

2026

Универзитет „Св. Кирил и Методиј“, Скопје

Зоран Мисајлески

Анета Велкоска

Веројатност и статистика

Универзитет „Св. Кирил и Методиј“, Скопје

Зоран Мисајлески
Анета Велкоска

Веројатност и статистика

Скопје, 2026 г.

Издавач:

Универзитет „Св. Кирил и Методиј“ во Скопје
бул. Гоце Делчев бр. 9, 1000 Скопје
www.ukim.edu.mk

Уредник за издавачка дејност на УКИМ:

проф. д-р Биљана Ангелова, ректор

Уредник на публикацијата:

проф. д-р Зоран Мисајлески, редовен професор на ГФ, УКИМ, Скопје
проф. д-р Анета Велкоска, редовен професор на УИИТ, Охрид

Рецензенти:

проф. д-р Даниел Велинов, редовен професор на ГФ, УКИМ, Скопје
проф. д-р Татјана Атанасова Пачемска, редовен професор на УГД, Штип

Техничка обработка:

проф. д-р Зоран Мисајлески, редовен професор на ГФ, УКИМ, Скопје
проф. д-р Анета Велкоска, редовен професор на УИИТ, Охрид

Лектура на македонски јазик:

доц. д-р Цутка Јованоска, доцент на ДУТ, Тетово

CIP - Каталогизација во публикација

Национална и универзитетска библиотека „Св. Климент Охридски“, Скопје

519.2(075.8)

МИСАЈЛЕСКИ, Зоран

Веројатност и статистика [Електронски извор] / Зоран Мисајлески,
Анета Велкоска. - Скопје: Универзитет „Св. Кирил и Методиј“, 2026

Начин на пристапување (URL): <https://ukim.edu.mk/e-book/00161> . - Текст
во PDF формат, содржи 320 стр. - Наслов преземен од екранот. - Опис на
изворот на ден 30.12.2025. - Библиографија: стр. 313-316

ISBN 978-9989-43-542-3

1. Велкоска, Анета [автор]

а) Веројатност и статистика -- Високошколски учебници

COBISS.MK-ID 67855877

Предговор

Учебникот „Веројатност и статистика“ целосно ја покрива содржината по предметот Веројатност и статистика (2 + 2) кој се изучува на Градежниот факултет на првиот циклус на студиската програма Геодезија. Учебникот можат да го користат и студентите кои го слушаат предметот Веројатност и статистика (1 + 1) и делот од Веројатност и статистика од предметите Одбрани поглавја од математиката (4 + 4) на вториот циклус студии и Одбрани поглавја од математиката (2 + 2) на третиот циклус со фонд 1 + 1 на Градежниот факултет; предметот Веројатност и статистика (3 + 2) на првиот циклус студии и делот од предметот Одлучување засновано на податоци на вториот циклус студии на студиската програма Дигитална бизнис-аналитика на Универзитетот по информатички науки „Св. Апостол Павле“, Охрид; како и од други факултети во кои се изучува предмет што има иста или слична наставна содржина.

Материјалот од првиот дел е од темата Веројатност. Поделен е на 10 глави: 1. Комбинаторика; 2. Простор на веројатност; 3. Условна веројатност; 4. Биномна веројатност; 5. Случајни променливи; 6. Математичко очекување и дисперзија; 7. Примери на распределби; 8. Дводимензионални случајни променливи; 9. Функции од случајни променливи; 10. Закон на големите броеви, централна гранична теорема.

Материјалот од делот Статистика е поделен на 4 глави: 1. Статистика; 2. Статистички оценки; 3. Статистички хипотези; и 4. Проста линеарна регресија и корелација.

Материјалот од темите 1-7.2.5 и 10.1.7-9 од првиот дел, 1.6, 3.4 и половина од 1.1.1-5 од вториот дел е напишан од Зоран Мисајлески, останатиот материјал е од Анета Велкоска.

Материјалот од Зоран Мисајлески е напишан во периодот од 2013 до 2024-та година. Во 2013-та авторот напиша материјали со решени задачи во ракопис за држење на вежбите, додека во 2016-та материјали обработени на компјутер за предавањата. Во 2024-та година материјалот е проширен и досреден.

Ова е прво издание на учебникот и авторите ќе им бидат благодарни на читателите кои со своите забелешки ќе придонесат во подобрувањето на квалитетот во наредните изданија.

Проф. д-р Зоран Мисајлески
Проф. д-р Анета Велкоска

Учебникот е посветен на мајка ми Верка Гелевска (1949 - 2024),
затоа што овозможи единствената грижа во животот да ми е
научно-образовната дејност

Зоран Мисајлески

I дел: Веројатност

1. КОМБИНАТОРИКА

При броењето се користат два принципа, т. е. правила, принципот на збир и принципот на производ.

Нека се дадени множествата A и B такви што $A \cap B = \emptyset$, со број на елементи $|A| = m$ и $|B| = n$, соодветно. **Принципот на производ** гласи: Бројот на различни начини да се формира подредениот пар (a, b) , $a \in A$, $b \in B$; е mn . **Принципот на збир** гласи: Бројот на различни начини од множествата A и B да се избере еден елемент е $m + n$.

1.1. ВАРИЈАЦИИ

Нека $A = \{1, 2, \dots, n\}$ е множество со n елементи.

Дефиниција 1.1.1. Варијација без повторување од n елементи, класа k , е секој елемент $(a_1, a_2, \dots, a_k)$ од Декартовиот производ A^k во кој елементите a_i , $i = 1, 2, \dots, k$; се различни.

Значи $a_i \neq a_j$ за $i \neq j$ во $(a_1, a_2, \dots, a_k) \in A^k$. Елементите a_i , $i = 1, 2, \dots, k$; ги викаме и координати.

Бројот на варијации без повторување од n елементи, класа k се означува со V_n^k .

Под терминот варијација подразбираме варијација без повторување.

Својство 1.1.2. $V_n^k = \frac{n!}{(n-k)!} = n(n-1) \dots (n-k+1)$.

Доказ. Првиот елемент од подредената k -торка $(a_1, a_2, \dots, a_k)$ може да се избере на n начини, вториот на $n-1$, ..., k -тиот на $n-k+1$ начин.

$$\underbrace{\quad \quad \quad \dots \quad \quad}_{n}.$$

Од принципот на производ следува дека:

$$V_n^k = n(n-1)\dots(n-k+1) = \frac{n!}{(n-k)!} \cdot \blacksquare$$

Значи дека бројот на варијации од n елементи, класа k ; се определува така што се множат k броеви каде што првиот множител е n , а секој следен множител е за 1 помал.

Ако бројот на елементи е едноцифрен, место како елемент од Декартовиот производ, координатите на варијацијата ќе ги пишуваме слеано, како n -цифрен број. Најпрегледен е записот во кој елементите на варијацијата ги означуваме со броеви.

Кога ги запишуваме варијациите пожелно е да имаме систем на броење за да не пропуштиме некоја од нив. Еден начин е следниов:

$$12\dots k, 12\dots(k-1)(k+1), \dots, 12\dots(k-1)n, 12\dots(k-2)\underline{k}(k-1), \\ 12\dots(k-2)\underline{k}(k+1), \dots, 12\dots(k-2)kn, 12\dots(k-2)(k+1)(k-1), \dots$$

Значи, првата варијација е претставена со последователни броеви. Потоа во следните варијации се зголемува за еден само последната цифра додека да го достигне n , па се зголемува претпоследната цифра за еден, додека последната почнува од најмалата можна вредност што не се совпаѓа со другите броеви, т. е. $k-1$ и. т. н. Ако бројот на елементи е поголем од девет, елементите во варијацијата ги одвојуваме со запирка, а варијациите со точка и запирка. Истиот договор ќе го применуваме и при останатите броења.

Задача 1.1.3. Напиши ги сите варијации без повторување на множеството $A = \{1, 2, 3, 4\}$, од класа $k = 2$. Со помош на формула, определи го нивниот број.

Решение. Имаме $n = 4$ и $k = 2$. Варијациите без повторување од 4 елементи, класа 2 се:

$$12, 13, 14, 21, 23, 24, 31, 32, 34, 41, 42 \text{ и } 43.$$

$$\text{Нивниот број е: } V_n^k = V_4^2 = 4 \cdot 3 = 12.$$

Задача 1.1.4. Напиши ги сите трицифрени броеви што ги содржат цифрите 1, 3, 5 и 7 и во кои цифрите не се повторуваат. Колку такви броеви постојат?

Решение. Секој трицифрен број формиран со помош на четири цифри кои не се повторуваат е варијација без повторување од 4 елементи, класа 3. Цифрите 1, 3, 5 и 7 одговараат на елементите 1, 2, 3 и 4; соодветно, од варијациите чии елементи се означени со последователни цифри. Затоа броевите се:

135, 137, 153, 157, 173, 175, 315, 317, 351, 357, 371, 375,
513, 517, 531, 537, 571, 573, 713, 715, 731, 735, 751 и 753.

Следува дека постојат вкупно 24 броеви што ги исполнуваат условите. За контрола, бројот на варијации може да се провери со формулата: $V_n^k = 4 \cdot 3 \cdot 2 = 24$.

Дефиниција 1.1.5. Варијација со повторување од n елементи, класа k е секој елемент $(a_1, a_2, \dots, a_k)$ од Декартовиот производ A^k .

Бројот на варијации со повторување од n елементи, класа k се означува со $\bar{V}_n^k$.

Варијациите со повторување се бројат така што се тргнува од варијацијата 11...1. Следни варијации се:

11...12, ..., 11...1n, 11...121, 11...122, ..., 11...12n, 11...131, и.т.н.

Својство 1.1.6. $\bar{V}_n^k = n^k$.

Доказ. Секој елемент $(a_1, a_2, \dots, a_k) \in A^k$ може да се избере на $\underbrace{n \cdot n \cdot \dots \cdot n}_k = n^k$ начини.

Задача 1.1.7. Напиши ги сите варијации со повторување на множеството $A = \{1, 2\}$, од класата $k = 4$. Со помош на формула, определи го нивниот број.

Решение. Имаме $n = 2$. Варијациите со повторување од 2 елемента, класа 4 се:

1111, 1112, 1121, 1122, 1211, 1212, 1221, 1222,
2111, 2112, 2121, 2122, 2211, 2212, 2221, 2222.

Нивниот број е: $\bar{V}_n^k = \bar{V}_2^4 = 2^4 = 16$. ■

Задача 1.1.8. Регистарската таблица на автомобил од Скопскиот регион содржи две фиксни букви SK, две произволни букви од англиската латиница (која има 28 букви) и четири цифри. Колку различни регистрации може да се формираат во скопскиот регион.



Цртеж 1.1.8. Изглед на регистарска таблица

Решение. Двете букви може да се изберат на $\bar{V}_{28}^2$ начини, додека четирите цифри на $\bar{V}_{10}^4$.

Според принципот на производ, вкупниот број начини е:

$$S = \bar{V}_{28}^2 \bar{V}_{10}^4 = 28^2 \cdot 10^4 = 784 \cdot 10000 = 7840000.$$

Заради прегледност, кога запишуваме големи броеви, после секои 3 цифри одејќи од десно кон лево, може да ставаме точки. Така бројот седум милиони и осумстотини четириесет илјади попрегледно ќе го прочитаеме ако го запишеме во облик 7.840.000.

Задача 1.1.8. Пет стрелци стрелаат еден по друг во 7 мети. Да се најде бројот на:

- 1) Сите можни исходи;
- 2) Сите можни исходи во кои барем двајца стрелци стрелаат во иста мета;
- 3) Сите можни исходи во кои првите двајца стрелци стрелаат во иста мета;
- 4) Сите можни исходи во кои првите двајца стрелци стрелаат во иста мета, а останатите тројца во мети, различни од метите на првите двајца.

Решение. 1) Бројот на мети е бројот на елементи $n = 7$. Од нив се бираат петорки кои ги погодиле (избрале) стрелците т.е. класата $k = 5$. Следува дека пет стрелци стрелаат во седум мети на:

$$S = \bar{V}_7^5 = 7^5 = 16807$$

начини.

2) При секој избор еден број се појавува најмалку два пати. Значи од сите варијации со повторување од 7 елементи, класа 5; ќе ги одземеме варијациите без повторување од 7 елементи класа 5. Следува:

$$S = \bar{V}_7^5 - V_7^5 = 7^5 - 7 \cdot 6 \cdot 5 \cdot 4 \cdot 3 = 16807 - 2520 = 14287.$$

3) Првите двајца стрелци стрелаат во една од седумте мети на 7 начини. Независно од нив, останатите 3 стрелци стрелаат во метите на $\bar{V}_7^3$ начини. Од принципот на производ кај броењето, следува дека бројот на можни исходи е:

$$S = 7 \cdot \bar{V}_7^3 = 7 \cdot 7^3 = 7^4 = 2401.$$

4) Првите двајца стрелци стрелаат во една од седумте мети на 7 начини. Независно од нив, останатите 3 стрелци стрелаат во преостанатите 6 мети на $\bar{V}_6^3$ начини. Затоа

$$S = 7 \cdot \bar{V}_6^3 = 7 \cdot 6^3 = 1512.$$

1.2. ПЕРМУТАЦИИ

Дефиниција 1.2.1. Пермутација без повторување од n елементи е варијација без повторување од n елементи, класа n .

Бројот на пермутации без повторување од n елементи се означува со P_n . Под терминот пермутација ќе подразбираме пермутација без повторување.

Својство 1.2.2. $P_n = n!$.

Доказ. Следува директно од својството 1.1.2 и дефиницијата 1.2.1.

Задача 1.2.3. Напиши ги сите пермутации без повторување од 3 елементи. Со помош на формула, определи го нивниот број.

Решение. Имаме $n = 3$. Пермутациите без повторување од 3 елементи, се:

123, 132, 213, 231, 312 и 321.

Нивниот број е $P_3 = 3! = 6$.

Задача 1.2.4. Колку петцифрени броеви може да се образуваат од цифрите 1, 2, 3, 4 и 5, така што цифрите да им се различни и ниту еден број да не почнува со 1 или 2.

Решение. Прв начин. На прво место може да стојат цифрите 3, 4 или 5.

$$\underline{3} \quad \underline{\quad}, \underline{4} \quad \underline{\quad}, \underline{5} \quad \underline{\quad}.$$

Независно првата цифра во бројот, останатите 4 места се пополнуваат од множество од 4 елементи (сите цифри освен првата) на P_4 начини. Следува дека бројот на бараните петоцифрени броеви е:

$$S = 3P_4 = 3 \cdot 4! = 72.$$

Втор начин. Вкупниот број на петцифрени броеви формирани со цифрите 1, 2, 3, 4 и 5 е P_5 . Од нив $2P_4$ броеви почнуваат на цифрите 1 или 2. Следува:

$$S = P_5 - 2P_4 = 5! - 2 \cdot 4! = 120 - 48 = 72.$$

Задача 1.2.5. Бројот на пермутации без повторување од n елементи спрема бројот на пермутации без повторување од $n+2$ елементи се однесува како 1:42. Пресметај го бројот на елементи.

Решение. Бидејќи $P_n = n!$ и $P_{n+2} = (n+2)!$, од условот на задачата следува дека:

$$\begin{aligned} \frac{P_n}{P_{n+2}} = \frac{1}{42} &\Leftrightarrow \frac{n!}{(n+2)!} = \frac{1}{42} \Leftrightarrow \frac{n!}{(n+2)(n+1)n!} = \frac{1}{42} \Leftrightarrow \\ \frac{1}{(n+2)(n+1)} &= \frac{1}{42} \Leftrightarrow (n+2)(n+1) = 42. \end{aligned}$$

Со проверка определуваме кој производ на последователни броеви е 42. Бидејќи $7 \cdot 6 = 42$ следува дека $n+2 = 7$ (или $n+1 = 6$) од каде што имаме $n = 5$.

Коментар. Ако равенката се развие:

$$(n+2)(n+1) = 42 \Leftrightarrow n^2 + 3n + 2 = 42 \Leftrightarrow n^2 + 3n - 40 = 0,$$

може да се реши со формулата за квадратна равенка:

$$n_{1,2} = \frac{-3 \pm \sqrt{3^2 + 4 \cdot 1 \cdot 40}}{2 \cdot 1} = \frac{-3 \pm \sqrt{169}}{2} = \frac{-3 \pm 13}{2} \Leftrightarrow n_1 = 5 \vee n_2 = -8.$$

Равенката може да се реши и со Виетовите формули:

$$(n+8)(n-5) = 0 \Leftrightarrow n = -8 \vee n = 5.$$

Бидејќи $n \in \mathbb{N}$ и во двата случаја негативното решение се отфрла, па решение е: $n = 5$.

Дефиниција 1.2.6. Пермутација со повторување од n елементи, тип $n_1, n_2, \dots, n_k$; е пермутација без повторување од n елементи во која елементите се повторуваат $n_1, n_2, \dots, n_k$ пати, соодветно.

Бидејќи збирот на елементите е n , важи:

$$n = n_1 + n_2 + \dots + n_k.$$

Бројот на пермутации со повторување од n елементи тип $n_1, n_2, \dots, n_k$ се означува со $P_n(n_1, n_2, \dots, n_k)$.

Својство 1.2.7. $P_n(n_1, n_2, \dots, n_k) = \frac{n!}{n_1! n_2! \dots n_k!}.$

Доказ. Имаме $P_n = n!$. Бидејќи $n_1! n_2! \dots n_k!$ пермутациите без повторување од n елементи определуваат една пермутација со повторување од n елементи, тип $n_1, n_2, \dots, n_k$; бројот на пермутации со повторување по $n_1! n_2! \dots n_k!$ го дава бројот на пермутации без повторување т.е.

$$n_1! n_2! \dots n_k! P_n(n_1, n_2, \dots, n_k) = P_n \Leftrightarrow P_n(n_1, n_2, \dots, n_k) = \frac{n!}{n_1! n_2! \dots n_k!}. \blacksquare$$

Задача 1.2.8. Напиши ги пермутациите што се формираат со користење само еднаш на еден од знаците: a , b , b и v . Со помош на формула, определи го нивниот број.

