zwykle symbolem H_0 .

Hipoteza alternatywna – hipoteza statystyczna, którą jesteśmy skłonni przyjąć, jeśli odrzucimy hipotezę zerową. Oznaczamy ją zwykle H_1 .

Błąd I rodzaju – możliwy do popełnienia przy weryfikacji hipotezy statystycznej błąd polegający na odrzuceniu testowanej hipotezy prawdziwej.

Poziom istotności α – prawdopodobieństwo popełnienia błędu I rodzaju w postępowaniu testującym hipotezę. Poziom istotności przyjmujemy z góry, zazwyczaj jako małe prawdopodobieństwo. Wartości α są bliskie zera i na ogół równe: 0,01; 0,05; 0,1. Odrzucenie testowanej hipotezy na poziomie istotności $\alpha = 0,05$ oznacza, że ryzyko popełnienia błędu I rodzaju przy tej decyzji wynosi 5% (co najwyżej 5 na 100 takich decyzji będzie błędnych).

Błąd II rodzaju – polega na przyjęciu hipotezy zerowej, gdy jest ona fałszywa. Prawdopodobieństwo popełnienia błędu II rodzaju oznaczamy literą β . Dobry test statystyczny powinien mieć tę własność, że β jest bliskie zeru.

Test statystyczny – reguła postępowania, która na podstawie wyników z próby ma doprowadzić do decyzji przyjęcia lub odrzucenia postawionej hipotezy statystycznej. Za pomocą testu weryfikujemy hipotezę statystyczną.

Moc testu – prawdopodobieństwo podjęcia prawidłowej decyzji przy weryfikacji hipotezy danym testem, a polegającej na odrzuceniu testowanej hipotezy fałszywej.

Test istotności – najczęściej używany w praktyce statystycznej typ testu, pozwalający na odrzucenie hipotezy z małym ryzykiem popełnienia błędu (mierzonym poziomem istotności α). Ponieważ w teście istotności uwzględnia się jedynie błąd I rodzaju, a nie rozpatruje szansy popełnienia błędu II rodzaju, to w wyniku tego testu możliwa jest decyzja odrzucenia hipotezy zerowej lub nie ma podstaw do jej odrzucenia (nie oznacza to jej przyjęcia).

Parametryczny test istotności – test istotności weryfikujący hipotezę zerową precyzującą wartość parametru w ustalonym typie rozkładu populacji generalnej.

Nieparametryczny test istotności – test istotności dla hipotezy zerowej precyzującej ogólny typ, postać rozkładu populacji generalnej.

Obszar krytyczny testu – podzbiór przestrzeni próby o tej własności, że jeżeli otrzymamy w próbie punkt przestrzeni próby należący do tego podzbioru, to podejmujemy decyzję o odrzuceniu hipotezy zerowej.

Obszar krytyczny testu dwustronny – obszar krytyczny złożony z dwu rozłącznych podzbiorów przestrzeni próby, wyznaczony najczęściej symetrycznie, w rozkładzie odpowiedniej statystyki. Testu dwustronnego używamy, gdy hipoteza alternatywna (parametryczna) występuje w postaci nierówności typu \neq .

Obszar krytyczny testu jednostronny – w zależności od hipotezy alternatywnej może być lewostronny lub prawostronny. Jest to obszar krytyczny złożony z jednego podzbioru przestrzeni próby, wybranego z jednej strony w rozkładzie odpowiedniej statystyki. Testu z jednostronnym obszarem krytycznym używa się zwykle wtedy, gdy hipoteza alternatywna występuje w postaci nierówności typu „ $<$ ” lub „ $>$ ”.

ZADANIA DO SAMODZIELNEGO ROZWIĄZANIA

Zadanie 1

Badanie ankietowe przeprowadzone wśród 500 rodzin w pewnym mieście dostarczyło informacji o liczbie posiadanych dzieci

Liczba dzieci	Liczba rodzin
0	20
1	45
2	265
3	93
4 i więcej	77
Razem	500

- a) obliczyć prawdopodobieństwa zajścia poszczególnych zdarzeń losowych,
b) jakie jest prawdopodobieństwo, że losowo wybrana rodzina posiada 4 lub więcej dzieci?
c) jakie jest prawdopodobieństwo, że losowo wybrana rodzina ma co najwyżej dwoje dzieci?

Odp.

a)

Liczba dzieci	Liczba rodzin
0	0,040
1	0,090
2	0,530
3	0,186
4 i więcej	0,154
Razem	1,000

- b) 0,154.
c) 0,87.

Zadanie 2

Prawdopodobieństwo zdania egzaminu ze statystyki przez studenta I roku ekonomii wynosi 0,75, zaś z matematyki 0,55. Prawdopodobieństwo zdania obu egzaminów wynosi 0,6. Jakie jest prawdopodobieństwo, że student ten zda przynajmniej jeden z egzaminów?

Odp. 0,7.

Zadanie 3

W firmie produkującej zabawki 5% mężczyzn osiągnęło wiek powyżej 50 lat, zaś 0,25% wśród kobiet. Wybrano losowo jedną osobę spośród 60 mężczyzn i 400 kobiet pracujących w tej firmie.

- a) jakie jest prawdopodobieństwo, że wylosowaną osobą będzie pracownik w wieku powyżej 50 lat?
- b) jakie jest prawdopodobieństwo wylosowania mężczyzny pod warunkiem, że będzie on osobą powyżej 50 lat?

Odp.

- a) 0,0087.
- b) 0,75.

Zadanie 4

Na poligonie artylerzystów odbywa się strzelanie do ruchomej makiety. Trzech działonowych mierzy jednocześnie do przesuwanego się celu. Ich umiejętności określa prawdopodobieństwo trafienia do celu, które wynosi odpowiednio: 0,8 0,6 0,5. Działonowi oddali po jednym strzale. Jakie jest prawdopodobieństwo, że:

- a) makieta została trafiona dokładnie raz,
- b) makieta została trafiona dokładnie dwa razy,
- c) makieta została trafiona dokładnie trzy razy?

Odp.

- a) 0,26, b) 0,46, c) 0,24.

Zadanie 5

W pewnym kurniku było 5 kur wysiadujących pisklęta. Okazało się, że w poszczególnych gniazdach wykluły się kogutki w następującej proporcji: 10:15:25:30:20. Niech X będzie zmienną losową, którą jest liczba kogutków w każdym z gniazd. Podaj rozkład tak określonej zmiennej losowej.

Odp.

x_i	1	2	3	4	5
p_i	0,10	0,15	0,25	0,30	0,20

Zauważamy, że $\sum_{i=1}^5 p_i = 1$.

Zadanie 6

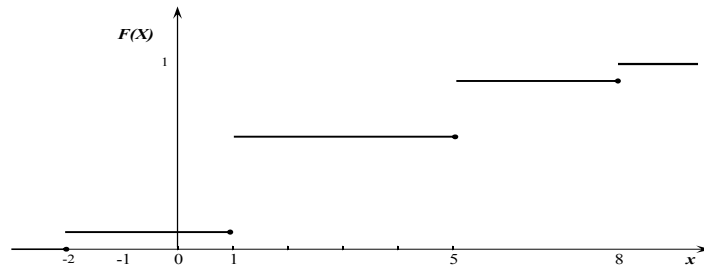
Zmienna losowa X ma następujący rozkład prawdopodobieństwa:

$X = k$	-2	1	5	8
$P(X = k)$	0,1	0,5	0,3	0,1

Wyznaczyć dystrybuantę tej zmiennej losowej oraz narysować jej wykres.

Odp.

X	...-2	...1	...5	...8	...
$F(X)$	0	0,1	0,6	0,9	1



Zadanie 7

Wyznaczyć stałą c tak, aby funkcja

$$f(x) = \begin{cases} \frac{x}{2} & \text{dla } 0 \leq x \leq 2 \\ 0 & \text{dla pozostałych } x \end{cases}$$

była funkcją gęstości prawdopodobieństwa zmiennej losowej X . Po wyznaczeniu stałej c obliczyć wartość oczekiwaną i wariancję zmiennej X .

Odp.

$$c = 1, \quad E(X) = 4/3, \quad S^2(X) = 2/9.$$

Zadanie 8

Dana jest następująca funkcja gęstości zmiennej losowej X

$$f(x) = \begin{cases} \frac{x^2}{9} & \text{dla } 0 \leq x \leq 2 \\ 0 & \text{dla pozostałych } x \end{cases}$$

Obliczyć medianę i dominantę.

Odp.

$$M_e = 2,38, \quad M_o = 0.$$

Zadanie 9

Dana jest funkcja gęstości zmiennej losowej X :

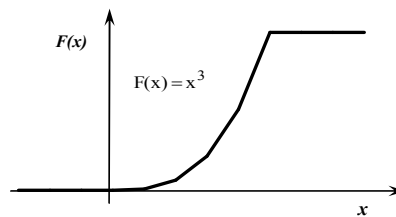
$$f(x) = \begin{cases} 0 & \text{dla } x < 0 \\ 3x^2 & \text{dla } 0 \leq x \leq 1 \\ 0 & \text{dla } x > 1 \end{cases}$$

- wyznaczyć dystrybuantę rozkładu, narysować jej wykres,
- obliczyć $P(0 < X < 0,5)$.

Odp.

a)

$$F(x) = \begin{cases} 0 & \text{dla } x < 0 \\ x^3 & \text{dla } 0 \leq x \leq 1 \\ 1 & \text{dla } x > 1 \end{cases}$$



b) $P(0 < X < 0,5) = 0,13$.

Zadanie 10

Prawdopodobieństwo, że student nie jest przygotowany do ćwiczeń wynosi 0,7. Obliczyć prawdopodobieństwo, że wśród pięciu zapytanych dwóch nie uzyskało oceny pozytywnej.

Odp.

$P(X = 2) = 0,13$.

Zadanie 11

Zmienna losowa przyjmuje trzy wartości z następującymi prawdopodobieństwami:

X = k	2	3
P(X = k)	0,5	0,3

Wyznaczyć x_3 i p_2 , wiedząc, że $E(X) = 4$.

Odp.

$p_2 = 0,2x_3 = 7,5$.

Zadanie 12

W pewnej firmie pracuje 5 automatów, które psują się niezależnie od siebie z prawdopodobieństwem $p = 1/3$. Obliczyć prawdopodobieństwo tego, że:

- a) zepsuje się jeden automat,
- b) nie zepsuje się żaden automat,
- c) zepsuje się co najmniej 1 automat,
- d) zepsuje się co najwyżej 1 automat.

Odp.

- a) 80/243, b) 32/243, c) 211/243, d) 112/243.

Zadanie 13

Prawdopodobieństwo wykrycia gruźlicy podczas prześwietleń małowzrostkowych wynosi $p = 0,6\%$. Niech X oznacza chorych na gruźlicę spośród 1 000 prześwietlonych. Obliczyć prawdopodobieństwo wykrycia gruźlicy:

- a) dokładnie w trzech przypadkach,
- b) co najmniej w trzech przypadkach,
- c) co najwyżej w trzech przypadkach.

Odp.

- a) 0,09, b) 0,94, c) 0,15.

Zadanie 14

Obliczyć prawdopodobieństwo średniego kursu euro (w zł) w kantorach przy założeniu, że kurs euro w kantorach ma rozkład jednostajny w przedziale $\langle 2,9; 3,0 \rangle$.

Odp.

2,95.

Zadanie 15

Zmienna losowa ma rozkład normalny $N(3,4)$. Obliczyć następujące prawdopodobieństwa:

- a) $P(X < -1)$, b) $P(2 < X < 7)$, c) $P(X > 8)$.

Odp.

- a) 0,1587, b) 0,4401, c) 0,1057.

Zadanie 16

Funkcja gęstości zmiennej losowej X ma postać:

$$f(x) = \begin{cases} \frac{x^2}{9} & \text{dla } 0 \leq x \leq 2 \\ 0 & \text{dla pozostałych } x \end{cases}$$

Obliczyć wartość oczekiwaną, wariancję oraz medianę.