Решение. Пермутациите се:

$$abbb, abvb, avbb, babv, babb, bbav, \\ bbva, bvab, bvba, vabb, vbab \text{ и } vbba.$$

Нивниот број е $P_4(1,2,1) = \frac{4!}{2!} = \frac{4 \cdot 3 \cdot 2}{2} = 12$.

Задача 1.2.9. Колку различни зборови може да се формираат од секој од зборовите:

1) математика, 2) градежништво, 3) геодезија?

Решение. 1) Зборот „математика“ содржи 6 различни букви кои редоследно се јавуваат: 2, 3, 2, 1, 1 и 1 пат, т. е. 10 пати. Затоа, бројот на зборови кои може да се формираат од „математика“ е:

$$P_{10}(2,3,2,1,1,1) = \frac{10!}{2!3!2!} = \frac{10 \cdot 9 \cdot 8 \cdot 7 \cdot 6 \cdot 5 \cdot 4 \cdot 3 \cdot 2}{2 \cdot 3 \cdot 2 \cdot 2} = \frac{10 \cdot 9 \cdot 8 \cdot 7 \cdot 6 \cdot 5}{1} = 151200.$$

2) Зборот градежништво содржи дури 12 букви и сите се различни, па бројот на пермутации е:

$$P_{12} = 12! = 479001600.$$

$$3) P_9(1,2,1,1,1,1,1,1,1) = \frac{9!}{2!} = \frac{9 \cdot 8 \cdot 7 \cdot 6 \cdot 5 \cdot 4 \cdot 3 \cdot \cancel{2}}{\cancel{2}} = 181440.$$

1.3. КОМБИНАЦИИ

Нека A е множество со n елементи.

Дефиниција 1.3.1. Комбинација без повторување од n елементи, класа k ; е секое подмножество од A со k елементи.

Бројот на комбинации без повторување од n елементи, класа k се означува со C_n^k . Под комбинација ќе подразбираме комбинација без повторување.

Комбинациите без повторување ги запишуваме во следниов редослед:

$$12...k, 12...(k-1)(k+1), ..., 12...(k-1)n, 12...(k-2)k(k+1), \\ 12...(k-2)k(k+2), ..., 12...(k-2)kn, 12...(k-2)(k+1)(k+2), ...$$

Броењето е како кај варијациите, само што елементите во комбинациите без повторување се во растечки редослед.

Својство 1.3.2. $C_n^k = \binom{n}{k}$, каде што: $\binom{n}{k} = \frac{n!}{k!(n-k)!}$.

Доказ. Нека $A = \{a_1, a_2, \dots, a_n\}$. На секоја комбинација $\{a_{i_1}, a_{i_2}, \dots, a_{i_k}\}$ ѝ одговараат $k!$ варијации без повторување од n елементи, класа k . Затоа бројот на варијации е $k!$ по бројот на комбинации т.е.

$$k!C_n^k = V_n^k \Leftrightarrow k!C_n^k = \frac{n!}{(n-k)!} \Leftrightarrow C_n^k = \frac{n!}{k!(n-k)!} \Leftrightarrow C_n^k = \binom{n}{k}. \blacksquare$$

Задача 1.3.3. Определи ги сите комбинации без повторување, класа 3, на множеството $A = \{1, 2, 3, 4\}$. Со помош на формула пресметај го нивниот број.

Решение. Имаме $n = 4$. Комбинациите без повторување се:

123, 124, 134, 234.

Нивниот број е $C_4^3 = \binom{4}{3} = 4$.

Задача 1.3.4. Колкав е бројот на комбинации што треба да се уплатат за сигурно за да се добие седмица на лото во кое од 39 броеви се влечат 7.

Решение. Бројот на комбинации е:

$$C_{39}^7 = \binom{39}{7} = \frac{39 \cdot 38 \cdot 37 \cdot \cancel{36} \cdot \cancel{35} \cdot 34 \cdot 33 \cdot \cancel{32}!}{\cancel{7} \cdot \cancel{6} \cdot \cancel{5} \cdot 4 \cdot \cancel{3} \cdot \cancel{2} \cdot \cancel{32}!} = \frac{39 \cdot 38 \cdot 37 \cdot 34 \cdot 33}{4 \cdot 3 \cdot 2} = 15380937.$$

Задача 1.3.5. Реши ја равенката $C_x^3 + C_x^2 = 26(x-1)$.

Решение. Имаме:

$$C_x^3 + C_x^2 = 26(x-1) \Leftrightarrow \binom{x}{3} + \binom{x}{2} = 26(x-1) \Leftrightarrow \frac{x(x-1)(x-2)}{3!} + \frac{x(x-1)}{2!} = 26(x-1).$$

За $x \neq 1$, равенката може да се подели со $x-1$, од каде што се добива:

$$\frac{x(x-2)}{6} + \frac{x}{2} = 26 \Leftrightarrow \frac{x^2 - 2x + 3x}{6} = 26 \Leftrightarrow x^2 + x = 156 \Leftrightarrow x(x+1) = 156.$$

Со проверка добиваме дека: $12 \cdot 13 = 156$, т. е. $x = 12$.

Задача 1.3.6. Наставничкиот совет на Градежниот факултет во 2007 година се состоел од 5 студенти, 4 асистенти и 40 професори. Притоа сите професори на Градежниот факултет се членови на наставничкиот совет. На колку начини можел да се избере наставничкиот совет, ако Градежниот факултет имал 1000 студенти и 20 асистенти.

Решение. Студентите може да се изберат на C_{1000}^5 , додека асистентите на C_{20}^4 начини. Следува дека бараниот број е:

$$S = C_{1000}^5 C_{20}^4 = \frac{1000 \cdot 999 \cdot 998 \cdot 997 \cdot 996}{5 \cdot 4 \cdot 3 \cdot 2} \cdot \frac{20 \cdot 19 \cdot 18 \cdot 17}{4 \cdot 3 \cdot 2} = 50 \cdot 999 \cdot 998 \cdot 997 \cdot 166 \cdot 5 \cdot 19 \cdot 3 \cdot 17 = 39972661107219000.$$

Задача 1.3.7. Во студентскиот дом „Стив Наумов“, тројца студенти делат соба. Студентите имаат 4 шољи, 5 тацни и 6 лажици. На колку начини можат да пијат чај ако секој од нив користи 1 шоља, 1 тацна и 1 лажица.

Решение. Вкупниот број начини е:

$$S = C_4^3 C_5^3 C_6^3 = 4 \cdot \frac{5 \cdot 4}{2} \cdot \frac{6 \cdot 5 \cdot 4}{3 \cdot 2} = 4 \cdot 10 \cdot 20 = 800.$$

Задача 1.3.8. Во кутија се наоѓаат 3 бели, 4 црвени и 5 црни топчиња. Влечеме 3 топчиња на следниов начин:

- 1) одеднаш,
- 2) едно по едно без враќање,
- 3) едно по едно со враќање.

На колку начини може да се изврши изборот, така што да бидат извлечени:

- а) топчиња во која било боја,
- б) 2 црни и 1 црвено топче,
- в) топчиња во различна боја.

Решение. Бројот на избори е:

$$\text{а) 1) } S = C_{12}^3 = \frac{12 \cdot 11 \cdot 10}{3 \cdot 2} = 6 \cdot 110 = 660,$$

$$2) S = V_{12}^3 = 12 \cdot 11 \cdot 10 = 1320,$$

$$3) S = \bar{V}_{12}^3 = 12^3 = 1728;$$

$$\text{б) 1) } S = C_5^2 C_3^1 = \frac{5 \cdot 4}{2} \cdot 3 = 30,$$

$$2) S = V_5^2 V_4^1 = 5 \cdot 4 \cdot 4 = 80,$$

$$3) S = \bar{V}_5^2 \bar{V}_4^1 = 5^2 \cdot 4 = 100.$$

в) Топчињата се во различна боја ако е извлечено 1 бело, 1 црвено и 1 црно топче. Следува:

$$1) S = C_3^1 C_4^1 C_5^1 = 3 \cdot 4 \cdot 5 = 60,$$

$$2) S = V_3^1 V_4^1 V_5^1 = 3 \cdot 4 \cdot 5 = 60,$$

$$3) S = \bar{V}_3^1 \bar{V}_4^1 \bar{V}_5^1 = 3 \cdot 4 \cdot 5 = 60.$$

Задача 1.3.9. Сакаме да избереме 6 од 8 лица, при што ако го избереме лицето a , мора да го избереме и лицето b . На колку начини може да се направи изборот.

Решение. Прв начин. Од сите избори на $k=6$ лица од $n=8$ елементи, C_8^6 , ги одземаме изборите што го содржат лицето a и не го содржат лицето b . Во тој случај елементот a е искористен т.е. класата е $k=5$, а од елементите отпаѓаат a и b т.е. $n=6$. Значи постојат C_6^5 начини. Следува:

$$S = C_8^6 - C_6^5 = \frac{8 \cdot 7}{2} - 6 = 28 - 6 = 22.$$

Втор начин. Избраните лица ќе ги поделиме на три дисјунктни групи: што ги содржат лицата a и b ; што го содржат лицето b и не го содржат лицето a ; и што не ги содржат ниту a , ниту b . Тогаш:

$$S = C_6^4 + C_6^5 + C_6^6 = \frac{6 \cdot 5}{2} + 6 + 1 = 22.$$

Трет начин. Кога имаме мал број комбинации, тие може да ги изброиме:

123456 123468 123578 124578 134568 145678 234678
 123457 123478 123678 124678 134578 234567 235678
 123458 ' 123567 ' 124567 ' 125678 ' 134678 ' 234568 ' 245678 '
 123467 123568 124568 134567 135678 234578 345678

Значи постојат вкупно 28 начини. Заради симетрија сеедно е кое е лицето a , а кое b . Ако земеме $a=1$, $b=2$; тогаш од изборите ќе ги отфрлиме тие што содржат 1, и не содржат 2. Такви се 6. Следува:

$$S = 28 - 6 = 22.$$

Дефиниција 1.3.10. Комбинација со повторување од n елементи, класа k , е секое подмножество од A во кое елементите се повторуваат, така што вкупниот број на елементи со повторувањето да е k .

Бројот на комбинации со повторување од n елементи, класа k , се означува со $\overline{C}_n^k$.

Броењето е како кај варијациите, само што елементите во комбинациите со повторување се во неопаѓачки редослед.

Својство 1.3.11. $\overline{C}_n^k = \binom{n+k-1}{k}.$

Доказ. Својството се докажува со индукција по $n+k-1$. Во $n+k-1$ -от чекор сите комбинации со повторување од n елементи, класа k ги делиме во две класи:

1. комбинации што го содржат елементот a_1 :

$a_1, \underbrace{\quad\quad\quad}_{k-1}$, каде што местата се пополнуваат со n елементи.

Нивниот број е: $\overline{C}_n^{k-1}$.

2. комбинации што не го содржат елементот a_1 :

$\underbrace{\quad\quad\quad}_k$, каде што местата се пополнуваат со $n-1$ елементи.

Нивниот број е: $\overline{C}_{n-1}^k$.

Класите се дисјунктни. Следува:

$$\bar{C}_n^k = \bar{C}_n^{k-1} + \bar{C}_{n-1}^k = \binom{n+k-2}{k-1} + \binom{n+k-2}{k} = \binom{n+k-1}{k},$$

што требаше да се докаже. Последното равенство кое лесно се докажува директно е својство на биномниот коефициент, докажано во лекцијата биномна формула во [1]. ■

Задача 1.3.12. Определи ги сите комбинации со повторување, класа 3 на множеството $A = \{1, 2, 3, 4\}$. Со помош на формула пресметај го нивниот број.

Решение. Комбинациите со повторување класа 3, 4 елементи се:

111, 112, 113, 114, 122, 123, 124, 133, 134, 144,
222, 223, 224, 233, 234, 244, 333, 334, 344, 444.

Нивниот број е $\bar{C}_4^3 = \binom{4+3-1}{3} = \binom{6}{3} = \frac{6 \cdot 5 \cdot 4 \cdot \cancel{3!}}{\cancel{3!} \cdot \cancel{3!}} = 5 \cdot 4 = 20$.

2. ПРОСТОР НА ВЕРОЈАТНОСТ

2.1. НАСТАНИ

Експеримент или **опит** е процес на испитување или набљудување при што се добива одреден резултат (исход). Ако секогаш се добива еден ист резултат, експериментот се нарекува **одреден** (детерминистички). Ако можните резултати на експериментот се повеќе од еден и не може да се определи точно кој од нив ќе се реализира, експериментот се нарекува **случаен** (стохастичен, недетерминистички).

Пример 2.1.1. Експеримент се состои во загревање вода на повеќе од 100° (при нормален притисок). Тогаш исходот-водата испарила, сигурно ќе се случи при секој експеримент бидејќи познато е дека водата врие на 100° . Следува дека експериментот е детерминистички. Од друга страна експериментот што се состои од набљудување дали тимот на Вардар ќе победи, изгуби или одигра нерешено одреден натпревар, не може со сигурност да се предвиди пред крајот на натпреварот. Затоа во овој случај експериментот е случаен.

Резултатите од случајниот експеримент се нарекуваат **елементарни настани** (исходи) и најчесто се означуваат со буквата ω , индексирана со индекси што припаѓаат на множество индекси I , т. е. со $\omega_i, i \in I$. Множеството од сите елементарни настани се нарекува **множество елементарни настани** и се означува со Ω . Значи: $\Omega = \{\omega_i | i \in I\}$.

Пример 2.1.2. Експеримент се состои од фрлање паричка. Се набљудува горната страна на паричката. Можни исходи се:

P - падна „петка“ или G - падна „глава“.

Следува дека множеството елементарни настани на овој експеримент е: $\Omega = \{P, G\}$.

Пример 2.1.3. Експеримент се состои од фрлање коцка за играње. Се набљудува бројот на горната страна на коцката. Можни исходи се да падне бројот 1, 2, ..., 6. Затоа:

$$\Omega = \{1, 2, \dots, 6\}.$$

Пример 2.1.4. Експеримент се состои од истовремено фрлање бела и црна коцка (за играње). На белата коцка може да се појави некој од броевите $1, 2, \dots, 6$, но и на црната коцка може да се појави некој од броевите $1, 2, \dots, 6$. Следува дека постојат 36 елементарни настани, односно:

$$\Omega = \{(i, j) | i, j \in \{1, 2, \dots, 6\}\}.$$

Случаен настан (настан) на даден експеримент е множество од елементарни настани на експериментот. Понатаму случајните настани ќе ги нарекуваме само настани и означуваме со големите латински букви

$$A, B, C, \dots.$$

Значи настаните се подмножества од Ω . При тоа не мора секое подмножество од Ω да биде случаен настан. Во понатамошниот дел од текстот ќе прецизираме кои подмножества може да бидат настани. Ќе велиме дека настанот A **се појавил (се реализирал)**, ако експериментот завршил со исход кој е елемент на настанот A . Во спротивно велиме дека настанот A **не се појавил (не се реализирал)**.

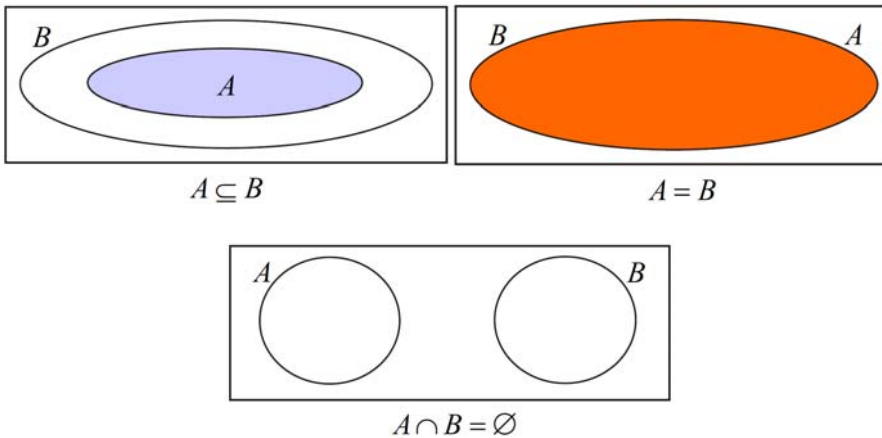
Множеството Ω се нарекува уште **сигурен настан**, бидејќи секогаш се реализира како резултат на кој било експеримент. Празното множество $\emptyset$ се нарекува **невозможен настан**, бидејќи никогаш не може да се појави како резултат на кој било експеримент.

Пример 2.1.5. Настанот A : „падна парен број при фрлање на коцка“ е: $A = \{2, 4, 6\}$.

Настаните се множества и со нив се оперира како што се оперира со множества.