Odp.

$$E(X) = 2,25, \quad S^2(X) = 0,34, \quad M_e = 2,38.$$

Zadanie 17

Wykonujemy 500 doświadczeń zgodnie ze schematem Bernoulliego. Prawdopodobieństwo sukcesu wynosi $p = 0,1$ w każdym doświadczeniu. Obliczyć prawdopodobieństwo tego, że częstość sukcesu odchyli się od 0,1 o mniej niż o 0,025.

Odp.

0,9372.

Zadanie 18

Z populacji studentów przyjętych na pewną uczelnię wylosowano niezależnie 20, których średni wzrost wynosił 175 cm, a odchylenie standardowe 4 cm. Oszacować przedziałowo średni wzrost dla całej populacji studentów z prawdopodobieństwem 0,90 przy założeniu, że rozkład wzrostu w populacji jest normalny.

Odp.

$$173,5 < m < 176,5.$$

Zadanie 19

Czas oczekiwania na mieszkanie (w miesiącach) w pewnej firmie developerskiej kształtował się następująco: 4,9; 5,1; 4,8; 4,9; 5,3. Oszacować przedziałowo wariancję czasu oczekiwania dla ogółu oczekujących w tej firmie, przyjmując współczynnik ufności na poziomie 0,98.

Odp.

$$0,012 < \sigma^2 < 0,546.$$

Zadanie 20

W pewnym mieście ogłoszono nabór na szkolenie w boksie. W celu wstępnej oceny, które kategorie wagowe cieszą się największym zainteresowaniem kandydatów wylosowano i zważono 120 spośród wszystkich potencjalnych uczestników kursu. Okazało się, że ich średnia waga wynosi 95 kg, a odchylenie standardowe 10 kg. Oszacować przedziałowo odchylenie standardowe w rozkładzie wagi wszystkich potencjalnych uczestników szkolenia, przyjmując poziom ufności 0,90.

Odp.

$$8,9 < \sigma < 11,1.$$

Zadanie 21

W celu sprawdzenia niezawodności żarówek wylosowano 20 sztuk i ustalono, że średni czas ich świecenia wynosi 3 000 godzin, a odchylenie standardowe 25 godzin. Zakładając, że czas świecenia żarówek ma rozkład normalny zbudować przedział ufności pokrywający średni czas świecenia żarówek w całej populacji, przyjmując współczynnik ufności 0,95.

Odp.

$$2988,8 < m < 3011,2.$$

Zadanie 22

Przebadano 1 000 gospodarstw domowych wg liczby osób pod względem przeciętnych miesięcznych wydatków na kulturę i rekreację. Wyniki przedstawiono w tabeli:

Liczba osób w gospodarstwie domowym	1	2	3	4	5 i więcej
Wydatki (w zł)	96	288	404	168	44

Oszacować przedział liczbowy dla odsetka 3-osobowych gospodarstw domowych w całej populacji, przyjmując współczynnik 0,95.

Odp.

(0,374; 0,434).

Zadanie 23

Do sklepu dostarczono 1 000 sztuk jaj. W celu zbadania ich jakości wylosowano niezależnie 400 sztuk i prześwietlono. Okazało się, że 5% jaj jest zepsutych. Oszacować procent jaj złej jakości w całej dostarczonej partii z prawdopodobieństwem 0,95.

Odp.

$2,9\% < p < 7,1\%$.

Zadanie 24

Pewna firma marketingowa chce przeprowadzić badanie wydatków klientów na określone dobro konsumpcyjne. Przyjęto, że wystarcza znajomość średnich wydatków z dokładnością nie większą niż 100 zł na poziomie ufności 95%. Z obserwacji wiadomo, że wydatki odchylają się od średniej o 400 zł. Obliczyć minimalną liczebność próby.

Odp.

62.

Zadanie 25

Norma przewiduje, że waga produkowanej sztabki pewnego stopu metali powinna wynosić 70g. Wysunięto przypuszczenie, że producent zawyża wagę sztabki. W celu potwierdzenia przypuszczenia wylosowano 20 sztabek, których średnia waga wynosiła 72g, przy odchyleniu standardowym 1,3g. Zakładamy, że rozkład wagi wyrobów jest normalny. Zweryfikować hipotezę, że waga sztabki według normy i waga faktyczna są równe wobec hipotezy alternatywnej, że waga faktyczna jest większa od zakładanej. Wykonać obliczenia, przyjmując poziom istotności $\alpha = 0,10$.

Odp.

$t_e = 6,77 > t_\alpha = 1,64$.

H_0 odrzucamy.

Zadanie 26

Automat do produkcji śrub powinien produkować śruby o średnicy wyrażonej rozkładem $N(m,0,2)$. Podczas kontroli technicznej pobrano losowo 20 śrub i otrzymano $s = 0,4$ mm. Czy na poziomie istotności $\alpha = 0,10$ można twierdzić, że wariancja średnicy śrub odbiega od założonego standardu?

Odp.

$H_0: \sigma^2 = 0,04$ wobec $H_1: \sigma^2 < 0,04$, stąd $\chi_e^2 = 80 > \chi_\alpha^2 = 10,12$.

H_0 odrzucamy, automat jest rozregulowany.

Zadanie 27

Czy prawdą jest, że średnie miesięczne spożycie chleba w 2-osobowych gospodarstwach domowych chłopskich jest mniejsze od 11 kg/osobę, jeśli w 17-elementowej próbie takich gospodarstw domowych otrzymano średnią pieczywa 10,6 kg i odchylenie standardowe 4 kg. Wykonać obliczenia przyjmując poziom istotności $\alpha = 0,02$.

Odp.

$t_e = -0,4 > t_\alpha = -2,584$. Nie ma podstaw do odrzucenia hipotezy H_0 .

Zadanie 28

Sondaż opinii publicznej na temat frekwencji oczekiwanej w wyborach samorządowych wykazał, że w losowo pobranej próbie 2 500 osób 1 600 zamierza uczestniczyć w głosowaniu. Zweryfikować opinię, że 60% ogółu uprawnionych do głosowania zamierza wziąć udział w wyborach do samorządu. Przyjąć poziom istotności $\alpha = 0,05$.

Odp.

$t_e = 4,08$. Dla $\alpha = 0,05$ H_0 należy odrzucić.

Zadanie 29

Przeciętne miesięczne wydatki w gospodarstwach domowych pracownicznych w 2008 r. w pewnym mieście w Polsce wynosiły 5630 zł. Pobrano losowo próbę 150 gospodarstw i otrzymano następujący rozkład wydatków:

Wydatki (w zł)	Odsetek gospodarstw domowych
3000 – 4000	13
4000 – 5000	25
5000 – 6000	30
6000 – 7000	16
7000 – 8000	11
8000 i więcej	5
Razem	100

Czy słuszna jest opinia, że wydatki w gospodarstwach domowych w tym mieście nie różnią się istotnie między sobą? Wykonać obliczenia, przyjmując poziom istotności $\alpha = 0,10$.

Odp.

$|t_e| = 0,999 < t_\alpha = 1,64$. Nie ma podstaw do odrzucenia H_0 .

Zadanie 30

W pewnej stołówce przeprowadzono kontrolę masy porcji obiadowej mięsa, która nominalnie powinna wynosić 120 g. Losowo wybrano, a następnie zważono 10 porcji, uzyskując następujące wyniki pomiarów (w gramach): 122, 118, 115, 116, 123, 125, 116, 114, 120, 121. Na poziomie istotności $\alpha = 0,05$ zweryfikować hipotezę, że konsumenci badanej stołówki są żywieni zgodnie z recepturą. Zakłada się, że rozkład masy porcji mięsa w całej populacji jest rozkładem normalnym.

Odp.

Nie ma podstaw do odrzucenia hipotezy zerowej. Średnia waga porcji mięsa nieistotnie różni się od jej wartości nominalnej.

BIBLIOGRAFIA

1. Aczel A., *Statystyka w zarządzaniu*, PWN, Warszawa 2000.
2. Bielecki J., Jurkiewicz B., Szymanowska Z., *Zbiór zadań ze statystyki ogólnej*, PWN, Warszawa 1978.
3. Buga J., *Wstęp do statystyki matematycznej*, WSI, Radom 1989.
4. Cieciura M., Zacharski J., *Rachunek prawdopodobieństwa w ujęciu praktycznym*, Wyd. Wizja Press&IT, Warszawa 2006.
5. Cieciura M., Zacharski J., *Metody probabilistyczne w ujęciu praktycznym*, Wyd. Wizja Press&IT, Warszawa 2007.
6. Gdakowicz A., Hozer-Koćmiel M., Markowicz J., *Zastosowanie metod opisu statystycznego do badania zjawisk społeczno-ekonomicznych*, CeDeWu, Warszawa 2022.
7. Gesternkorn T., Śródka T., *Kombinatoryka i rachunek prawdopodobieństwa*, PWN, Warszawa 1972.
8. Greń J., *Statystyka matematyczna. Modele i zadania*, PWN, Warszawa 1984.
9. Kassyk-Rokicka H., *Statystyka. Zbiór zadań*, PWE, Warszawa 1996.
10. Krzysztofiak M., Urbanek D., *Metody statystyczne*, PWN, Warszawa 1977.
11. Kukuła K., *Elementy statystyki w zadaniach*, PWN, Warszawa 1998.
12. Makać W., Urbanek-Krzysztofiak D., *Metody opisu statystycznego*, Wyd. Uniwersytetu Gdańskiego, Gdańsk 1996.
13. Ostasiewicz S., Rusnak Z., Siedlecka U., *Statystyka. Elementy teorii i zadania*, AE, Wrocław 1995.
14. Olbrych B., Sagan T., *Statystyka*, WSH, Radom 2003.
15. Panek T., (red), *Statystyka społeczna*, PWE, Warszawa 2007.
16. Olbrych B., *Wnioskowanie statystyczne w przykładach*, Wyd. PR, Radom 2009.
17. Paradysz J., *Statystyka w przykładach i zadaniach*, AE, Poznań 1995.
18. Sobczyk M., *Statystyka*, Wyd. UMCS, Lublin 1994.
19. Sobczyk M., *Statystyka*, PWN, Warszawa 1997.
20. Starzyńska W., *Statystyka praktyczna*, PWN, Warszawa 2000.