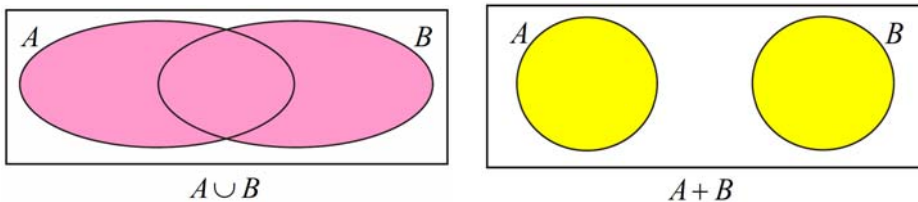
Релации со настани. Настанот A е **поднастан** од B (A е подмножество од B , A го имплицира B), ознака $A \subseteq B$, ако при секое појавување на настанот A се појавува и настанот B . Настаните A и B се **еднакви**, ознака $A = B$, ако настаните A и B се појавуваат истовремено. Настаните A и B се **дисјунктни**, ако

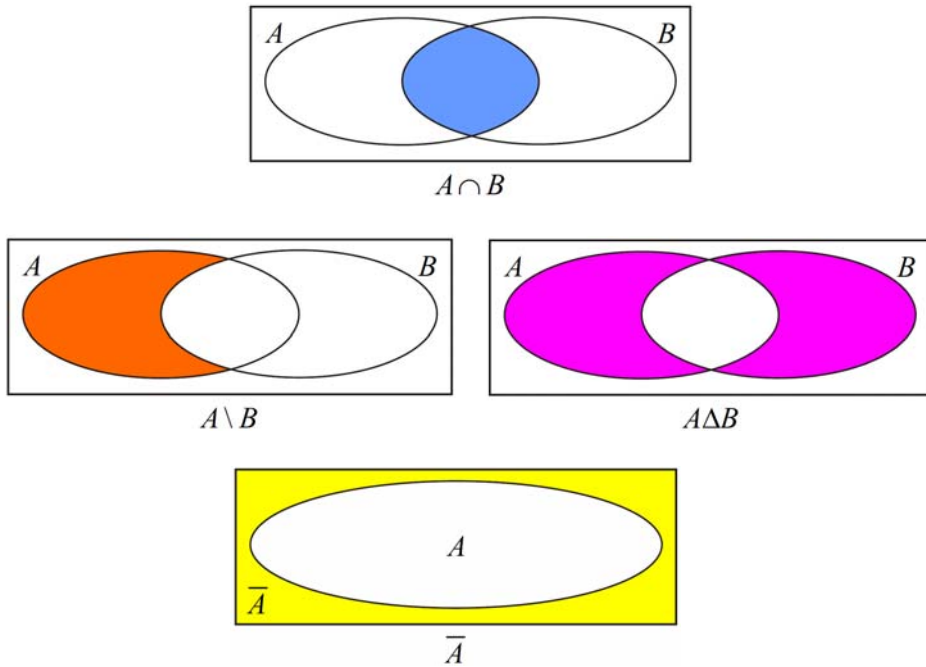
при појавување на настанот A не се појавува настанот B и обратно. Значи настаните A и B се дисјунктни ако: $A \cap B = \emptyset$.



Цртеж 2.1.1-3. Релации со настани

Операции со настани. Комплемент (спротивен настан) на настанот A , ознака $\bar{A}$, е настанот што се појавува кога нема да се појави настанот A . Значи, $\bar{A} = \Omega \setminus A$. **Унија** на настаните A и B , ознака $A \cup B$, е настанот што се појавува ако се појави барем еден од настаните A и B . Ако $A \cap B = \emptyset$, тогаш унијата на настаните A и B ја означуваме со: $A + B$ и ја нарекуваме **збир**. **Разлика** на настанот A со настанот B , ознака $A \setminus B$, е настанот што се појавува ако се појави настанот A и не се појави настанот B . **Симетрична разлика** на настаните A и B , ознака $A \Delta B$, е настанот што се реализира ако се реализира настанот $A \setminus B$ или настанот $B \setminus A$. Значи, $A \Delta B = (A \setminus B) \cup (B \setminus A)$. **Производ** на настаните A и B , ознака $A \cap B$ или AB , е настанот што се појавува кога се појавуваат истовремено настанот A и настанот B .





Цртеж 2.1.4-9. Операции со настани

Теорема 2.1.1. Нека Ω е множество елементарни настани и $A, B, C \subseteq \Omega$ се настани. Тогаш:

- 1) $A \cup \Omega = \Omega$, $A \cap \Omega = A$; 2) $A \cup A = A$, $A \cap A = A$;
- 3) $A \cup B = B \cup A$; 4) $A \cap B = B \cap A$;
- 5) $(A \cup B) \cup C = A \cup (B \cup C)$;
- 6) $(A \cap B) \cap C = A \cap (B \cap C)$;
- 7) $A \cap (B \cup C) = (A \cap B) \cup (A \cap C)$;
- 8) $A \cup (B \cap C) = (A \cup B) \cap (A \cup C)$;
- 9) $A \cap \bar{A} = \emptyset$, $A \cup \bar{A} = \Omega$; 10) $\bar{\bar{A}} = A$;
- 11) $\overline{A \cup B} = \bar{A} \cap \bar{B}$; 12) $\overline{A \cap B} = \bar{A} \cup \bar{B}$;
- 13) $A \subseteq B \Rightarrow \bar{B} \subseteq \bar{A}$; 14) $A \cap B \subseteq A \subseteq A \cup B$.

Доказ. Имаме:

1) $\omega \in A \cup \Omega \Leftrightarrow \omega \in A \vee \omega \in \Omega \Leftrightarrow \omega \in \Omega$. Оттука: $A \cup \Omega = \Omega$.
Слично, $A \cap \Omega = A$,

2) $\omega \in A \cup A \Leftrightarrow \omega \in A \vee \omega \in A \Leftrightarrow \omega \in A$. Следува: $A \cup A = A$.
Слично, $A \cap A = A$.

3) $\omega \in A \cup B \Leftrightarrow \omega \in A \vee \omega \in B \Leftrightarrow \omega \in B \vee \omega \in A \Leftrightarrow \omega \in B \cup A$.

Следува дека: $A \cup B = B \cup A$.

4) Се покажува аналогно како 3).

$$\begin{aligned} 5) \quad \omega \in (A \cup B) \cup C &\Leftrightarrow \omega \in A \cup B \vee \omega \in C \Leftrightarrow \\ \omega \in A \vee \omega \in B \vee \omega \in C &\Leftrightarrow \omega \in A \vee \omega \in B \cup C \Leftrightarrow \omega \in A \cup (B \cup C). \end{aligned}$$

Следува дека: $(A \cup B) \cup C = A \cup (B \cup C)$.

6) Се покажува аналогно како 5).

$$\begin{aligned} 7) \quad \omega \in A \cap (B \cup C) &\Leftrightarrow \omega \in A \wedge \omega \in B \cup C \Leftrightarrow \\ \omega \in A \wedge (\omega \in B \vee \omega \in C) &\Leftrightarrow (\omega \in A \wedge \omega \in B) \vee (\omega \in A \wedge \omega \in C) \Leftrightarrow \\ \omega \in A \cap B \vee \omega \in A \cap C &\Leftrightarrow \omega \in (A \cap B) \cup (A \cap C). \end{aligned}$$

Следува дека: $A \cap (B \cup C) = (A \cap B) \cup (A \cap C)$.

8) Се покажува аналогно како 7).

$$9) \quad \omega \in A \cap \bar{A} \Leftrightarrow \omega \in A \wedge \omega \in \bar{A} \Leftrightarrow \omega \in A \wedge \omega \notin A \Leftrightarrow \omega \in \emptyset.$$

$$10) \quad \omega \in \bar{\bar{A}} \Leftrightarrow \omega \notin \bar{A} \Leftrightarrow \omega \in A.$$

$$\begin{aligned} 11) \quad \omega \in \overline{A \cup B} &\Leftrightarrow \omega \notin A \cup B \Leftrightarrow \\ \omega \notin A \wedge \omega \notin B &\Leftrightarrow \omega \in \bar{A} \wedge \omega \in \bar{B} \Leftrightarrow \omega \in \bar{A} \cap \bar{B}. \end{aligned}$$

12) Се покажува аналогно како 11).

13) Нека од $\omega \in A$ следува $\omega \in B$. Сега нека $\omega \in \bar{B}$ т. е. $\omega \notin B$. Следува дека мора $\omega \in \bar{A}$, бидејќи ако $\omega \notin \bar{A}$ т. е. $\omega \in A$ би добиле дека $\omega \in B$, што противречи на претпоставката.

14) Ако $\omega \in A \cap B$ т. е. $\omega \in A$ и $\omega \in B$, тогаш $\omega \in A$. Од $\omega \in A$ следува дека $\omega \in A$ или $\omega \in B$. Значи: $A \cap B \subseteq A \subseteq A \cup B$. ■

Задача 2.1.6. Една машина за тестирање исправност на производи е направена така што ако производот е исправен да светне зелена сијалица, додека ако е неисправен да светне црвена сијалица. Најди го множеството од сите елементарни настани и настаните:

- A : точно еден производ е исправен;
- B : точно два производа се исправни;
- C : барем (најмалку) еден производ е исправен;
- D : најмногу два производа се неисправни;

ако се тестирани 3 производи.

Решение. Нека појавата на исправен производ т. е. светнувањето на зелена сијаличка ја означиме со $З$, додека појавата на неисправна сијаличка со $Ц$. Тогаш множеството елементарни настани е:

$$\Omega = \{333, 33Ц, 3ЦЗ, 3ЦЦ, ЦЗЗ, ЦЗЦ, ЦЦЗ, ЦЦЦ\}.$$

1) Настанот A значи дека првиот производ е исправен, а другите неисправни или вториот производ е исправен, другите неисправни или третиот производ е исправен, другите неисправни т. е.:

$$A = \{3ЦЦ, ЦЗЦ, ЦЦЗ\}$$

2) Настанот B значи дека (првиот и вториот) или (првиот и третиот) или (вториот и третиот) производ се исправни, а преостанатиот производ е неисправен, т.е.

$$B = \{33Ц, 3ЦЗ, ЦЗЗ\}.$$

Настанот B може да се најде и ако се определи еднаквиот на него настан: точно еден производ (првиот или вториот или третиот) е неисправен.

3) Настанот C е унија на настаните: точно еден производ (A), точно два производа (B) и точно три производа ($\{333\}$) се исправни т. е.:

$$C = \{3ЦЦ, ЦЗЦ, ЦЦЗ, 33Ц, 3ЦЗ, ЦЗЗ, 333\}.$$

Настанот C може да се определи и ако се определи неговиот спротивен настан $\bar{C}$: ниту еден производ не е исправен, $\bar{C} = \{ЦЦЦ\}$, од каде што: $C = \Omega \setminus \bar{C} = \Omega \setminus \{ЦЦЦ\}$.

4) Настанот D е еднаков со настанот C . Следува:

$$D = \{3ЦЦ, ЦЗЦ, ЦЦЗ, 33Ц, 3ЦЗ, ЦЗЗ, 333\}.$$

Настанот D можеше да се најде и ако се најдеше спротивниот настан $\bar{D}$: сите производи се неисправни, $\bar{D} = \{ЦЦЦ\}$, па: $D = \Omega \setminus \{ЦЦЦ\}$.

Задача 2.1.7. Нека A , B и C се три настани од дадено множество елементарни настани. Со помош на нив да се изразат настаните:

- 1) се појавил само настанот A ;
- 2) се појавил барем еден од дадените настани;
- 3) се појавил точно еден од дадените настани;
- 4) не се појавил ниту еден од дадените настани;
- 5) се појавиле најмногу два од дадените настани.

Решение. Бараните настани изразени преку дадените настани се:

- 1) $\overline{A}\overline{B}\overline{C}$; 2) $A \cup B \cup C$; 3) $\overline{A}\overline{B}\overline{C} \cup \overline{A}B\overline{C} \cup \overline{A}\overline{B}C$; 4) $\overline{A}\overline{B}\overline{C}$;
- 5) $\overline{A}\overline{B}\overline{C} \cup \overline{A}B\overline{C} \cup \overline{A}\overline{B}C \cup \overline{A}BC \cup A\overline{B}\overline{C} \cup A\overline{B}C \cup A\overline{B}C = \Omega \setminus \{ABC\}$.

Коментар. Во продолжение зборот „точно“ или „само“ некаде нема да се запишува, само ќе се подразбира. Така, „еден“ исправен производ ќе значи „точно еден“, не „барем еден“ исправен производ.

Задача 2.1.8. Во една цел се стрела 3 пати. Се разгледуваат настаните:

A_i : „целта е погодена во i -тото стрелање“, $i = 1, 2, 3$.

- а) Најди го множеството елементарни настани;
- б) Најди ги настаните:

B : постигнати се 3 погодоци;

C : целта е 3 пати промашена;

D : до третото стрелање немало погодок;

E : целта е промашена барем еднаш;

F : целта е погодена повеќе од 2 пати.

Решение. а) Множеството елементарни настани е:

$$\Omega = \{A_1A_2A_3, A_1A_2\overline{A}_3, A_1\overline{A}_2A_3, \overline{A}_1A_2A_3, A_1\overline{A}_2\overline{A}_3, \overline{A}_1A_2\overline{A}_3, \overline{A}_1\overline{A}_2A_3, \overline{A}_1\overline{A}_2\overline{A}_3\}.$$

б) Бараните настани се:

$$B = A_1A_2A_3, C = \overline{A}_1\overline{A}_2\overline{A}_3, D = \overline{A}_1\overline{A}_2A_3 \cup \overline{A}_1\overline{A}_2\overline{A}_3,$$

$$E = \Omega \setminus A_1A_2A_3 \text{ и } F = A_1A_2A_3.$$

Задача 2.1.9. Метална паричка се фрла два пати. Ако се паднат различни страни, тогаш паричката се фрла трет пат, а ако се падне иста, се фрла коцка. Определи го множеството елементарни настани и настанот A : не е паднат број. Колку елементи имаат множествата?

Решение. Множеството елементарни настани е:

$$\Omega = \{PGP, PGG, GPP, GPG, PP1, PP2, PP3, PP4, PP5, PP6, GG1, GG2, GG3, GG4, GG5, GG6\}, |\Omega| = 16;$$

додека:

$$A = \{PGP, PGG, GPP, GPG\}, |A| = 4.$$

2.2. СТАТИСТИЧКА ВЕРОЈАТНОСТ

Нека Ω е множество елементарни настани при некој експеримент и $A \subseteq \Omega$ е настан. Ако при серија од n експерименти настанот A се појавил $m = n(A)$ пати велиме дека серијата експерименти има **апсолутна честота** (честота) m . Бројот m/n се нарекува **релативна честота** на настанот A при дадената серија.

Кај статистичкиот експеримент, неопходен услов е да се повторува неограничен број пати. Тогаш:

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

Формулата следува од аксиома на веројатноста и законот на големите броеви, за кои ќе стане збор подоцна. Формулата вели дека кога бројот n е голем, количникот $n(A)/n$ е приближно ист, односно се групира околу некој број. При друга серија од n експерименти се добиваат други вредности за честотите m и m/n , но и тие тежат кон $P(A)$. Веројатноста се нарекува **статистичка веројатност**.

Во практиката се конструира математички модел така што на основа на податоците од серија од голем број n експерименти, на секој настан A му се доделува веројатност еднаква на неговата релативната честота, т. е.:

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

За друг математички модел со серија n експерименти се добива друга, но приближна по вредност веројатност на настанот. Статистичката веројатност на настанот A ќе биде бројот, околу кој се групираат релативните честоти на настанот A .

Статистичката веројатност уште се нарекува **емпириска** (апостериори) **веројатност**, бидејќи веројатноста за појавување на настанот A се утврдува после завршувањето на експериментите.

Пример 2.2.1. Нека при одреден број n експерименти, настанот A се појавува $n(A)$ пати, според следнава табела:

n	100	1000	10000	100000
$n(A)$	74	758	7528	74777

Статистичката веројатност на настанот A , односно бројот околу кој осцилира релативната честота на настанот A заокружен на две децимали е 0,75.

2.3. КЛАСИЧНА ВЕРОЈАТНОСТ

Нека $\Omega = \{\omega_1, \omega_2, \dots, \omega_n\}$ е конечно множество од n елементарни настани и сите подмножества од Ω се случајни настани. Нека A е произволен случаен настан со $n(A)$ елементи. Тогаш веројатноста за појава на настанот A е:

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

Ако елементите на Ω ги наречеме можни, додека на A поволни случаи (можности), тогаш:

$$P(A) = \frac{\text{број на поволни случаи}}{\text{број на можни случаи}}.$$

Класичната дефиниција на веројатност можеме да ја користиме кога множеството на елементарни настани е конечно и веројатноста за појава на секој од нив е иста.

На пример, класичната веројатност ја применивме при експериментот фрлање хомогена коцка или паричка, бидејќи веројатностите за појава на броевите 1, 2, ..., 6; кај коцката, или „петка“ и „глава“ кај паричката, се еднакви. Но, доколку коцката или монетата не се хомогени, тогаш нема иста веројатност за појава на секој од елементарните настани и класичната дефиниција на веројатност не може да се примени.

Кога разгледуваме математички модел во серија од n експерименти кај статистичката веројатност, веројатностите на моделот се класични веројатности.

Теорема 2.3.1. (Својства на класичната веројатност).

Нека Ω е конечно множество од n елементарни настани и веројатноста за појава на секој од нив е еднаква. Тогаш:

1) $0 \leq P(A) \leq 1$, за секој настан $A \subseteq \Omega$.

Притоа, $P(A) = 0$ ако и само ако $A = \emptyset$.

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

3) Ако A и B се дисјунктни настани, тогаш:

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

Доказ. 1) Нека $A \subseteq \Omega$ е настан. Тогаш:

$$0 \leq n(A) \leq n \Leftrightarrow 0 \leq \frac{n(A)}{n} \leq 1 \Leftrightarrow 0 \leq P(A) \leq 1.$$

$$2) P(\Omega) = \frac{n}{n} = 1.$$

3) Нека $n(A)$ и $n(B)$ се броевите на повољните можности на A и B . Бидејќи A и B се дисјунктни, бројот на повољни можности на настанот $A \cup B$ е $n(A \cup B) = n(A) + n(B)$. Тогаш:

$$P(A \cup B) = \frac{n(A \cup B)}{n} = \frac{n(A) + n(B)}{n} = \frac{n(A)}{n} + \frac{n(B)}{n} = P(A) + P(B). \blacksquare$$

Од 3) и принципот на математичка индукција следува дека ако $A_1, A_2, \dots, A_n \subseteq \Omega$ се попарно дисјунктни настани, тогаш:

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

Да повториме дека под исказот „се појавиле n настани или настанот се појавил n пати“, подразбираме дека „се појавиле точно n настани или настанот се појавил точно n пати“. Ако настанот се појавил најмалку n пати велиме: настанот се појавил барем n пати.