TABLICE STATYSTYCZNE

Gęstość rozkładu normalnego

u	0,00	0,01	0,02	0,03	0,04	0,05	0,06	0,07	0,08	0,09	u
0,0	,3989	,3989	,3989	,3988	,3986	,3984	,3982	,3980	,3977	,3973	0,0
0,1	,3970	,3965	,3961	,3956	,3951	,3945	,3939	,3932	,3925	,3918	0,1
0,2	,3910	,3902	,3894	,3885	,3876	,3867	,3857	,3847	,3836	,3825	0,2
0,3	,3814	,3802	,3790	,3778	,3765	,3752	,3739	,3725	,3712	,3697	0,3
0,4	,3683	,3668	,3653	,3637	,3621	,3605	,3589	,3572	,3555	,3538	0,4
0,5	,3521	,3503	,3485	,3467	,3448	,3429	,3410	,3391	,3372	,3352	0,5
0,6	,3332	,3312	,3292	,3271	,3251	,3230	,3209	,3187	,3166	,3144	0,6
0,7	,3123	,3101	,3079	,3056	,3034	,3011	,2989	,2966	,2943	,2920	0,7
0,8	,2897	,2874	,2850	,2827	,2803	,2780	,2756	,2732	,2709	,2685	0,8
0,9	,2661	,2637	,2613	,2589	,2565	,2541	,2516	,2492	,2468	,2444	0,9
1,0	,2420	,2396	,2371	,2347	,2323	,2299	,2275	,2251	,2227	,2203	1,0
1,1	,2179	,2155	,2131	,2107	,2083	,2059	,2036	,2012	,1989	,1965	1,1
1,2	,1942	,1919	,1895	,1872	,1849	,1826	,1804	,1781	,1758	,1736	1,2
1,3	,1714	,1691	,1669	,1647	,1626	,1604	,1582	,1561	,1539	,1518	1,3
1,4	,1497	,1476	,1456	,1435	,1415	,1394	,1374	,1354	,1334	,1315	1,4
1,5	,1295	,1276	,1257	,1238	,1219	,1200	,1182	,1163	,1145	,1127	1,5
1,6	,1109	,1092	,1074	,1057	,1040	,1023	,1006	,99728	,99728	,99566	1,6
1,7	,09405	,09246	,09089	,08933	,08780	,08628	,08478	,08329	,08183	,08038	1,7
1,8	,07895	,07754	,07614	,07477	,07341	,07206	,07074	,06943	,06814	,06687	1,8
1,9	,06562	,06438	,06316	,06195	,06077	,05959	,05844	,05730	,05618	,05508	1,9
2,0	,05399	,05292	,05186	,05082	,04980	,04879	,04780	,04682	,04586	,04491	2,0
2,1	,04398	,04307	,04217	,04128	,04041	,03955	,03871	,03788	,03706	,03626	2,1
2,2	,03547	,03470	,03394	,03319	,03246	,03174	,03103	,03034	,02965	,02898	2,2
2,3	,02833	,02768	,02705	,02643	,02582	,02522	,02463	,02406	,02349	,02294	2,3
2,4	,02239	,02186	,02134	,02083	,02033	,01984	,01936	,01888	,01842	,01797	2,4

Gęstość rozkładu normalnego (ciąg dalszy)

u	0,00	0,01	0,02	0,03	0,04	0,05	0,06	0,07	0,08	0,09	u
2,5	,01753	,01709	,01667	,01625	,01585	,01545	,01506	,01468	,01431	,01394	2,5
2,6	,01358	,01323	,01289	,01256	,01223	,01191	,01160	,01130	,01100	,01071	2,6
2,7	,01042	,01014	,0 ⁹ 9871	,0 ⁹ 9606	,0 ⁹ 9347	,0 ⁹ 9094	,0 ⁸ 8846	,0 ⁸ 8605	,0 ⁸ 8370	,0 ⁸ 8140	2,7
2,8	,0 ⁷ 7915	,0 ⁷ 7697	,0 ⁷ 7483	,0 ⁷ 7274	,0 ⁷ 7071	,0 ⁶ 6873	,0 ⁶ 6679	,0 ⁶ 6491	,0 ⁶ 6307	,0 ⁶ 6127	2,8
2,9	,0 ⁵ 5953	,0 ⁵ 5782	,0 ⁵ 5616	,0 ⁵ 5454	,0 ⁵ 5296	,0 ⁵ 5143	,0 ⁴ 4993	,0 ⁴ 4847	,0 ⁴ 4705	,0 ⁴ 4567	2,9
3,0	,0 ⁴ 4432	,0 ⁴ 4301	,0 ⁴ 4173	,0 ⁴ 4049	,0 ⁴ 3928	,0 ⁴ 3810	,0 ⁴ 3695	,0 ⁴ 3584	,0 ⁴ 3475	,0 ⁴ 3370	3,0
3,1	,0 ³ 3267	,0 ³ 3167	,0 ³ 3070	,0 ³ 2975	,0 ³ 2884	,0 ³ 2794	,0 ³ 2707	,0 ³ 2623	,0 ³ 2541	,0 ³ 2461	3,1
3,2	,0 ² 2384	,0 ² 2309	,0 ² 2236	,0 ² 2165	,0 ² 2096	,0 ² 2029	,0 ² 1964	,0 ² 1901	,0 ² 1840	,0 ² 1780	3,2
3,3	,0 ¹ 1723	,0 ¹ 1667	,0 ¹ 1612	,0 ¹ 1560	,0 ¹ 1508	,0 ¹ 1459	,0 ¹ 1411	,0 ¹ 1364	,0 ¹ 1319	,0 ¹ 1275	3,3
3,4	,0 ¹ 1232	,0 ¹ 1191	,0 ¹ 1151	,0 ¹ 1112	,0 ¹ 1075	,0 ¹ 1038	,0 ¹ 1003	,0 ⁰ 9689	,0 ⁰ 9358	,0 ⁰ 9037	3,4
3,5	,0 ⁰ 8727	,0 ⁰ 8426	,0 ⁰ 8135	,0 ⁰ 7853	,0 ⁰ 7581	,0 ⁰ 7317	,0 ⁰ 7061	,0 ⁰ 6814	,0 ⁰ 6575	,0 ⁰ 6343	3,5
3,6	,0 ⁰ 6119	,0 ⁰ 5902	,0 ⁰ 5693	,0 ⁰ 5490	,0 ⁰ 5294	,0 ⁰ 5105	,0 ⁰ 4921	,0 ⁰ 4744	,0 ⁰ 4573	,0 ⁰ 4408	3,6
3,7	,0 ⁰ 4248	,0 ⁰ 4093	,0 ⁰ 3944	,0 ⁰ 3800	,0 ⁰ 3661	,0 ⁰ 3526	,0 ⁰ 3396	,0 ⁰ 3271	,0 ⁰ 3149	,0 ⁰ 3032	3,7
3,8	,0 ⁰ 2919	,0 ⁰ 2810	,0 ⁰ 2705	,0 ⁰ 2604	,0 ⁰ 2506	,0 ⁰ 2411	,0 ⁰ 2320	,0 ⁰ 2232	,0 ⁰ 2147	,0 ⁰ 2065	3,8
3,9	,0 ⁰ 1987	,0 ⁰ 1910	,0 ⁰ 1837	,0 ⁰ 1766	,0 ⁰ 1698	,0 ⁰ 1633	,0 ⁰ 1569	,0 ⁰ 1508	,0 ⁰ 1449	,0 ⁰ 1393	3,9
4,0	,0 ⁰ 1338	,0 ⁰ 1286	,0 ⁰ 1235	,0 ⁰ 1186	,0 ⁰ 1140	,0 ⁰ 1094	,0 ⁰ 1051	,0 ⁰ 1009	,0 ⁰ 9687	,0 ⁰ 9299	4,0
4,1	,0 ⁰ 8926	,0 ⁰ 8567	,0 ⁰ 8222	,0 ⁰ 7890	,0 ⁰ 7570	,0 ⁰ 7263	,0 ⁰ 6967	,0 ⁰ 6683	,0 ⁰ 6410	,0 ⁰ 6147	4,1
4,2	,0 ⁰ 5894	,0 ⁰ 5652	,0 ⁰ 5418	,0 ⁰ 5194	,0 ⁰ 4979	,0 ⁰ 4772	,0 ⁰ 4573	,0 ⁰ 4382	,0 ⁰ 4199	,0 ⁰ 4023	4,2
4,3	,0 ⁰ 3854	,0 ⁰ 3691	,0 ⁰ 3535	,0 ⁰ 3386	,0 ⁰ 3242	,0 ⁰ 3104	,0 ⁰ 2972	,0 ⁰ 2845	,0 ⁰ 2723	,0 ⁰ 2606	4,3
4,4	,0 ⁰ 2494	,0 ⁰ 2387	,0 ⁰ 2284	,0 ⁰ 2185	,0 ⁰ 2090	,0 ⁰ 1999	,0 ⁰ 1912	,0 ⁰ 1829	,0 ⁰ 1749	,0 ⁰ 1672	4,4
4,5	,0 ⁰ 1528	,0 ⁰ 1528	,0 ⁰ 1461	,0 ⁰ 1396	,0 ⁰ 1334	,0 ⁰ 1275	,0 ⁰ 1218	,0 ⁰ 1164	,0 ⁰ 1112	,0 ⁰ 1062	4,5
4,6	,0 ⁰ 1014	,0 ⁰ 9684	,0 ⁰ 9248	,0 ⁰ 8830	,0 ⁰ 8430	,0 ⁰ 8047	,0 ⁰ 7681	,0 ⁰ 7331	,0 ⁰ 6996	,0 ⁰ 6676	4,6
4,7	,0 ⁰ 6370	,0 ⁰ 6077	,0 ⁰ 5797	,0 ⁰ 5530	,0 ⁰ 5274	,0 ⁰ 5030	,0 ⁰ 4796	,0 ⁰ 4573	,0 ⁰ 4360	,0 ⁰ 4156	4,7
4,8	,0 ⁰ 3961	,0 ⁰ 3775	,0 ⁰ 3598	,0 ⁰ 3428	,0 ⁰ 3267	,0 ⁰ 3112	,0 ⁰ 2965	,0 ⁰ 2824	,0 ⁰ 2690	,0 ⁰ 2561	4,8
4,9	,0 ⁰ 2439	,0 ⁰ 2322	,0 ⁰ 2211	,0 ⁰ 2105	,0 ⁰ 2003	,0 ⁰ 1907	,0 ⁰ 1814	,0 ⁰ 1727	,0 ⁰ 1643	,0 ⁰ 1563	4,9