Задача 2.3.2. Во една кутија има 7 бели и 3 црни топчиња. Од кутијата се извлекуваат две топчиња, наеднаш. Колкава е веројатноста двете извлечени топчиња да се бели.

Решение. Нека настанот A е „извлечени се две бели топчиња“. Бројот на можни исходи на експериментот, т. е. бројот на

елементарни настани е $n = C_{10}^2 = \frac{10 \cdot 9}{2} = 45$. Бројот на поволни

случаи е: $n(A) = C_7^2 = \frac{7 \cdot 6}{2} = 21$. Следува дека веројатноста за појавување на настанот A изнесува:

$$P(A) = \frac{n(A)}{n} = \frac{21}{45} = \frac{7}{15} = 0,4(6).$$

Задача 2.3.3. На еден состанок на Студентскиот парламент на Градежниот факултет учествувале 50 студенти, од кои 30 биле од прв циклус. Од нив по пат на ждрепка се избираат 5 претставници за членови на Наставничкиот совет на Факултетот. Пресметај ја веројатноста:

- 1) меѓу избраните претставници да има 3 студенти од прв циклус;
- 2) меѓу избраните претставници да има барем 3 студенти од прв циклус;
- 3) меѓу избраните претставници да нема студенти од прв циклус.

Решение. Бројот на исходи е:

$$n = C_{50}^5 = \frac{50 \cdot 49 \cdot 48 \cdot 47 \cdot 46}{5 \cdot 4 \cdot 3 \cdot 2} = 10 \cdot 49 \cdot 2 \cdot 47 \cdot 46 = 2118760.$$

1) Нека настанот A е „меѓу избраните претставници има 3 студенти од прв циклус“. Од 30 членови од прв циклус 3 претставници може да се изберат на C_{30}^3 начини. За секој избор, од преостанатите 20 студенти, 2 претставника може да се изберат на C_{20}^2 начини. Затоа бројот на исходи на A е:

$$n(A) = C_{30}^3 C_{20}^2 = \frac{30 \cdot 29 \cdot 28}{3 \cdot 2} \cdot \frac{20 \cdot 19}{2} = 10 \cdot 29 \cdot 14 \cdot 10 \cdot 19 = 771400.$$

Следува дека веројатноста е:

$$P(A) = \frac{n(A)}{n} = \frac{771400}{2118760} \approx 0,364.$$

2) Нека настанот B е „меѓу избраните претставници има барем 3 студенти од прв циклус“. Значи дека меѓу избраните претставници може да има 3, 4 или 5 студенти од прв циклус, кои може да се изберат на $n(B) = C_{30}^3 C_{20}^2 + C_{30}^4 C_{20}^1 + C_{30}^5$ начини. Следува:

$$P(B) = \frac{n(B)}{n} = \frac{C_{30}^3 C_{20}^2 + C_{30}^4 C_{20}^1 + C_{30}^5}{C_{50}^5} =$$

$$\frac{\frac{30 \cdot 29 \cdot 28}{3 \cdot 2} \cdot \frac{20 \cdot 19}{2} + \frac{30 \cdot 29 \cdot 28 \cdot 27}{4 \cdot 3 \cdot 2} \cdot 20 + \frac{30 \cdot 29 \cdot 28 \cdot 27 \cdot 26}{5 \cdot 4 \cdot 3 \cdot 2}}{\frac{50 \cdot 49 \cdot 48 \cdot 47 \cdot 46}{5 \cdot 4 \cdot 3 \cdot 2}} =$$

$$\frac{10 \cdot 29 \cdot 7 \cdot 20 \cdot 19 + 5 \cdot 29 \cdot 7 \cdot 27 \cdot 20 + 29 \cdot 7 \cdot 27 \cdot 26}{10 \cdot 49 \cdot 2 \cdot 47 \cdot 46} =$$

$$\frac{771400 + 548100 + 142506}{2118760} = \frac{1462006}{2118760} \approx 0,69.$$

3) Нека настанот C е: „меѓу избраните претставници нема студенти од прв циклус“. Тогаш:

$$P(C) = \frac{n(C)}{n} = \frac{C_{20}^5}{C_{50}^5} = \frac{\frac{20 \cdot 19 \cdot 18 \cdot 17 \cdot 16}{5 \cdot 4 \cdot 3 \cdot 2}}{\frac{50 \cdot 49 \cdot 48 \cdot 47 \cdot 46}{5 \cdot 4 \cdot 3 \cdot 2}} =$$

$$\frac{19 \cdot 3 \cdot 17 \cdot 16}{5 \cdot 49 \cdot 4 \cdot 47 \cdot 46} = \frac{19 \cdot 3 \cdot 17 \cdot 2}{5 \cdot 49 \cdot 47 \cdot 23} = \frac{1938}{264845} \approx 0,0073.$$

Задача 2.3.4. Испитот за одреден предмет се состои од 7 прашања. На секое прашање се дадени по 4 понудени одговори, 1 точен и 3 неточни. Студентот го заокружува редниот број на еден од понудените одговори.

1) На колку различни начини може да се одговорат сите 7 прашања?

2) Колкава е веројатноста сите прашања точно да се одговорени?

3) Колкава е веројатноста точно да се одговорат 4 прашања?

4) Ако испитот се положува со најмалку 4 точни одговори, колкава е веројатноста испитот да се положи?

Решение. Нека со 1, 2, 3 и 4 ги означиме редните броеви на одговорите на секое од прашањата. Тогаш $A = \{1, 2, 3, 4\}$ е множеството елементи, додека бројот на прашања $k = 7$.

1) Бројот на различни начини на кои може да се одговорат 7-те прашања, односно бројот на елементарни настани е:

$$S = \overline{V}_4^7 = 4^7 = 16384.$$

2) Веројатноста на настанот A : „сите прашања се точно одговорени“; е:

$$P(A) = \frac{1}{\bar{V}_4^7} = \frac{1}{16384} = 0,000061.$$

3) Веројатноста на настанот B : „точно се одговорени 4 прашања“ е:

$$P(B) = \frac{C_7^4 3^3}{\bar{V}_4^7} = \frac{\frac{7 \cdot 6 \cdot 5}{3 \cdot 2} 3^3}{16384} = \frac{945}{16384} \approx 0,058.$$

4) Веројатноста на настанот C : „одговорени се барем 4 прашања“ е:

$$\begin{aligned} P(C) &= \frac{C_7^4 3^3 + C_7^5 3^2 + C_7^6 3 + 1}{\bar{V}_4^7} = \frac{\frac{7 \cdot 6 \cdot 5}{3 \cdot 2} 3^3 + \frac{7 \cdot 6}{2} 3^2 + 7 \cdot 3}{16384} = \\ &= \frac{945 + 189 + 21 + 1}{16384} = \frac{1156}{16384} \approx 0,071. \end{aligned}$$

Коментар. Задачата се решава и со биномна веројатност која ќе ја изучиме понатаму.

Задача 2.3.5. Група од $n \geq 3$ луѓе седат на кружна маса. Колкава е веројатноста двајца (претходно определени) пријатели да седат еден до друг.

Решение. Имаме, n пермутации: $a_1, a_2, \dots, a_n$; $a_2, a_3, \dots, a_n, a_1$; $\dots$, $a_n, a_1, a_2, \dots, a_{n-1}$; кои определуваат една циклична пермутација P_n^c од n елементи т. е.:

$$nP_n^c = P_n \Leftrightarrow nP_n^c = n! \Leftrightarrow P_n^c = \frac{n!}{n} = (n-1)!.$$

Нека настанот A е „пријателите седат еден до друг“. Пријателите можат да седат на два начина еден до друг (првиот да биде од лево а вториот од десно или обратно). Ако пријателите ги разгледаме како еден елемент тогаш заедно со другите $n-2$ членови формираат циклична пермутација од $n-1$ елемент. Вкупниот број на вакви пермутации е $(n-2)!$. Затоа:

$$P(A) = \frac{|A|}{|\Omega|} = \frac{2(n-2)!}{(n-1)!} = \frac{2(n-2)!}{(n-1)(n-2)!} = \frac{2}{n-1}.$$

2.4. ГЕОМЕТРИСКА ВЕРОЈАТНОСТ

Множеството елементарни настани Ω можеме да го определиме со помош на точки од фигури (на права, во рамнина, во простор) на кои им е дефинирана ненулта мера (вредност) $m = m(\Omega) \neq 0$ за густина на точките (како што се: должина на права, плоштина во рамнина, волумен во простор). Ако множеството поволни исходи на настанот A има мера $m(A)$, тогаш веројатноста за појава на настанот A е:

$$P(A) = \frac{m(A)}{m}.$$

Веројатноста се нарекува **геометриска веројатност**, бидејќи се темели на геометриската интерпретација на настаните. Ако ознаките l (P , V) се должината (плоштината, волуменот) на множеството елементарни настани, соодветно, тогаш веројатноста да се појави настанот A е:

$$P(A) = \frac{l(A)}{l} \quad (P(A) = \frac{P_{\text{пл.}}(A)}{P_{\text{пл.}}}, P(A) = \frac{V(A)}{V}),$$

каде што $l(A)$ ($P_{\text{пл.}}(A)$, $V(A)$) се должината (плоштината, волуменот) на настанот A .

Теорема 2.4.1. (Својства на геометриската веројатност). Нека Ω множество елементарни настани и веројатноста за појава на секој од нив е еднаква. Тогаш:

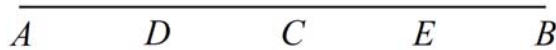
- 1) $0 \leq P(A) \leq 1$, за секој настан $A \subseteq \Omega$.
- 2) $P(\Omega) = 1$;
- 3) Ако A и B се сечат најмногу на множество со мера 0, тогаш: $P(A \cup B) = P(A) + P(B)$.

Доказ. Доказот следува директно од својствата на мерата.

Задача 2.4.2. Точка се фрла на отсечката AB , $\overline{AB} \neq 0$. Колкава е веројатноста точката да падне:

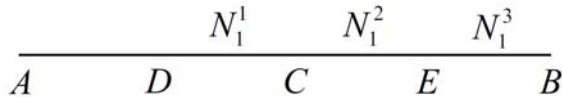
- 1) поблиску до средината отколку до крајната точка A ;
- 2) поблиску до средината отколку до една од крајните точки;
- 3) на една од крајните точки.

Решение. Нека C е средина на отсечката AB , додека D и E се средини на AC и CB соодветно.



Цртеж 2.4.2.0.

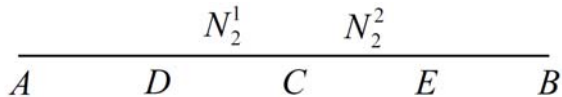
1) Нека N_1 е настанот „точката да падне поблиску до средината отколку до крајната точка A “. Тогаш:



Цртеж 2.4.2.1.

$$P(N_1) = \frac{\overline{DB}}{\overline{AB}} = \frac{\frac{3}{4}\overline{AB}}{\overline{AB}} = \frac{3}{4}.$$

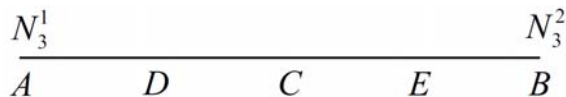
2) Нека N_2 е настанот „точката да падне поблиску до средината отколку до една од крајните точки“. Тогаш:



Цртеж 2.4.2.2.

$$P(N_2) = \frac{\overline{DE}}{\overline{AB}} = \frac{\frac{1}{2}\overline{AB}}{\overline{AB}} = \frac{1}{2}.$$

3) Нека N_3 е настанот „точката да падне на една од крајните точки“. Должинатана секоја точка X е $l(X) = 0$. Затоа веројатноста точката да падне на дадена точка X е: $P(X) = l(X) / \overline{AB} = 0$.



Цртеж 2.4.2.3.

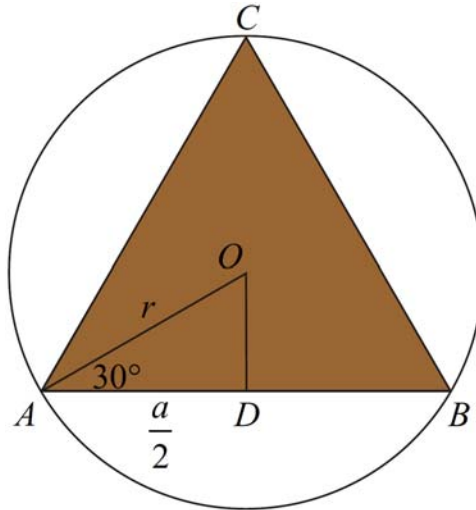
Следува:

$$P(N_3) = \frac{l(A) + l(B)}{AB} = \frac{0}{AB} = 0.$$

Задача 2.4.3. Во круг со радиус r е впишан рамностран триаголник. Случајно се избира точка од кругот. Колкава е веројатноста точката да лежи во триаголникот.

Решение. Страната на триаголникот е: $a = \sqrt{3}r$. Формулата може да се изведе на пример ако се разгледа правоаголниот триаголник ADO , каде што O е центарот на кружницата, а D е подножјето на виината спуштена од темето A . Аголот OAD е 30° бидејќи AO се симетрала на аголот CAB . Значи:

$$\overline{OD} = r \sin 30^\circ = \frac{1}{2}r.$$



Цртеж 2.4.4.

Од Питагоровата теорема важи:

$$\left(\frac{a}{2}\right)^2 + \left(\frac{r}{2}\right)^2 = r^2 \Leftrightarrow \frac{a^2}{4} + \frac{r^2}{4} = r^2 \Leftrightarrow \frac{a^2}{4} = \frac{3r^2}{4} \Leftrightarrow a = \sqrt{3}r.$$

Плоштината на правоаголниот триаголник е:

$$P_t = a^2 \frac{\sqrt{3}}{4} = (\sqrt{3}r)^2 \frac{\sqrt{3}}{4} = \frac{3\sqrt{3}}{4}r^2.$$

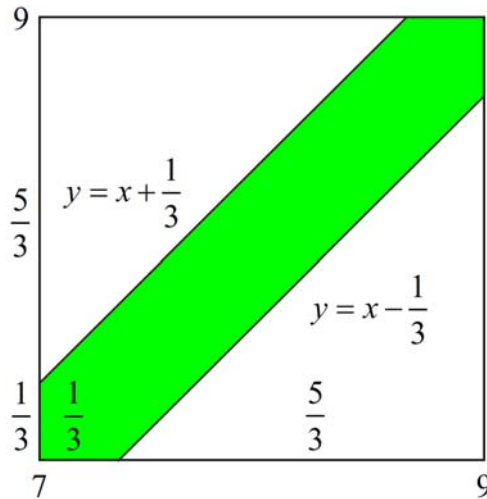
Конечно, веројатноста точката да биде во триаголникот е:

$$P = \frac{P_t}{P_k} = \frac{\frac{3\sqrt{3}}{4}r^2}{r^2\pi} = \frac{3\sqrt{3}}{4\pi} \approx 0,41.$$

Задача 2.4.4. Двајца колеги се договориле да се чекаат пред Рекорд меѓу 7 и 9 часот. Секој од нив доаѓа на Рекорд чека 20 минути и ако во тоа време не се сретне со колегата, си оди. Колкава е веројатноста колегите да се сретнат на Рекорд.

Решение. Нека со A го означиме настанот: лицата се сретнале. Нека со x го означиме времето на доаѓање на едниот колега, а со y на другиот. Важи:

$$7 \leq x \leq 9 \text{ и } 7 \leq y \leq 9.$$



Цртеж 2.4.5.

Множеството елементарни настани Ω е квадрат со страна 2 т.е. $m(\Omega) = 4$. Ако колегите се сретнале, тогаш разликата од нивните времиња на доаѓање е непоголема од $\frac{1}{3}$ саати т.е.

$$|x - y| \leq \frac{1}{3} \Leftrightarrow -\frac{1}{3} \leq x - y \leq \frac{1}{3} \Leftrightarrow y \geq x - \frac{1}{3}, y \leq x + \frac{1}{3}.$$

Мерата на настанот A е плоштината на ликот ограничен со правите:

$$x=7, x=9, y=7, y=9, y=x-\frac{1}{3} \text{ и } y=x+\frac{1}{3};$$

кој што се добива како разлика на плоштината на квадратот со плоштината на двата складни рамнокраки триаголници со страна $\frac{5}{3}$.

Значи:

$$m(A) = 2^2 - 2 \cdot \frac{1}{2} \left(\frac{5}{3} \right)^2 = 4 - \frac{25}{9} = \frac{11}{9}.$$

Конечно:

$$P(A) = \frac{m(A)}{m} = \frac{\frac{11}{9}}{4} = \frac{11}{36} = 0,30(5).$$

Задача 2.4.5. Прачка со должина a е скршена на две места, така што се добиваат три нови прачки. Колкава е веројатноста со прачките да се конструира триаголник.

Решение. Прв начин. Нека должините на првите две прачки се x и y . Тогаш должината на третата прачка е $a - x - y$.



Цртеж 2.4.3.1.

а) Првата прачка има должина поголема од нула, а помала од a т.е. $0 < x < a$. Тогаш бидејќи должината на остатокот од прачката е $a - x$, втората прачка има должина: $0 < y < a - x$. Затоа множеството елементарни настани Ω е внатрешноста на ликот ограничен со правите:

$$x=0, x=a, y=0 \text{ и } y=a-x.$$

Неговата мера е плоштината на ликот, кој што е рамнокрак правоаголен триаголник со крак a , е:

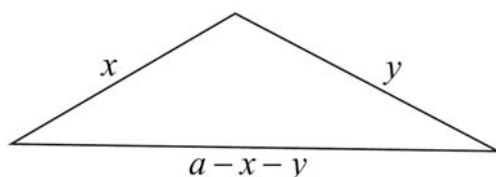
$$m(\Omega) = \frac{a^2}{2}.$$

б) За да новодобиените прачки формираат триаголник треба збирот на две страни да е поголем од третата, односно да важи:

$$x + y > a - x - y \Leftrightarrow 2(x + y) > a \Leftrightarrow x + y > \frac{a}{2} \quad (1);$$

$$x + a - x - y > y \Leftrightarrow a > 2y \Leftrightarrow y < \frac{a}{2} \quad (2); \text{ и}$$

$$y + a - x - y > x \Leftrightarrow a > 2x \Leftrightarrow x < \frac{a}{2} \quad (3).$$

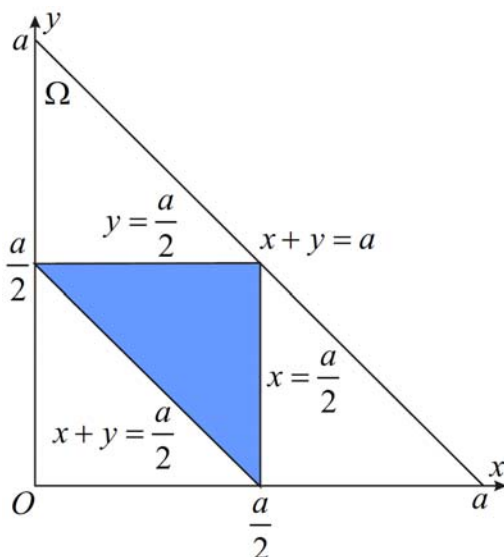


Цртеж 2.4.3.2.

Да забележиме дека третата неравенка можевме да ја запишеме директно од втората заради симетрија на x и y .

Ако со A го означиме настанот: со прачките е конструиран триаголник, тогаш мерата на A е плоштината на триаголникот ограничен со правите: $x = \frac{a}{2}$, $y = \frac{a}{2}$ и $x + y = \frac{a}{2}$ т.е.

$$m(A) = \frac{1}{2} \left(\frac{a}{2} \right)^2 = \frac{a^2}{8}.$$



Цртеж 2.4.3.3.

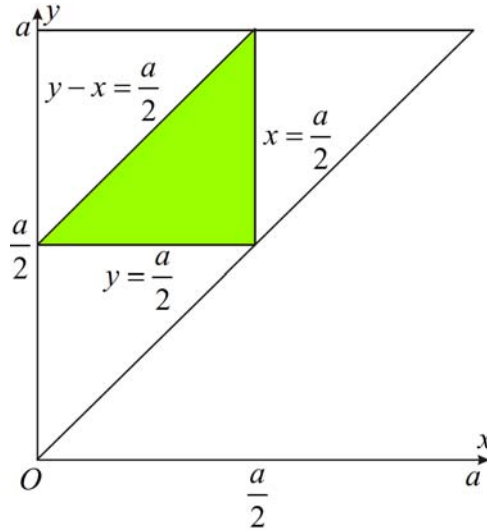
Следува:

$$P(A) = \frac{m(A)}{m} = \frac{\frac{a^2}{8}}{\frac{a^2}{2}} = \frac{1}{4}.$$

Втор начин. Точките x и y не ги разгледуваме како должини, туку како координати, точки од множеството $[0, a]$. Притоа претпоставуваме дека $x < y$. Тогаш Ω е определена со:

$$x \in (0, a), y \in (0, a) \text{ и } x < y;$$

и претставува внатрешноста на горниот триаголник на цртеж 2.4.3.5.



Цртеж 2.4.3.4.

Во тој случај должините на деловите од прачката се: x , $y-x$ и $a-y$. За да од нив може да се формира триаголник треба да важи:

$$x + y - x > a - y \Leftrightarrow 2y > a \Leftrightarrow y > \frac{a}{2},$$

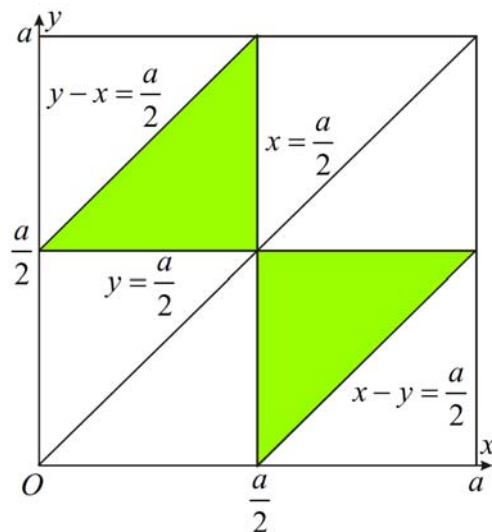
$$x + a - y > y - x \Leftrightarrow a > 2(y - x) \Leftrightarrow y - x > \frac{a}{2},$$

$$y - x + a - y > x \Leftrightarrow a > 2x \Leftrightarrow x < \frac{a}{2}.$$

Значи областа поволни исходи е обоениот триаголник (види цртеж 2.4.3.4). Тогаш:

$$P(A) = \frac{m(A)}{m} = \frac{\frac{a^2}{8}}{\frac{a^2}{2}} = \frac{1}{4}.$$

Трет начин. Без претпоставката $x < y$, Ω ќе биде внатрешноста на квадратот т.е. $x \in (0, a)$, $y \in (0, a)$, додека областа поволни настани ќе биде дополнета со уште еден триаголник кој што се добива од условите, со замена на x и y .



Цртеж 2.4.3.5.

Четврт начин. Проблемот може да се сведе на линеарен случај т.е. на должина на отсечка. Најголемата нова прачка x е непомала од $\frac{a}{3}$, бидејќи во спротивно ќе постои поголема прачка од неа. Исто

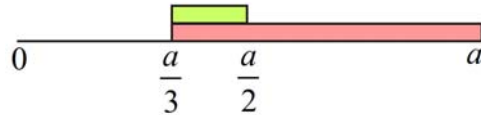
така $x < a$. Значи $x \in \left[\frac{a}{3}, a\right)$. Должината на соодветната отсечка е

$l(x) = a - \frac{a}{3} = \frac{2a}{3}$. За да се формира триаголник најголемата страна

треба да биде помала од $\frac{a}{2}$, во спротивно би била поголема од

збирот на останатите страни. Затоа множеството поволни исходи е интервалот $x \in \left[\frac{a}{3}, \frac{a}{2}\right)$ кој има мера:

$$l(x) = \frac{a}{2} - \frac{a}{3} = \frac{a}{6}.$$



Цртеж 2.4.3.5.

Оттука веројатноста е:

$$P(A) = \frac{m(A)}{m} = \frac{\frac{a}{6}}{\frac{2a}{3}} = \frac{1}{4}.$$

Петти начин. Да разгледаме рамностран триаголник со висина a . Според теоремата на Вивиани збирот од растојанијата на која било внатрешна точка до страните на триаголникот е еднаква на висината a . Теоремата лесно се покажува ако се споредат плоштините на триаголникот и на трите триаголници со теме во внатрешната точка и по две темиња од секоја страна. Значи Ω е внатрешноста на триаголникот. За да исходот е поволен треба секое растојание да е помало од $\frac{a}{2}$, односно множеството поволни настани е внатрешноста на триаголникот формиран од средините. Бидејќи средините го делат триаголникот на 4 складни триаголници важи:

$$P(A) = \frac{m(A)}{m} = \frac{1}{4}.$$

Шести начин. Дел од претходните начини може да се разгледаат во простор. На пример во првиот начин ако се разгледа и третата должина, важи:

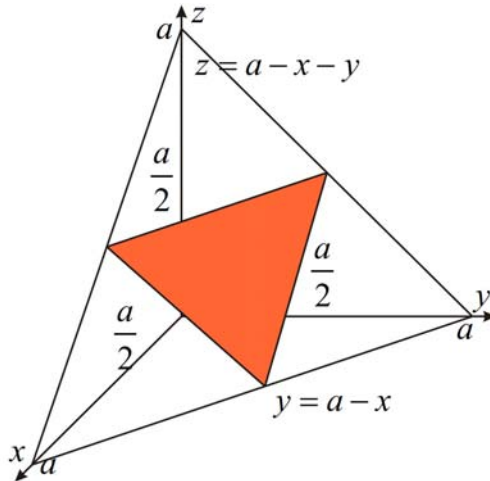
$$0 < x < a, \quad 0 < y < a - x, \quad z = a - x - y;$$

што ја претставува горната страна од тетраедарот на цртежот. Поволните исходи го исполнуваат условот:

$$x, y, z < \frac{a}{2},$$

што значи дека множеството поволни резултати е обоениот триаголникот од цртежот 2.4.3.6, кој што е еден од четирите складни триаголници што го сочинуваат Ω . Затоа:

$$P(A) = \frac{1}{4}.$$



Цртеж 2.4.3.6.

Седми начин. Рамнинските случаи може да се анализираат во простор, аналогно на шестиот начин.

Задачата покажа дека пристапот на определување на геометриската веројатност може да биде различен.

2.5. АКСИОМАТСКА ДЕФИНИЦИЈА НА ВЕРОЈАТНОСТА

Нека F е фамилија од подмножества од множеството на елементарни настани Ω која се состои од настаните при даден експеримент.

Дефиниција 2.5.1. Фамилијата F се нарекува σ -алгебра од подмножества на Ω , ако:

- 1) $\Omega \in F$;
- 2) Ако $A \in F$, тогаш $\bar{A} \in F$;
- 3) Ако $A_1, A_2, \dots \in F$, тогаш $\bigcup_{i=1}^{\infty} A_i \in F$.

Елементите на σ -алгебрата се нарекуваат **случајни настани (настани)**.

Од 1) и 2) следува дека Ω и $\emptyset = \overline{\Omega}$ се настани.

Ако $A_1, A_2, \dots$ се настани, тогаш нивниот преброив пресек $\bigcap_{i=1}^{\infty} A_i$ е настан. Имено од Деморгановите закони $\bigcap_{i=1}^{\infty} A_i = \overline{\bigcup_{i=1}^{\infty} \bar{A}_i}$, при што бидејќи $A_i, i \in \mathbb{N}$; се настани, тогаш од 2) и $\bar{A}_i, i \in \mathbb{N}$; се настани и од 3) $\bigcup_{i=1}^{\infty} \bar{A}_i$ е настан, од каде што од 2) $\overline{\bigcup_{i=1}^{\infty} \bar{A}_i}$ е настан.

Ако $A_1, A_2, \dots, A_n$ се настани, тогаш $\bigcup_{i=1}^n A_i$ е настан. Имено ако избереме $A_{n+1} = A_{n+2} = \dots = \emptyset$, од 3) следува дека: $\bigcup_{i=1}^{\infty} A_i = \bigcup_{i=1}^n A_i$ е настан. Аналогно се покажува дека $\bigcap_{i=1}^n A_i$ настан.

Ако A и B се настани, тогаш $A \cup B$, $A \cap B$ и $A \setminus B = A \cap \bar{B}$ се настани. Значи дека секоја σ -алгебра е затворена во однос на операциите комплемент, преброива и конечна унија и пресек. Затоа понатаму ќе работиме само со сигма алгебри кои овозможуваат затвореност на овие операции со настани.

Без доказ ќе наведеме дека за секоја фамилија A од подмножества од Ω , постои најмала σ -алгебра F што ја содржи фамилијата A . Фамилијата F се нарекува σ -алгебра генерирана од A .

Ако за настаните $A_i, i \in F$; важи $A_i \cap A_j = \emptyset$ за секои $i, j \in F, i \neq j$; тогаш настаните ги нарекуваме **попарно дисјунктни**, а нивната унија **дисјунктна сума** и место симболите „ $\cup$ “ и „ $\bigcup$ “ ги користиме симболите „ $+$ “ и „ Σ “, соодветно.

Дефиниција 2.5.2. (Аксиоматска дефиниција на веројатноста). Нека F е σ -алгебра од подмножества на Ω . Пресликувањето $P: F \rightarrow [0,1]$ се нарекува **веројатност** над F ако:

- 1) $P(A) \geq 0$ за секој $A \in F$;
- 2) $P(\Omega) = 1$;

3) Ако $A_i \in F$, $i = 1, 2, \dots$; и $A_i \cap A_j = \emptyset$ за $i \neq j$, тогаш:

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

Важи: Ако $A_1, A_2, \dots, A_n$ се попарно дисјунктни настани, тогаш:

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

Имено ако $A_{n+1} = A_{n+2} = \dots = \emptyset$ од 3) следува точноста на тврдењето.

Дефиниција 2.5.3. (Простор на веројатност).

Подредената тројка (Ω, F, P) каде што Ω е произволно непразно множество, F е σ -алгебра од подмножества на Ω и P е веројатност над F , се нарекува **простор на веројатност**.

Просторот на веројатност (Ω, F, P) во кој множеството елементарни настани Ω е дискретно се нарекува **дискретен простор на веројатност**. Ако Ω е конечно, говориме за **конечен простор на веројатност**. Во спротивно (Ω, F, P) се нарекува **апсолутно непрекинат простор на веројатност**.

Класичната веројатност е специјален случај на конечен простор на веројатност (Ω, F, P) . Кај конечниот простор на веројатност не мора веројатностите на елементарните настани да се еднакви.

Теорема 2.5.4. (Својства на веројатноста). Нека

(Ω, F, P) е простор на веројатност. Тогаш:

- 1) $P(\emptyset) = 0$;
- 2) $P(A) \leq P(B)$ кога $A \subseteq B$;
- 3) $0 \leq P(A) \leq 1$;
- 4) $P(\bar{A}) = 1 - P(A)$.

Доказ. 1) Настаните $\emptyset$ и Ω се попарно дисјунктни и $\Omega = \Omega + \emptyset$. Според аксиома 3),

$$P(\Omega) = P(\Omega) + P(\emptyset) \Leftrightarrow P(\emptyset) = 0.$$

2) Настаните A и $\bar{A}B$ се дисјунктни и кога $A \subseteq B$, важи: $A + \bar{A}B = B$. Следува:

$$P(B) = P(A) + P(\bar{A}B) \geq P(A).$$

3) За секој A настан Од $\emptyset \subseteq A \subseteq \Omega$ следува дека:

$$P(\emptyset) \leq P(A) \leq P(\Omega) \Leftrightarrow 0 \leq P(A) \leq 1.$$

4) Настаните A и $\bar{A}$ се дисјунктни и $A + \bar{A} = \Omega$. Следува:

$$P(A) + P(\bar{A}) = P(\Omega) \Leftrightarrow P(A) + P(\bar{A}) = 1 \Leftrightarrow P(\bar{A}) = 1 - P(A). \blacksquare$$

Теорема 2.5.5. (Веројатност од унија од два настана).

Ако A и B се настани, тогаш:

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

Доказ. Настаните A и $\bar{A}B$ се дисјунктни и $A \cup B = A + \bar{A}B$.
Следува дека: $P(A \cup B) = P(A) + P(\bar{A}B)$.

Настаните AB и $\bar{A}B$ се дисјунктни и $B = AB \cup \bar{A}B$. Следува дека: $P(B) = P(AB) + P(\bar{A}B)$ од каде што: $P(\bar{A}B) = P(B) - P(AB)$.

Значи:

$$P(A \cup B) = P(A) + P(\bar{A}B) = P(A) + P(B) - P(AB). \blacksquare$$

*Од претходната теорема користејќи го принципот на математичка индукција се покажува точноста на следната теорема:

Теорема 2.5.6. (Веројатност од конечна унија настани). Ако $A_1, A_2, \dots, A_n$ се настани, тогаш:

$$P\left(\bigcup_{i=1}^n A_i\right) = \sum_{i=1}^n P(A_i) - \sum_{1 \leq i < j \leq n} P(A_i A_j) + \sum_{1 \leq i < j < k \leq n} P(A_i A_j A_k) - \dots + (-1)^{n-1} P(A_1 A_2 \dots A_n). \blacksquare$$

Специјално, за $n = 3$, ако A , B и C се настани, важи:

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

$$P(A) + P(B) + P(C) - P(AB) - P(AC) - P(BC) + P(ABC).$$

Теорема 2.5.7. (Лема за покривање). Нека F е σ -алгебра. За произволни настани $A_i \in F$, $i \in \mathbb{N}$; важи:

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

Доказ. Настанот $\bigcup_{i=1}^{\infty} A_i$ е дисјунктен збир на настаните:
 $A_1, A_2 \bar{A}_1, A_3 \bar{A}_2 \bar{A}_1, \dots$

Следува:

$$P\left(\bigcup_{i=1}^{\infty} A_i\right) = P(A_1) + P(A_2 \bar{A}_1) + P(A_3 \bar{A}_2 \bar{A}_1) + \dots$$

Бидејќи $A_n \bar{A}_{n-1} \bar{A}_{n-2} \dots \bar{A}_1 \subseteq A_n$ за секое $n \in \mathbb{N}$, важи:

$$P(A_n \bar{A}_{n-1} \bar{A}_{n-2} \dots \bar{A}_1) \leq P(A_n), \quad n \in \mathbb{N}.$$

Затоа:

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

Теорема 2.5.8. (Непрекинатост на веројатноста). Нека F е σ -алгебра од подмножества на Ω . За $A_i \in F$, $i \in \mathbb{N}$; важи:

1) Ако $A_1 \subseteq A_2 \subseteq \dots \subseteq A_n \subseteq \dots$, тогаш: $P\left(\bigcup_{i=1}^{\infty} A_i\right) = \lim_{n \rightarrow +\infty} P(A_n)$;

2) Ако $A_1 \supseteq A_2 \supseteq \dots \supseteq A_n \supseteq \dots$, тогаш: $P\left(\bigcap_{i=1}^{\infty} A_i\right) = \lim_{n \rightarrow +\infty} P(A_n)$.