Dystrybuanta rozkładu normalnego

$$\Phi(t) = \frac{1}{\sqrt{2\pi}} \int_0^t e^{-\frac{t^2}{2}} dt$$

t	Φ(t)	t	Φ(t)	t	Φ(t)
0,00	0,0000	1,00	0,3413	2,00	0,4772
0,05	0,0199	1,05	0,3531	2,05	0,4798
0,10	0,0398	1,10	0,3643	2,10	0,4821
0,15	0,0596	1,15	0,3749	2,15	0,4842
0,20	0,0793	1,20	0,3849	2,20	0,4861
0,25	0,0987	1,25	0,3944	2,25	0,4878
0,30	0,1179	1,30	0,4032	2,30	0,4893
0,35	0,1368	1,35	0,4115	2,35	0,4906
0,40	0,1554	1,40	0,4192	2,40	0,4918
0,45	0,1736	1,45	0,4265	2,45	0,4929
0,50	0,1915	1,50	0,4332	2,50	0,4938
0,55	0,2088	1,55	0,4394	2,55	0,4946
0,60	0,2257	1,60	0,4452	2,60	0,4953
0,65	0,2422	1,65	0,4505	2,65	0,4960
0,70	0,2580	1,70	0,4554	2,70	0,4965
0,75	0,2734	1,75	0,4599	2,75	0,4970
0,80	0,2881	1,80	0,4641	2,80	0,4974
0,85	0,3023	1,85	0,4678	2,85	0,4978
0,90	0,3159	1,90	0,4713	2,90	0,4981
0,95	0,3289	1,95	0,4744	2,95	0,4984
				3,00	0,4987