Доказ. 1) Бидејќи $A_i \subseteq A_{i+1}$, за секое $i \in \mathbb{N}$; настаните:

$$A_i = \bar{A}_0 A_i \text{ каде што } \bar{A}_0 = \Omega, \bar{A}_1 A_2, \bar{A}_2 A_3, \dots$$

се попарно дисјунктни,

$$\bigcup_{i=1}^n A_i = A_n \text{ и } \bigcup_{i=1}^n A_i = \bigcup_{i=1}^n \bar{A}_{i-1} A_i \text{ за секое } n \in \mathbb{N}; \text{ и } \bigcup_{i=1}^{\infty} A_i = \bigcup_{i=1}^{\infty} \bar{A}_{i-1} A_i.$$

Следува дека:

$$P\left(\bigcup_{i=1}^{\infty} A_i\right) = P\left(\bigcup_{i=1}^{\infty} \bar{A}_{i-1} A_i\right) = \sum_{i=1}^{\infty} P(\bar{A}_{i-1} A_i) =$$

$$\lim_{n \rightarrow +\infty} \sum_{i=1}^n P(\bar{A}_{i-1} A_i) = \lim_{n \rightarrow +\infty} P\left(\bigcup_{i=1}^n \bar{A}_{i-1} A_i\right) = \lim_{n \rightarrow +\infty} P\left(\bigcup_{i=1}^n A_i\right) = \lim_{n \rightarrow +\infty} P(A_n).$$

2) Нека $A_1 \supseteq A_2 \supseteq \dots \supseteq A_n \supseteq \dots$. Тогаш $\bar{A}_1 \subseteq \bar{A}_2 \subseteq \dots \subseteq \bar{A}_n \subseteq \dots$ и заради $\bigcup_{i=1}^{\infty} \bar{A}_i = \overline{\bigcap_{i=1}^{\infty} A_i}$ и заради 1) добиваме:

$$P\left(\bigcup_{i=1}^{\infty} \bar{A}_i\right) = \lim_{n \rightarrow +\infty} P(\bar{A}_n) \Leftrightarrow P\left(\overline{\bigcap_{i=1}^{\infty} A_i}\right) = \lim_{n \rightarrow +\infty} (1 - P(A_n)) \Leftrightarrow$$

$$1 - P\left(\bigcap_{i=1}^{\infty} A_i\right) = 1 - \lim_{n \rightarrow +\infty} P(A_n) \Leftrightarrow P\left(\bigcap_{i=1}^{\infty} A_i\right) = \lim_{n \rightarrow +\infty} P(A_n).$$

Се покажува дека **класичната, статистичката и геометриската веројатност** се веројатности.

1. Имено кај класичната дефиниција $P(A) = \frac{n(A)}{n} \geq 0$ за секој настан A подмножество до Ω , каде што Ω е множеството елементарни настани. Следува дека првото својство е исполнето. Понатаму:

$$P(\Omega) = \frac{n(\Omega)}{n} = \frac{n}{n} = 1.$$

Значи второто својство е исполнето. Ќе покажеме дека важи и третото својство. Нека $A_1, A_2, \dots, A_n$ се конечно многу попарно дисконктни настани со поволни исходи:

$$n(A_1), n(A_2), \dots, n(A_n).$$

Тогаш:

$$P\left(\bigcup_{i=1}^n A_i\right) = \frac{n(A_1) + n(A_2) + \dots + n(A_n)}{n} = \frac{n(A_1)}{n} + \frac{n(A_2)}{n} + \dots + \frac{n(A_n)}{n} =$$

$$P(A_1) + P(A_2) + \dots + P(A_n) = \sum_{i=1}^n P(A_i).$$

2. Простор на веројатност

2. Нека P е статистичка веројатност дефинирана на партитивното множество на Ω . Тогаш:

$$1) P(A) = \lim_{n \rightarrow +\infty} \frac{n(A)}{n} \geq 0, \text{ бидејќи } \frac{n(A)}{n} \geq 0 \text{ за секој } n \in \mathbb{N};$$

$$2) P(\Omega) = \lim_{n \rightarrow +\infty} \frac{n(\Omega)}{n} = \lim_{n \rightarrow +\infty} \frac{n}{n} = \lim_{n \rightarrow +\infty} 1 = 1;$$

и

3) За дисјунктните настани $A_1, A_2, \dots, A_n$ такви што:

$$P(A_i) = \lim_{n \rightarrow +\infty} \frac{n(A_i)}{n}, \quad i = 1, 2, \dots, n;$$

важи:

$$\begin{aligned} P\left(\bigcup_{i=1}^n A_i\right) &= \lim_{n \rightarrow +\infty} \frac{n(A_1) + n(A_2) + \dots + n(A_n)}{n} = \\ &= \lim_{n \rightarrow +\infty} \frac{n(A_1)}{n} + \lim_{n \rightarrow +\infty} \frac{n(A_2)}{n} + \dots + \lim_{n \rightarrow +\infty} \frac{n(A_n)}{n} = \\ &= P(A_1) + P(A_2) + \dots + P(A_n) = \sum_{i=1}^n P(A_i). \end{aligned}$$

Веројатноста на настаните во математичкиот модел од голем број n експерименти е класична веројатност и според тоа функцијата на веројатност P е веројатност.

3. Нека P е геометриска веројатност, нека Ω е множество елементарни настани и нека е дадена σ -алгебра од подмножества на Ω на кои е дефинирана мера таква што $m(\Omega) = m$, тогаш:

$$1) P(A) = \frac{m(A)}{m} \geq 0 \text{ бидејќи мерата (за овој материјал тоа е}$$

должината, плоштината и волуменот) е ненегативна;

$$2) P(\Omega) = \frac{m(\Omega)}{m} = \frac{m}{m} = 1;$$

и

3) Ако $A_1, A_2, \dots$ се дисјунктни настани такви што:

$$P(A_i) = \frac{m(A_i)}{m}, \quad i \in \mathbb{N};$$

тогаш:

$$P\left(\bigcup_{i=1}^{\infty} A_i\right) = \frac{m(A_1) + m(A_2) + \dots}{m} = \frac{m(A_1)}{m} + \frac{m(A_2)}{m} + \dots =$$
$$P(A_1) + P(A_2) + \dots = \sum_{i=1}^{\infty} P(A_i).$$

3. УСЛОВНА ВЕРОЈАТНОСТ. НЕЗАВИСНИ НАСТАНИ. ТОТАЛНА ВЕРОЈАТНОСТ. БЕЈЕСОВИ ФОРМУЛИ.

3.1. УСЛОВНА ВЕРОЈАТНОСТ

Често се случува веројатноста за појавување на еден настан да зависи од појавата на друг настан.

Пример 3.1.1. Нека се дадени настаните

A : врне снег во Скопје;

B : денот е 21 декември.

Веројатноста за појава на настанот A зависи од појавата на настанот B . Имено, информацијата дека се појавил настанот B , ја зголемува веројатноста за појава на A .

Примерот покажува дека се јавува потреба за дефинирање на веројатноста за појава на еден настан A , при услов да се појавил друг настан B . Оваа веројатност ќе ја наречеме **условна веројатност** и ќе ја означиме со $P(A|B)$.

Условна веројатност при класичната дефиниција на веројатност. Нека се изведени n експерименти во кои настанот A се појавил $n(A)$ пати; настанот B , $n(B)$ пати; додека настанот AB , $n(AB)$ пати. Тогаш релативната честота за појава на настанот A при услов да се појавил настанот B е:

$$\frac{n(AB)}{n(B)} = \frac{\frac{n(AB)}{n}}{\frac{n(B)}{n}} \approx \frac{P(AB)}{P(B)}. \text{ Следува: } P(A|B) = \frac{P(AB)}{P(B)}.$$

Задача 3.1.2. Експеримент се состои во фрлање коцка. Определи ја веројатноста да падне 6-ка, ако е познато дека паднал парен број.

Решение. Да ги означиме настаните со:

A : паднала 6-ка, B : паднал парен број.

Тогаш бараната веројатност е:

$$P(A|B) = \frac{P(AB)}{P(B)} = \frac{\frac{1}{6}}{\frac{3}{6}} = \frac{1}{3}. \blacksquare$$

Задача 3.1.3. Во една продавница за храна на полицата има 2 бели и 2 црни леба од производителот X и 1 бел и 3 црни леба од производителот Y . Тројца купувачи купуваат леб. На првиот купувач не му е битно од кој производител ќе купи леб, вториот купувач ќе купи леб од производителот X , додека третиот од производителот Y . Колкави се веројатностите лебот што е купен од секој од купувачите да биде бел.

Решение. Веројатностите првиот, вториот и третиот купувач да купат бел леб, соодветно се:

$$P_1 = \frac{3}{8} = 0,375, \quad P_2 = \frac{2}{4} = \frac{1}{2} = 0,5 \quad \text{и} \quad P_3 = \frac{1}{4} = 0,25.$$

Условна веројатност во произволен простор на веројатност. Ќе дефинираме условна веројатност во произволен простор на веројатност која ќе се совпаѓа со условната веројатност при класичната дефиниција на веројатност.

Дефиниција 3.1.4. Нека (Ω, F, P) е простор на веројатност и нека се $A, B \in F$, така што $P(B) > 0$. **Условна веројатност за појава на настанот A при услов да се случил настанот B ,** ознака $P(A|B)$, е бројот $P(A|B) = \frac{P(AB)}{P(B)}$.

Нека $F_B, B \in F$ е множество од сите множества од видот $AB, A \in F$. Тогаш F_B е σ -алгебра во однос на множеството B . Нека Ω_B е множеството од елементарни настани што го сочинуваат настанот B и $P_B = P(A|B)$. Комплементот на AB во Ω_B е елементот $\bar{A}B$.

Теорема 3.1.5. Нека B е даден настан со веројатност $P(B) > 0$. Пресликувањето:

$$P_B : F_B \rightarrow [0, 1], \quad P_B = P(A|B)$$

е веројатност над F .

Доказ. Нека A е настан. Бидејќи $P(AB) \geq 0$ и $P(B) > 0$, следува дека:

$$P_B(A) = P(A|B) = \frac{P(AB)}{P(B)} \geq 0$$

Значи дека првата аксиома е задоволена. Заради:

$$P_B(\Omega) = P(\Omega|B) = \frac{P(\Omega B)}{P(B)} = \frac{P(B)}{P(B)} = 1,$$

втората аксиома е исполнета. Ако $A_i, i = 1, 2, \dots$; се настани такви што и $A_i \cap A_j = \emptyset$ за $i \neq j$, тогаш $A_i B \cap A_j B = \emptyset$ за $i \neq j$ и имајќи предвид дека P е веројатност, добиваме:

$$\begin{aligned} P_B\left(\bigcup_{i=1}^{\infty} A_i\right) &= P_B\left(\sum_{i=1}^{\infty} A_i\right) = \frac{P\left(\left(\sum_{i=1}^{\infty} A_i\right)B\right)}{P(B)} = \frac{P\left(\sum_{i=1}^{\infty} A_i B\right)}{P(B)} = \frac{\sum_{i=1}^{\infty} P(A_i B)}{P(B)} = \\ &= \sum_{i=1}^{\infty} \frac{P(A_i B)}{P(B)} = \sum_{i=1}^{\infty} P_B(A_i). \end{aligned}$$

Значи, и третата аксиома е исполнета, т. е. P_B е веројатност. ■

Значи, тројката (Ω_B, F_B, P_B) е простор на веројатност.

Теорема 3.1.4. (Формула за веројатност од производ на два настани). Нека се дадени настани A и B . Тогаш веројатноста на нивниот производ е:

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

Доказ. Од дефиницијата на условна веројатност, важат формулата за веројатност на појава на настанот A при услов да се појавил настанот B , $P(A|B) = \frac{P(AB)}{P(B)}$, и веројатноста за појава на

настанот B при услов да се појавил настанот A , $P(B|A) = \frac{P(AB)}{P(A)}$.

Оттука:

$$P(AB) = P(A|B)P(B) = P(B|A)P(A). \blacksquare$$

Задача 3.1.5. Во една кутија има 7 бели и 3 црни топчиња. Од кутијата се извлекуваат две топчиња, едно по друго без враќање. Определи ја веројатноста двете извлечени топчиња да се бели.

Решение. Да ги означиме настаните:

A : во првото извлекување е извлечено бело топче.

B : во второто извлекување е извлечено бело топче, откако претходно во првото извлекување било извлечено бело топче.

Важи:

$$P(A) = \frac{7}{10} \text{ и } P(B|A) = \frac{6}{9} = \frac{2}{3}.$$

Оттука, бараната веројатност е:

$$P(AB) = P(B|A)P(A) = \frac{2}{3} \frac{7}{10} = \frac{7}{15} = 0,4(6). \blacksquare$$

Задача 3.1.6. Ако:

$$P(A|B) = \frac{6}{13}, \quad P(A) = \frac{9}{20} \text{ и } P(A|\bar{B}) = \frac{3}{7};$$

да се најде: $P(B)$ и $P(\bar{A} \cup \bar{B})$.

Решение. Нека:

$$a = P(AB), \quad b = P(A|B), \quad c = P(B|A) \text{ и } d = P(\overline{A \cup B}).$$

Од условите на задачата добиваме:

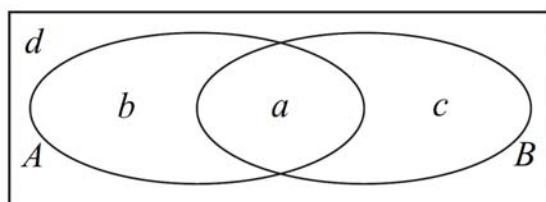
$$P(A|B) = \frac{a}{a+c} = \frac{6}{13} \Leftrightarrow a = \frac{6}{13}(a+c) \Leftrightarrow c = \frac{7}{6}a;$$

$$P(A|\bar{B}) = \frac{b}{b+d} = \frac{3}{7} \Leftrightarrow b = \frac{3}{7}(b+d) \Leftrightarrow d = \frac{4}{3}b \text{ и}$$

$$P(A) = a+b = \frac{9}{20} \Leftrightarrow b = \frac{9}{20} - a.$$

Оттука:

$$d = \frac{4}{3} \left(\frac{9}{20} - a \right) = \frac{3}{5} - \frac{4}{3}a.$$



Цртеж 3.1.6.

Настаните чии веројатности се: a, b, c и d ; се попарно дисјунктни и нивната унија е Ω . Следува дека:

$$a + b + c + d = 1 \Leftrightarrow \frac{9}{20} + \frac{7}{6}a + \frac{3}{5} - \frac{4}{3}a = 1 \Leftrightarrow \frac{1}{20} = \frac{1}{6}a \Leftrightarrow a = \frac{3}{10}.$$

Затоа:

$$b = \frac{9}{20} - \frac{3}{10} = \frac{3}{20}, \quad c = \frac{7}{6} \cdot \frac{3}{10} = \frac{7}{20} \quad \text{и} \quad d = \frac{3}{5} - \frac{4}{3} \cdot \frac{3}{10} = \frac{1}{5}.$$

Конечно:

$$P(B) = a + c = \frac{3}{10} + \frac{7}{20} = \frac{13}{20}$$

и

$$P(\bar{A} \cap \bar{B}) = b + c + d = \frac{3}{20} + \frac{7}{20} + \frac{1}{5} = \frac{14}{20} = \frac{7}{10}.$$

Со помош на принципот на математичка индукција се докажува точноста на формулата за веројатност од конечен производ на настани.

Теорема 3.1.7. (Формула за веројатност од производ на настани). Нека се дадени настаните $A_1, A_2, \dots, A_n$. Тогаш веројатноста на нивниот производ е:

$$P(A_1, A_2, \dots, A_n) = P(A_1)P(A_2 | A_1) \dots P(A_n | A_1 A_2 \dots A_{n-1}). \blacksquare$$

3.2. НЕЗАВИСНИ НАСТАНИ

Независност на два настани. Претходно уврдивме дека информацијата за појава на настан B може да влијае на веројатноста за појава на друг настан A . Но често појавата на настанот B не влијае на веројатноста за појава на настанот A . На пример веројатноста за појава на настанот „ A : падна шестка“, не се менува ако се појави настанот „ B : коцката е од производителот X “. Во тој случај важи $P(A|B) = P(A)$. Да ги разгледаме настаните што го поседуваат ова својство во произволен простор на веројатност.

Дефиниција 3.2.1. Нека (Ω, F, P) е простор на веројатност и нека се $A, B \in F$ такви што $P(A), P(B) > 0$. Настанот A е **независен** од настанот B , ако $P(A|B) = P(A)$.

Значи, настанот A е **независен** од настанот B , ако појавата на настанот B не ја менува веројатноста за појава на настанот A .

Ако настанот A е независен од настанот B , тогаш со примена на формулата за производ на настани и претходната дефиниција, добиваме:

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

Следува дека и настанот B е независен од настанот A . Затоа понатаму ќе велиме дека настаните A и B се **независни**.

Теорема 3.2.2. (Критериум за независност на два настани). Настаните A и B се независни ако и само ако:

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

Доказ. Нека настаните A и B се независни. Следува дека: $P(A|B) = P(A)$. Тогаш:

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

Нека $P(AB) = P(A)P(B)$. Тогаш:

$$P(A|B) \stackrel{\text{def}}{=} \frac{P(AB)}{P(B)} = \frac{P(A)P(B)}{P(B)} = P(A). \blacksquare$$

Теорема 3.2.3. Ако настаните A и B се независни, тогаш независни се и настаните:

$$1) A \text{ и } \bar{B}; 2) \bar{A} \text{ и } B; \text{ и } 3) \bar{A} \text{ и } \bar{B}.$$

Доказ. 1) Нека A и B се независни настани, т. е.:

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

Тогаш заради:

$$A = A\Omega = A(B + \bar{B}) = AB + A\bar{B},$$

добиваме:

$$\begin{aligned} P(A) &= P(AB) + P(A\bar{B}) \Leftrightarrow \\ P(A\bar{B}) &= P(A) - P(AB) = P(A) - P(A)P(B) = P(A)(1 - P(B)) = P(A)P(\bar{B}). \end{aligned}$$

Следува дека настаните A и $\bar{B}$ се независни.

2-3) Другите случаи се докажуваат аналогно, а може да се докажат и со помош на 1). Имено, бидејќи B и A се независни тогаш независни се и B и $\bar{A}$, а бидејќи $\bar{A}$ и B се независни, такви се и $\bar{A}$ и $\bar{B}$. $\blacksquare$

Независност на конечен број на настани. Во овој дел преку следниве две дефиниции ќе го обопштите поимот за независност на два настани.

Дефиниција 3.2.4. Настаните $A_1, A_2, \dots, A_n$ се **независни по парови** ако сите парови од настани A_i и A_j , $i, j \in \{1, 2, \dots, n\}$, $i \neq j$, се независни.

Дефиниција 3.2.5. Настаните $A_1, A_2, \dots, A_n$ се **потполно независни** ако важи:

$$P(A_{i_1}, A_{i_2}, \dots, A_{i_k}) = P(A_{i_1})P(A_{i_2}) \dots P(A_{i_k})$$

за секој број k , $k \in \{2, 3, \dots, n\}$ и за кои било индекси $i_1, i_2, \dots, i_k \in \{1, 2, \dots, n\}$, $i_1 < i_2 < \dots < i_k$.

Од дефинициите следува дека ако n настани се потполно независни, тогаш се и независни по парови. Следниов пример покажува дека обратното тврдење не мора да важи.

Задача 3.2.6. Три страни на правилен хомоген тетраедар се обоени со црвена, жолта и зелена боја соодветно, додека четвртата страна ги содржи сите три бои. Тетраедарот се фрла на маса и се набљудува бојата на страната што лежи на масата. Да се провери дали настаните:

A : на страната има црвена боја,
 B : на страната има жолта боја,
 C : на страната има зелена боја.

се независни по парови или потполно независни.

Решение. Важи:

$$P(A) = P(B) = P(C) = \frac{2}{4} = \frac{1}{2},$$

од каде што:

$$P(A)P(B) = P(A)P(C) = P(B)P(C) = \frac{1}{2} \frac{1}{2} = \frac{1}{4} \text{ и}$$

$$P(AB) = P(AC) = P(BC) = \frac{1}{4}.$$

Значи:

$$P(AB) = P(A)P(B), \quad P(AC) = P(A)P(C) \text{ и } P(BC) = P(B)P(C),$$

односно A , B и C се по парови независни настани.

Од друга страна:

$$P(ABC) = \frac{1}{4}, \text{ но } P(A)P(B)P(C) = \frac{1}{2} \frac{1}{2} \frac{1}{2} = \frac{1}{8} \text{ па:}$$

$$P(ABC) \neq P(A)P(B)P(C),$$

т.е. A , B и C не се потполно независни настани. ■

Од принципот на математичка индукција и теорема 3.2.2, следува точноста на следнава теорема:

Теорема 3.2.3. Ако настаните $A_1, A_2, \dots, A_n$ се потполно независни, тогаш потполно независни се и настаните $B_1, B_2, \dots, B_n$, каде што $B_k = A_k$ или $B_k = \overline{A_k}$, $k = 1, 2, \dots, n$. ■

3.3. ТОТАЛНА ВЕРОЈАТНОСТ

Дефиниција 3.3.1. Нека Ω е множество од елементарни настани. Настаните $H_1, H_2, \dots, H_n$ се **разбивање** (дисјунктна унија) на Ω , ако:

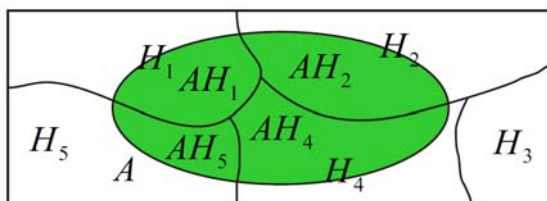
$$H_i \cap H_j = \emptyset \text{ за } i \neq j \text{ и } H_1 \cup H_2 \cup \dots \cup H_n = \Omega.$$

Со други зборови $H_1, H_2, \dots, H_n$ се разбивање на Ω ако целосно го покриваат и не го препокриваат Ω .

За настаните $H_1, H_2, \dots, H_n$ важи:

$$P(H_1) + P(H_2) + \dots + P(H_n) = 1.$$

Настаните H_i , $i = 1, 2, \dots, n$; уште се нарекуваат **хипотези**. Нивните веројатности $P(H_i)$ се нарекуваат **априорни веројатности** бидејќи се добиени пред спроведувањето на експериментот во кој се набљудува појавата на настанот A . Ако се знаат и веројатностите $P(A|H_i)$, $i = 1, 2, \dots, n$; тогаш следната теорема овозможува да се определи веројатноста $P(A)$.



Цртеж 3.3.1.

Теорема 3.3.2. (Формула за тотална веројатност). Нека настаните $H_1, H_2, \dots, H_n$ имаат позитивни веројатности и се разбивање на Ω . Тогаш за секој настан A ,

$$P(A) = P(A|H_1)P(H_1) + P(A|H_2)P(H_2) + \dots + P(A|H_n)P(H_n) = \sum_{i=1}^n P(A|H_i)P(H_i).$$

Доказ. Имаме:

$$A = A\Omega = A\left(\bigcup_{i=1}^n H_i\right) = \bigcup_{i=1}^n AH_i.$$

Од $H_i \cap H_j = \emptyset$ за $i \neq j$, следува дека $AH_i \cap AH_j = \emptyset$.

Значи настаните $AH_1, AH_2, \dots, AH_n$ се попарно дисјунктни, поради што:

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

Притоа точноста на последното равенство следува од дефиницијата за производ на настани. ■

Задача 3.3.3. Седум кутии ги имаат следниве состави: 2 кутии по 3 бели и 2 црни топчиња; 3 кутии по 3 бели и 3 црни топчиња; 1 кутија со 2 бели и 4 црни топчиња; и 1 кутија со 2 бели и 2 црни топчиња. Од произволно избрана кутија се избира едно топче. Колкава е веројатноста избраното топче да биде бело?

Решение. Да ги означиме настаните со:

A - извлеченото топче е бело.

H_1 - избрана е кутија со 3 бели и 2 црни топчиња;

H_2 - избрана е кутија со 3 бели и 3 црни топчиња;

H_3 - избрана е кутија со 2 бели и 4 црни топчиња;

H_4 - избрана е кутија со 2 бели и 2 црни топчиња.

Во задачата се бара веројатноста за појава на настанот A . Настаните $H_1, H_2, \dots, H_n$ имаат позитивни веројатности и се разбивање на Ω . Следува дека се исполнети условите од формулата за тотална веројатност.

Ги определуваме веројатностите:

$$P(H_1) = \frac{2}{7}, P(H_2) = \frac{3}{7}, P(H_3) = \frac{1}{7} \text{ и } P(H_4) = \frac{1}{7}; \text{ и}$$

$$P(A|H_1) = \frac{3}{5}, P(A|H_2) = \frac{3}{6} = \frac{1}{2},$$

$$P(A|H_3) = \frac{2}{6} = \frac{1}{3} \text{ и } P(A|H_4) = \frac{2}{4} = \frac{1}{2}.$$

Оттука:

$$\begin{aligned} P(A) &= P(A|H_1)P(H_1) + P(A|H_2)P(H_2) + \\ &P(A|H_3)P(H_3) + P(A|H_4)P(H_4) = \end{aligned}$$

$$\frac{3}{5} \cdot \frac{2}{7} + \frac{1}{2} \cdot \frac{3}{7} + \frac{1}{3} \cdot \frac{1}{7} + \frac{1}{2} \cdot \frac{1}{7} = \frac{6}{35} + \frac{3}{14} + \frac{1}{21} + \frac{1}{14} =$$

$$\frac{36 + 45 + 10 + 15}{210} = \frac{106}{210} = \frac{53}{105} \approx 0,505.$$

Задача 3.3.4. Во вториот круг од локалните избори во една општина имало два кандидати за градоначалник, едниот од партија X , другиот од Y . Пред изборите 55 % од гласачите биле симпатизери на партијата X , додека 45 % на партијата Y кои планирале да гласаат за градоначалникот од својата партија. После кампањата меѓу двата круга мислењето на дел од гласачите се променило и 80 % од симпатизерите на партијата X и 10 % од симпатизерите на партијата Y гласале за кандидатот на партијата X ; додека 20 % од симпатизерите на партијата X и 90 % од симпатизерите на партијата Y гласале за кандидатот на партијата Y . Колкава е веројатноста произволно избрано лице да гласа за кандидатот од партијата Y .

Решение. Нека настаните се:

A - избраното лице гласа за кандидат на партијата Y ;

H_1 - избраното лице било симпатизер на партијата X ;

H_2 - избраното лице било симпатизер на партијата Y .

Веројатностите на настаните кои се бараат во формулата за тотална веројатност се:

$$P(H_1) = \frac{55}{100} = 0,55, \quad P(H_2) = \frac{45}{100} = 0,45;$$

$$P(A|H_1) = \frac{20}{100} = 0,2 \text{ и } P(A|H_2) = \frac{90}{100} = 0,9.$$

Следува:

$$P(A) = P(A|H_1)P(H_1) + P(A|H_2)P(H_2) =$$

$$0,55 \cdot 0,2 + 0,45 \cdot 0,9 = 0,11 + 0,405 = 0,515.$$

3.4. БЕЈЕСОВИ ФОРМУЛИ

Ако се реализира настанот A за хипотезите H_i , можат да се пресметаат нивните условни веројатности $P(H_i | A)$, наречени **апостериорни веројатности** или веројатности добиени по спроведувањето на експериментот. Бејесовите формули овозможуваат пресметување на апостериорните веројатности, во случаите кога сакаме да утврдиме како појавата на настанот A влијаела на веројатноста за појава на хипотезите.

Теорема 3.4.1. (Бејесови формули). Нека настаните $H_1, H_2, \dots, H_n$ се разбивање на Ω и имаат позитивни веројатности. Тогаш за секој настан A ,

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

Доказ. Од дефиницијата за производ на настани и претходната теорема следува дека:

$$P(H_j | A) = \frac{P(AH_j)}{P(A)} = \frac{P(A | H_j)P(H_j)}{\sum_{i=1}^n P(A | H_i)P(H_i)}. \blacksquare$$

Задача 3.4.2. Во магацин е извршено тестирање на еден производ. Се покажало дека тестираниот производ е неисправен. Притоа сите производи биле добиени од три фабрики претходниот ден. Бројот на производите добиени од првата фабрика бил три пати поголем од бројот на производите добиени од втората фабрика. Бројот на производите добиени од втората фабрика бил два пати поголем од бројот на производите добиени од третата фабрика. Колкава е веројатноста тестираниот производ да е изработен во третата фабрика, ако се знае дека во првата фабрика се изработуваат 2 % неисправни производи, во втората 3% и во третата 4 %.

Решение. Ги означуваме следниве настани:

A - тестираниот производ е неисправен;

H_1 - производот е изработен во првата фабрика;

H_2 - производот е изработен во втората фабрика;

H_3 - производот е изработен во третата фабрика.

Нека $P(H_1) = x$. Тогаш $P(H_2) = \frac{x}{3}$ и $P(H_3) = \frac{1}{2} \frac{x}{3} = \frac{x}{6}$. Од:

$$P(H_1) + P(H_2) + P(H_3) = 1,$$

следува дека:

$$x + \frac{x}{3} + \frac{x}{6} = 1 \Leftrightarrow \frac{6+2+1}{6}x = 1 \Leftrightarrow \frac{9}{6}x = 1 \Leftrightarrow \frac{3}{2}x = 1 \Leftrightarrow x = \frac{2}{3}.$$

Значи, $P(H_1) = \frac{2}{3}$, $P(H_2) = \frac{2}{9}$ и $P(H_3) = \frac{1}{9}$. Понатаму:

$$P(A|H_1) = 0,02, \quad P(A|H_2) = 0,03 \text{ и } P(A|H_3) = 0,04.$$

Според Бејесовите формули добиваме:

$$P(H_3|A) = \frac{P(A|H_3)P(H_3)}{P(A|H_1)P(H_1) + P(A|H_2)P(H_2) + P(A|H_3)P(H_3)} =$$

$$\frac{\frac{4}{100} \frac{1}{9}}{\frac{2}{100} \frac{2}{3} + \frac{3}{100} \frac{2}{9} + \frac{4}{100} \frac{1}{9}} = \frac{\frac{4}{900}}{\frac{12+6+4}{900}} = \frac{4}{22} = \frac{2}{11} \approx 0,18.$$

Задача 3.4.3. Двајца стрелци, независно еден од друг стрелаат во определена цел. Веројатноста целта да биде погодена од првиот стрелец со еден истрел е 0,8; додека соодветната веројатност од вториот стрелец е 0,4. Двата стрелци стрелале само еднаш и истовремено, при што целта била погодена само еднаш. Колкава е веројатноста целта да ја погодил првиот стрелец.

Решение. Да ги означиме настаните со:

A - целта е погодена еднаш;

H_1 - целта не е погодена;

H_2 - целта е погодена од првиот стрелец;

H_3 - целта е погодена од вториот стрелец;

H_4 - целта е погодена од двајцата стрелци.

Настанот H_1 се случува ако првиот стрелец промаши и вториот стрелец промаши. Следува:

$$P(H_1) = (1 - 0,8)(1 - 0,4) = 0,2 \cdot 0,6 = 0,12.$$

Настанот H_2 се случува ако првиот стрелец погоди и вториот стрелец промаши. Следува:

$$P(H_2) = 0,8 \cdot (1 - 0,4) = 0,8 \cdot 0,6 = 0,48.$$

Слично:

$$P(H_3) = (1 - 0,8) \cdot 0,4 = 0,2 \cdot 0,4 = 0,08 \text{ и}$$

$$P(H_4) = 0,8 \cdot 0,4 = 0,32.$$

Условните веројатности за појава на настанот A при услов да се појавила секоја од хипотезите се:

$$P(A|H_1) = 0, \quad P(A|H_2) = 1 \text{ и } P(A|H_3) = 1 \text{ и } P(A|H_4) = 0.$$

Конечно бараната веројатност е:

$$P(H_2|A) = \frac{P(A|H_2)P(H_2)}{P(A|H_1)P(H_1) + P(A|H_2)P(H_2) + P(A|H_3)P(H_3)} = \frac{0,48}{0,48 + 0,08} = \frac{6}{7} \approx 0,857.$$

4. СЕРИЈА НЕЗАВИСНИ ЕКСПЕРИМЕНТИ (БЕРНУЛИЕВА ШЕМА). НАЈВЕРОЈАТЕН БРОЈ. ПРИБЛИЖНИ ФОРМУЛИ.

4.1. БЕРНУЛИЕВА ШЕМА

При примената на теоријата на веројатност, често пати се јавува потреба еден експеримент да се повтори повеќе пати под исти услови. Притоа се набљудува само појавата на еден настан A . Серијата од n вакви независни експерименти се нарекува **Бернулиева шема**.

Нека p е веројатноста за појава на настанот A во Бернулиевата шема од n експерименти. Тогаш $q = 1 - p$ е веројатноста да не се појави настанот A , односно да се појави настанот $\bar{A}$.

Теорема 4.1.1. (Биномна веројатност). Нека е дадена серија од n независни експерименти, p е веројатноста за појава на настанот A во секој од експериментите, $q = 1 - p$ и $P_n(k)$ е веројатноста настанот A да се појави k пати во серијата, $k \in \{0, 1, \dots, n\}$. Тогаш:

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

Доказ. Настанот A во Бернулиевата шема ќе се појави k пати, ако се појави k пати и не се појави во $n - k$ пати во сите експерименти. Во секој од n -те експерименти настанот A може k пати да се појави на:

$$C_n^k = \binom{n}{k} \text{ начини (распореда).}$$

(На пример, еден распоред е да се појави A во првите k и да не се појави во следните $n - k$ експерименти).

Поради независноста на настаните, веројатноста настанот A да се појави при конкретен распоред е $\underbrace{pp \dots p}_k \underbrace{qq \dots q}_{n-k} = p^k q^{n-k}$. Затоа:

$$P_n(k) = \binom{n}{k} p^k q^{n-k}. \blacksquare$$

Задача 4.1.2. Коцка се фрла 7 пати. Која е веројатноста да се паднат 3 шестки.

Решение. Задачата се решава според формулата за биномна веројатност. Бројот на фрлања на коцка (експерименти) е $n = 7$. Бројот на паднати шестки $k = 3$. Веројатноста да падне шестка е $p = \frac{1}{6}$, додека да не падне $q = \frac{5}{6}$. Следува дека бараната веројатност е:

$$P_n(k) = \binom{n}{k} p^k q^{n-k} \Leftrightarrow$$

$$P_7(3) = \binom{7}{3} \left(\frac{1}{6}\right)^3 \left(\frac{5}{6}\right)^4 = \frac{7!}{4!3!} \frac{5^4}{6^7} = \frac{7 \cdot 5^4}{2^7 3^7} \approx 0,078. \blacksquare$$

Задача 4.1.3. Еден прибор за работа се состои од 8 различни елементи. Приборот може да се користи ако барем 6 од елементите се исправни. Веројатноста во текот на еден час работа да се расипе некој од елементите е 0,2. Да се одреди веројатноста во текот на еден час работа приборот да се оштети толку да понатаму не ќе може да биде употреблив.

Решение. Веројатноста после еден час работа еден елемент да се расипе е $p = 0,2$, додека да не се расипе е $q = 1 - 0,2 = 0,8$. Приборот нема да биде употреблив ако се расипат барем 3 елементи, што значи да се расипат 3, 4, ..., 8 елементи. Помалку пресметки ќе има ако се пресмета веројатноста на комплементарниот настан: да се расипат најмногу 2 елементи, односно 0, 1 или 2 елементи. Значи:

$$P = 1 - \bar{P} = 1 - (P_8(0) + P_8(1) + P_8(2)) =$$

$$1 - \binom{8}{0} 0,8^8 - \binom{8}{1} 0,2 \cdot 0,8^7 - \binom{8}{2} 0,2^2 \cdot 0,8^6 =$$

$$1 - 0,1678 - 0,3355 - 0,2936 = 1 - 0,7969 = 0,2031.$$

Следува дека веројатноста во текот на еден час работа приборот да стане неупотреблив е 0,2031.