Rozkład χ^2
 $P(\chi^2 > \chi^2_\alpha)$

k \ P	0,99	0,95	0,90	0,80	0,50	0,30	0,20	0,10	0,02	0,01
1	0,000	0,004	0,016	0,064	0,455	1,074	1,642	2,706	5,412	6,635
2	0,020	0,103	0,211	0,446	1,386	2,408	3,219	4,605	7,824	9,210
3	0,115	0,352	0,584	1,005	2,366	3,665	4,642	6,251	9,837	11,345
4	0,297	0,711	1,064	1,649	3,357	4,878	5,989	7,779	11,668	13,277
5	0,554	1,145	1,610	2,343	4,351	6,064	7,289	9,236	13,388	15,086
6	0,872	1,635	2,204	3,070	5,348	7,231	8,558	10,645	15,033	16,812
7	1,239	2,167	2,833	3,822	6,346	8,383	9,803	12,017	16,622	18,475
8	1,647	2,733	3,490	4,594	7,344	9,524	11,030	13,362	18,168	20,090
9	2,088	3,325	4,168	5,380	8,343	10,656	12,242	14,684	19,679	21,666
10	2,558	3,940	4,865	6,179	9,342	11,781	13,442	15,987	21,161	23,209
11	3,053	4,575	5,578	6,989	10,341	12,899	14,631	17,275	22,618	24,725
12	3,571	5,226	6,304	7,807	11,340	14,011	15,812	18,549	24,054	26,217
13	4,107	5,892	7,041	8,634	12,340	15,119	16,985	19,812	25,471	27,688
14	4,660	6,571	7,790	9,467	13,339	16,222	18,151	21,064	26,873	29,141
15	5,229	7,261	8,547	10,307	14,339	17,322	19,311	22,307	28,259	30,578
16	5,812	7,962	9,312	11,152	15,338	18,418	20,465	23,542	29,633	32,000
17	6,408	8,672	10,085	12,002	16,338	19,511	21,615	24,769	30,995	33,409
18	7,015	9,390	10,865	12,857	17,338	20,601	22,760	25,989	32,346	34,805
19	7,633	10,117	11,651	13,716	18,338	21,689	23,900	27,204	33,687	36,191
20	8,260	10,851	12,443	14,578	19,337	22,775	25,038	28,412	35,020	37,566
21	8,897	11,591	13,240	15,445	20,337	23,858	26,171	29,615	36,343	38,932
22	9,542	12,338	14,041	16,314	21,337	24,939	27,301	30,813	37,659	40,289
23	10,196	13,091	14,848	17,187	22,337	26,018	28,429	32,007	38,968	41,638
24	10,856	13,848	15,659	18,062	23,337	27,096	29,553	33,196	40,270	42,980
25	11,524	14,611	16,473	18,940	24,337	28,172	30,675	34,382	41,566	44,314
26	12,198	15,379	17,292	19,820	25,336	29,246	31,795	35,563	42,856	45,642
27	12,878	16,151	18,114	20,703	26,336	30,319	32,912	36,741	44,140	46,963
28	13,565	16,928	18,939	21,588	27,336	31,391	34,027	37,916	45,419	48,278
29	14,256	17,708	19,768	22,475	28,336	32,461	35,139	39,087	46,693	49,588
30	14,953	18,493	20,599	23,364	29,336	33,530	36,250	40,256	47,962	50,892

Rozkład Studenta

$$P(|t| > t_{\alpha})$$

n	P	0,9	0,8	0,7	0,5	0,3	0,2	0,1	0,05	0,02	0,01
1		0,158	0,325	0,510	1,000	1,963	3,078	6,314	12,706	31,821	63,656
2		0,142	0,289	0,445	0,816	1,386	1,886	2,920	4,303	6,965	9,925
3		0,137	0,277	0,424	0,765	1,250	1,638	2,353	3,182	4,541	5,841
4		0,134	0,271	0,414	0,741	1,190	1,533	2,132	2,776	3,747	4,604
5		0,132	0,267	0,408	0,727	1,156	1,476	2,015	2,571	3,365	4,032
6		0,131	0,265	0,404	0,718	1,134	1,440	1,943	2,447	3,143	3,707
7		0,130	0,263	0,402	0,711	1,119	1,415	1,895	2,365	2,998	3,499
8		0,130	0,262	0,399	0,706	1,108	1,397	1,860	2,306	2,896	3,355
9		0,129	0,261	0,398	0,703	1,100	1,383	1,833	2,262	2,821	3,250
10		0,129	0,260	0,397	0,700	1,093	1,372	1,812	2,228	2,764	3,169
11		0,129	0,260	0,396	0,697	1,088	1,363	1,796	2,201	2,718	3,106
12		0,128	0,259	0,395	0,695	1,083	1,356	1,782	2,179	2,681	3,055
13		0,128	0,259	0,394	0,694	1,079	1,350	1,771	2,160	2,650	3,012
14		0,128	0,258	0,393	0,692	1,076	1,345	1,761	2,145	2,624	2,977
15		0,128	0,258	0,393	0,691	1,074	1,341	1,753	2,131	2,602	2,947
16		0,128	0,258	0,392	0,690	1,071	1,337	1,746	2,120	2,583	2,921
17		0,128	0,257	0,392	0,689	1,069	1,333	1,740	2,110	2,567	2,898
18		0,127	0,257	0,392	0,688	1,067	1,330	1,734	2,101	2,552	2,878
19		0,127	0,257	0,391	0,688	1,066	1,328	1,729	2,093	2,539	2,861
20		0,127	0,257	0,391	0,687	1,064	1,325	1,725	2,086	2,528	2,845
21		0,127	0,257	0,391	0,686	1,063	1,323	1,721	2,080	2,518	2,831
22		0,127	0,256	0,390	0,686	1,061	1,321	1,717	2,074	2,508	2,819
23		0,127	0,256	0,390	0,685	1,060	1,319	1,714	2,069	2,500	2,807
24		0,127	0,256	0,390	0,685	1,059	1,318	1,711	2,064	2,492	2,797
25		0,127	0,256	0,390	0,684	1,058	1,316	1,708	2,060	2,485	2,787
26		0,127	0,256	0,390	0,684	1,058	1,315	1,706	2,056	2,479	2,779
27		0,127	0,256	0,389	0,684	1,057	1,314	1,703	2,052	2,473	2,771
28		0,127	0,256	0,389	0,683	1,056	1,313	1,701	2,048	2,467	2,763
29		0,127	0,256	0,389	0,683	1,055	1,311	1,699	2,045	2,462	2,756
30		0,127	0,256	0,389	0,683	1,055	1,310	1,697	2,042	2,457	2,750
40		0,126	0,255	0,388	0,681	1,050	1,303	1,684	2,021	2,423	2,704
60		0,126	0,254	0,387	0,679	1,045	1,296	1,671	2,000	2,390	2,660
120		0,126	0,254	0,386	0,677	1,041	1,289	1,658	1,980	2,358	2,617

Rozkład λ Kolmogorowa (graniczny)

λ	$Q(\lambda)$	λ	$Q(\lambda)$	λ	$Q(\lambda)$
0,30	0,000 000	0,70	0,268 765	1,10	0,822 282
0,31	0,000 021	0,71	0,305 401	1,11	0,829 950
0,32	0,000 046	0,72	0,322 265	1,12	0,837 356
0,33	0,000 091	0,73	0,339 113	1,13	0,844 502
0,34	0,000 171	0,74	0,355 981	1,14	0,851 394
0,35	0,000 303	0,75	0,372 833	1,15	0,586 038
0,36	0,000 511	0,76	0,389 640	1,16	0,864 442
0,37	0,000 826	0,77	0,406 372	1,17	0,870 612
0,38	0,001 285	0,78	0,423 002	1,18	0,876 548
0,39	0,001 929	0,79	0,439 505	1,19	0,882 258
0,40	0,002 808	0,80	0,455 857	1,20	0,887 750
0,41	0,003 972	0,81	0,472 041	1,21	0,893 030
0,42	0,005 476	0,82	0,488 030	1,22	0,898 104
0,43	0,007 377	0,83	0,503 806	1,23	0,902 972
0,44	0,009 730	0,84	0,519 366	1,24	0,907 648
0,45	0,012 590	0,85	0,534 682	1,25	0,912 132
0,46	0,016 005	0,86	0,549 744	1,26	0,916 432
0,47	0,020 022	0,87	0,564 546	1,27	0,920 556
0,48	0,024 682	0,88	0,279 070	1,28	0,924 505
0,49	0,030 0170	0,89	0,593 316	1,29	0,928 288
0,50	0,036 055	0,90	0,607 270	1,30	0,931 908
0,51	0,042 814	0,91	0,620 928	1,31	0,935 370
0,52	0,050 306	0,92	0,634 286	1,32	0,938 682
0,53	0,058 534	0,93	0,647 338	1,33	0,941 848
0,54	0,067 497	0,94	0,660 082	1,34	0,944 872
0,55	0,077 183	0,95	0,672 516	1,35	0,947 756
0,56	0,087 577	0,96	0,684 636	1,36	0,950 512
0,57	0,098656	0,97	0,696 444	1,37	0,953 142
0,58	0,110 395	0,98	0,707 940	1,38	0,955 650
0,59	0,122 760	0,99	0,719 126	1,39	0,958 040
0,60	0,135 718	1,00	0,730 000	1,40	0,960 318
0,61	0,149 229	1,01	0,740 566	1,41	0,962 486
0,62	0,163 225	1,02	0,750 826	1,42	0,964 552
0,63	0,177 753	1,03	0,760 780	1,43	0,966 516
0,64	0,192 677	1,04	0,770 434	1,44	0,968 382
0,65	0,207 987	1,05	0,779 794	1,45	0,970 158
0,66	0,223 637	1,06	0,788 860	1,46	0,971 846
0,67	0,239 582	1,07	0,797 636	1,47	0,973 448
0,68	0,255 780	1,08	0,806 128	1,48	0,974 970
0,69	0,272 189	1,09	0,814 342	1,49	0,976 412

Rozkład λ Kolmogorowa (graniczny) cd.

λ	$Q(\lambda)$	λ	$Q(\lambda)$	λ	$Q(\lambda)$
1,50	0,977 782	1,90	0,998 536	2,30	0,999 949
1,51	0,979 080	1,91	0,998 644	2,31	0,999 954
1,52	0,980 310	1,92	0,998 744	2,32	0,999 958
1,53	0,981 476	1,93	0,998 837	2,33	0,999 962
1,54	0,982 578	1,94	0,998 924	2,34	0,999 965
1,55	0,983 622	1,95	0,999 004	2,35	0,999 968
1,56	0,984 610	1,96	0,999 079	2,36	0,999 970
1,57	0,985 544	1,97	0,999 149	2,37	0,999 973
1,58	0,986 426	1,98	0,999 213	2,38	0,999 976
1,59	0,987 260	1,99	0,999 273	2,39	0,999 978
1,60	0,988 048	2,00	0,999 329	2,40	0,999 980
1,61	0,988 791	2,01	0,999 380	2,41	0,999 982
1,62	0,989 492	2,02	0,999 428	2,42	0,999 984
1,63	0,990 154	2,03	0,999 474	2,43	0,999 986
1,64	0,990 777	2,04	0,999 516	2,44	0,999 987
1,65	0,991 364	2,05	0,999 552	2,45	0,999 988
1,66	0,991 917	2,06	0,999 588	2,46	0,999 989
1,67	0,992 438	2,07	0,999 620	2,47	0,999 990
1,68	0,992 928	2,08	0,999 650	2,48	0,999 991
1,69	0,993 389	2,09	0,999 680	2,49	0,999 992
1,70	0,993 828	2,10	0,999 705	2,50	0,999 9925
1,71	0,994 230	2,11	0,999 723	2,55	0,999 9956
1,72	0,994 612	2,12	0,999 750	2,60	0,999 9974
1,73	0,994 972	2,13	0,999 770	2,65	0,999 9984
1,74	0,995 309	2,14	0,999 790	2,70	0,999 9990
1,75	0,995 625	2,15	0,999 806	2,75	0,999 9994
1,76	0,995 922	2,16	0,999 822	2,80	0,999 9997
1,77	0,996 200	2,17	0,999 838	2,85	0,999 99982
1,78	0,996 460	2,18	0,999 852	2,90	0,999 99990
1,79	0,996 704	2,19	0,999 864	2,95	0,999 99994
1,80	0,996 932	2,20	0,999 874	3,00	0,999 99997
1,81	0,997 146	2,21	0,999 886		
1,82	0,997 346	2,22	0,999 896		
1,83	0,997 533	2,23	0,999 904		
1,84	0,997 707	2,24	0,999 912		
1,85	0,997 870	2,25	0,999 920		
1,86	0,998 023	2,26	0,999 926		
1,87	0,998 145	2,27	0,999 934		
1,88	0,998 297	2,28	0,999 940		
1,89	0,998 421	2,29	0,999 944		