Задача 4.1.4. Веројатноста за појавување на настанот A во 4 независни експерименти е 0,59. Која е веројатноста за

појавување на настанот A во секој од експериментите, ако се знае дека таа е иста за сите експерименти.

Решение. Нека веројатноста за појавување на настанот A во еден експеримент е p . Од условот на задачата добиваме:

$$1 - P_4(0) = 0,59 \Leftrightarrow P_4(0) = 0,41 \Leftrightarrow \binom{4}{0}(1-p)^4 = 0,41 \Leftrightarrow$$

$$(1-p)^4 = 0,41 \Leftrightarrow 1-p = \sqrt[4]{0,41} \Leftrightarrow 1-p \approx 0,8 \Leftrightarrow p \approx 0,2.$$

Следува дека веројатноста за појавување на настанот A во еден експеримент приближно е $0,2$.

Задача 4.1.5. Извршени се 4 независни стрелања од еден авион во друг авион. Веројатноста за погодок во секоја од стрелбите е $0,3$. За да се собори авионот доволни се два погодока, додека веројатноста за соборување со еден погодок е $0,6$. Да се определи веројатноста авионот да биде соборен.

Решение. Ги означуваме следниве настани:

A -авионот е соборен;

H_i -авионот е погоден i -пати, $i = 0, 1, 2, 3, 4$.

Настаните H_i , $i = 0, 1, 2, 3, 4$; се дисјунктна унија на множеството елементарни настани. Затоа веројатноста на настанот A ќе се најде со помош на формулата за тотална веројатност. Веројатноста на хипотезите ќе се пресмета со помош на формулата за биномна веројатност. Веројатноста за погодок во една стрелба е $p = 0,3$. Оттука $q = 0,7$. Затоа:

$$P(H_0) = P_4(0) = C_4^0 \cdot 0,3^0 \cdot 0,7^4 = 0,24;$$

$$P(H_1) = P_4(1) = C_4^1 \cdot 0,3^1 \cdot 0,7^3 = 4 \cdot 0,3 \cdot 0,343 = 0,4116;$$

$$P(H_2) = P_4(2) = C_4^2 \cdot 0,3^2 \cdot 0,7^2 = 6 \cdot 0,09 \cdot 0,49 = 0,2646;$$

$$P(H_3) = P_4(3) = C_4^3 \cdot 0,3^3 \cdot 0,7^1 = 4 \cdot 0,027 \cdot 0,7 = 0,0756;$$

$$P(H_4) = P_4(4) = C_4^4 \cdot 0,3^4 \cdot 0,7^0 = 0,0081;$$

и

$$P(A | H_0) = 0, \quad P(A | H_1) = 0,6 \text{ и}$$

$$P(A | H_2) = P(A | H_3) = P(A | H_4) = 1$$

Конечно:

$$\begin{aligned} P(A) &= P(A|H_0)P(H_0) + P(A|H_1)P(H_1) + \\ &P(A|H_2)P(H_2) + P(A|H_3)P(H_3) + P(A|H_4)P(H_4) = \\ &0,6 \cdot 0,4116 + 0,2646 + 0,0756 = 0,595. \end{aligned}$$

Значи дека веројатноста авионот да биде соборен е 0,595.

4.2. НАЈВЕРОЈАТЕН БРОЈ

Во Бернулиевата шема настанот A може да се појави од 0 до n -пати со веројатности $P_n(k)$, $k = 0, 1, \dots, n$; соодветно. Сакаме да утврдиме за кој број $k = k_0$ веројатноста ќе биде најголема. Тој број ќе го наречеме **најверојатен број**. Имаме:

$$\begin{aligned} \frac{P_n(k+1)}{P_n(k)} &> 1 \Leftrightarrow \frac{\binom{n}{k+1}}{\binom{n}{k}} \frac{p^{k+1}q^{n-k-1}}{p^k q^{n-k}} > 1 \Leftrightarrow \\ &\frac{n!}{(k+1)!(n-k-1)!} \frac{p}{q} > 1 \Leftrightarrow \frac{n-k}{k+1} \frac{p}{q} > 1 \Leftrightarrow \\ &\frac{n!}{k!(n-k)!} \frac{p}{q} > 1 \Leftrightarrow np - kp > kq + q \Leftrightarrow np - kp > k(1-p) + q \Leftrightarrow \\ &np - \underline{kp} > \underline{k} - \underline{kp} + q \Leftrightarrow k < np - q \quad (1). \end{aligned}$$

Значи за $k < np - q$, важи: $P_n(k) < P_n(k+1)$.

Од друга страна заради еквивалентност на чекорите во еквиваленциите (1) за:

$$k > np - q, P_n(k) > P_n(k+1) \text{ и за } k = np - q, P_n(k) = P_n(k+1).$$

Значи дека:

- Ако $np - q$ не е природен број, тогаш:

$$\begin{aligned} P_n(0) &< P_n(1) < \dots < P_n([np - q]) < P_n([np - q] + 1) > \\ &P_n([np - q] + 2) > \dots > P_n(n), \end{aligned}$$

па $[np - q] + 1$ е најверојатниот број.

- Ако $np - q$ е природен број, тогаш:

$$P_n(0) < P_n(1) < \dots < P_n(np - q) = P_n(np - q + 1) > \dots > P_n(n),$$

па $np - q$ и $np - q + 1$ се најверојатните броеви.

Следува дека е покажана точноста на следнава теорема:

Теорема 4.2.1. (Најверојатен број). Нека е дадена Бернулиева шема од n експерименти во која p е веројатноста за појава на настанот A , и $q = 1 - p$.

- 1) Ако $np - q$ не е природен број, тогаш најверојатниот број е:

$$[np - q] + 1;$$

- 2) Ако $np - q$ е природен број, тогаш најверојатни броеви се:

$$np - q \text{ и } np - q + 1.$$

Задача 4.2.2. Веројатноста еден кошаркар да постигне погодок при слободно фрлање е $0,7$. Определи го најверојатниот број на погодоци и неговата веројатност, ако кошаркарот 10 пати изведувал слободни фрлања.

Решение. Бројот на експерименти е $n = 10$, веројатноста да погоди при едно фрлање е $p = 0,7$, додека да промаши $q = 1 - 0,7 = 0,3$. Бидејќи:

$$np - q = 10 \cdot 0,7 - 0,3 = 6,7$$

не е цел број, следува дека најверојатниот број на погодоци е најмалиот цел број поголем од $6,7$ т.е. $k_0 = 7$. Веројатноста да се погодат 7 погодоци е:

$$P_n(k_0) = \binom{n}{k_0} p^{k_0} q^{n-k_0} \text{ т.е.}$$

$$P_{10}(7) = \binom{10}{7} 0,7^7 0,3^3 = \frac{10!}{3!7!} 0,7^7 0,3^3 \approx 0,267.$$

Задача 4.2.3. Колку пати треба да се фрли на маса коцка за играње, за да најверојатниот број на појавување на 6-ки е 10.

Решение. Веројатноста да падне 6-ка е $p = \frac{1}{6}$, а да не падне е $q = \frac{5}{6}$. Да го означиме бројот на фрлања на коцката со n . Тогаш:

$$np - q = n \frac{1}{6} - \frac{5}{6} = \frac{n-5}{6}.$$

Бидејќи има еден најверојатен број $k_0 = 10$ важи:

$$k_0 = [np - q] + 1 \Leftrightarrow k_0 - 1 = [np - q] \stackrel{np-q \notin \mathbb{N}}{\Leftrightarrow} k_0 - 1 < np - q < k_0 \Leftrightarrow 9 < \frac{n-5}{6} < 10 \Leftrightarrow n-5 > 54, n-5 < 60 \Leftrightarrow n > 59, n < 65.$$

Следува дека бројот на фрлања на коцката е меѓу 60 и 64 односно:

$$n \in \{60, 61, 62, 63, 64\}.$$

4.3-5. Формули за приближно пресметување на веројатностите во Бернулиевата шема.

При пресметувањето на биомните веројатности може да се соочиме со сериозни технички потешкотии во случаите кога бројот на експерименти n е голем. Затоа во продолжение без доказ ќе дадеме три приближни формули. Теоремите определуваат кон кои броеви тежат веројатностите $P_n(k)$ и $P_n(a \leq k \leq b)$, кога бројот на експерименти n тежи кон бесконечност. Тоа значи дека за доволно голем број на експерименти можеме да ја најдеме веројатноста на колку сакаме точни децимали. За конкретен број на експерименти n , теоремите не го кажуваат отстапувањето од точната веројатност, кое зависи и од веројатноста p за појава на настанот во секој експеримент, но ќе продискутираме кога приближувањето е доволно добро.

4.3. ЛОКАЛНА ТЕОРЕМА НА МОАВР-ЛАПЛАС

Теоремата се користи за приближно пресметување на веројатноста $P_n(k)$ во Бернулиевата шема.

Теорема 4.3.1. (Локална теорема на Моавр-Лаплас).

Нека p , $0 < p < 1$ е веројатноста за појавување на настанот A во Бернулиеви шеми со n експерименти, $n = 1, 2, \dots$. Тогаш:

$$\lim_{n \rightarrow \infty} P_n(k) = \lim_{n \rightarrow \infty} \frac{1}{\sqrt{2\pi npq}} e^{-\frac{x^2}{2}}, \text{ каде што: } x = \frac{k - np}{\sqrt{npq}}.$$

Локалната теорема на Моавр-Лаплас тврди дека за големи вредности на n ,

$$P_n(k) \approx \frac{1}{\sqrt{2\pi npq}} e^{-\frac{x^2}{2}} \text{ каде што } x = \frac{k - np}{\sqrt{npq}}, k = 0, 1, \dots, n.$$

Задача 4.3.2. Фабрика произведува исправни производи со веројатност $p = 0,9$. Колкава е веројатноста од 100 производи, 90 да бидат исправни.

Решение. Бројот на изведени експерименти е $n = 100$, додека бројот на појавувања на настанот „произведен е исправен производ“ е $k = 90$. Се бара веројатноста $P_{100}(90)$. Според локалната теорема на Моавр-Лаплас, добиваме:

$$x = \frac{k - np}{\sqrt{npq}} = \frac{90 - 100 \cdot 0,9}{\sqrt{100 \cdot 0,9 \cdot 0,1}} = 0,$$

од каде што:

$$P_{100}(90) \approx \frac{1}{\sqrt{2\pi npq}} e^{-\frac{x^2}{2}} = \frac{1}{\sqrt{2\pi \cdot 100 \cdot 0,9 \cdot 0,1}} e^0 = \frac{1}{\sqrt{18\pi}} \approx 0,133.$$

Задача 4.3.3. Колку пати со веројатност 0,0484 може да се очекува појавување на еден настан A во една серија од 100 независни експерименти, ако веројатноста за негово појавување во секој експеримент е 0,5?

Решение. Ќе ја искористиме локалната теорема на Моавр-Лаплас. Имаме:

$$P_n(k) = 0,0484, \quad n = 100, \quad p = 0,5 \quad q = 0,5, \\ npq = 100 \cdot 0,5 \cdot 0,5 = 25 \quad \text{и} \quad \sqrt{npq} = 5.$$

Следува:

$$P_n(k) \approx \frac{1}{\sqrt{2\pi npq}} e^{-\frac{x^2}{2}} \Leftrightarrow 0,0484 \approx \frac{1}{5\sqrt{2\pi}} e^{-\frac{x^2}{2}} \Leftrightarrow 0,242\sqrt{2\pi} \approx e^{-\frac{x^2}{2}} \Leftrightarrow \\ \ln 0,242\sqrt{2\pi} \approx -\frac{x^2}{2} \Leftrightarrow x \approx \sqrt{-2 \ln 0,242\sqrt{2\pi}} \Leftrightarrow x \approx 1.$$

Оттука:

$$x = \frac{k - np}{\sqrt{npq}} \Leftrightarrow 1 = \frac{k - 100 \cdot 0,5}{5} \Leftrightarrow 5 = k - 50 \Leftrightarrow k = 55.$$

Следува дека се очекува настанот A да се појави 55 пати.

4.4. Интегрална теорема на Моавр-Лаплас.

Теоремата се користи за определување на веројатноста $P_n(a \leq k \leq b)$ во Бернулиевата шема.

Теорема 4.4.1. (Интегрална теорема на Моавр-Лаплас). Нека p , $0 < p < 1$ е веројатноста за појавување на настанот A во Бернулиеви шеми со n експерименти, $n = 1, 2, \dots$. Тогаш:

$$P_n(a \leq k \leq b) = \lim_{n \rightarrow \infty} \frac{1}{\sqrt{2\pi}} \int_{x_a}^{x_b} e^{-\frac{x^2}{2}} dx, \quad \text{каде што} \quad x_a = \frac{a - np}{\sqrt{npq}}, \quad x_b = \frac{b - np}{\sqrt{npq}} \quad (1).$$

Пресметувањето на интегралот во (1) е тешко. Затоа во практика вредностите на функцијата $\Phi(x) = \frac{1}{\sqrt{2\pi}} \int_0^x e^{-\frac{x^2}{2}} dx$ се дадени во таблица. Тогаш:

$$P_n(a \leq k \leq b) = \lim_{n \rightarrow \infty} (\Phi(x_b) - \Phi(x_a)).$$

Значи интегралната теорема на Моавр-Лаплас вели дека за големи вредности на n :

$$P_n(a \leq k \leq b) \approx \Phi(x_b) - \Phi(x_a), \text{ каде што } x_a = \frac{a - np}{\sqrt{npq}}, x_b = \frac{b - np}{\sqrt{npq}}.$$

*Нешто подобро приближување дава формулата:

$$P_n(a \leq k \leq b) \approx \Phi\left(x_{b+\frac{1}{2}}\right) - \Phi\left(x_{a-\frac{1}{2}}\right) \quad (2).$$

која кога $n > 100$ и p не е близу 0 или 1, дава резултат со точност до трета-четврта децимала, што не е случај со формулата (1). Сепак заради поедноставните пресметки и бидејќи и (1) дава добро приближување, почесто ќе ја користиме формулата (1).

$\Phi(x) = \frac{1}{\sqrt{2\pi}} \int_0^x e^{-\frac{x^2}{2}} dx, \quad \Phi(-x) = -\Phi(x) \quad \Phi(4,5) = 0,499997$ $\Phi(4) = 0,49997', \quad \Phi(5) = 0,4999997$										
	0	1	2	3	4	5	6	7	8	9
0,0	0,00000	0,00399	0,00798	0,01197	0,01595	0,01994	0,02392	0,02790	0,03188	0,03586
0,1	0,03983	0,04380	0,04776	0,05172	0,05567	0,05962	0,06356	0,06749	0,07142	0,07535
0,2	0,07926	0,08317	0,08706	0,09095	0,09483	0,09871	0,10257	0,10642	0,11026	0,11409
0,3	0,11791	0,12172	0,12552	0,12930	0,13307	0,13683	0,14058	0,14431	0,14803	0,15173
0,4	0,15542	0,15910	0,16276	0,16640	0,17003	0,17364	0,17724	0,18082	0,18439	0,18793
0,5	0,19146	0,19497	0,19847	0,20194	0,20540	0,20884	0,21226	0,21566	0,21904	0,22240
0,6	0,22575	0,22907	0,23237	0,23565	0,23891	0,24215	0,24537	0,24857	0,25175	0,25490
0,7	0,25804	0,26115	0,26424	0,26730	0,27035	0,27337	0,27637	0,27935	0,28230	0,28524
0,8	0,28814	0,29103	0,29389	0,29673	0,29955	0,30234	0,30511	0,30785	0,31057	0,31327
0,9	0,31594	0,31859	0,32121	0,32381	0,32639	0,32894	0,33147	0,33398	0,33646	0,33891
1,0	0,34124	0,34373	0,34614	0,34850	0,35083	0,35314	0,35534	0,35769	0,35993	0,36214
1,1	0,36433	0,36650	0,36864	0,37076	0,37286	0,37493	0,37698	0,37900	0,38100	0,38298
1,2	0,38493	0,38696	0,38877	0,39065	0,39251	0,39435	0,39617	0,39796	0,39973	0,40147
1,3	0,40320	0,40490	0,40658	0,40824	0,40988	0,41149	0,41309	0,41466	0,41621	0,41774
1,4	0,41924	0,42073	0,42220	0,42364	0,42507	0,42647	0,42786	0,42922	0,43056	0,42189
1,5	0,43319	0,43448	0,43574	0,43699	0,43822	0,43943	0,44062	0,44179	0,44179	0,44408
1,6	0,44520	0,44630	0,44738	0,44845	0,44950	0,45053	0,45154	0,45254	0,45352	0,45449
1,7	0,45543	0,45637	0,45728	0,45818	0,45907	0,45994	0,46080	0,46164	0,46246	0,46327
1,8	0,46407	0,46485	0,46562	0,46638	0,46712	0,46784	0,46856	0,46926	0,46995	0,47062
1,9	0,47128	0,47193	0,47257	0,47320	0,47381	0,47441	0,47500	0,47558	0,47615	0,47670
2,0	0,47725	0,47778	0,47831	0,47882	0,47932	0,47982	0,48030	0,48077	0,48124	0,48169

2,1	0,48214	0,48257	0,48300	0,48341	0,48382	0,48422	0,48461	0,48500	0,4837	0,48574
2,2	0,48610	0,48645	0,48679	0,48713	0,48745	0,48778	0,48809	0,48840	0,48870	0,48899
2,3	0,48928	0,48959	0,48983	0,49010	0,49036	0,49061	0,49086	0,49111	0,49134	0,49158
2,4	0,49180	0,49202	0,49224	0,49245	0,49266	0,49286	0,49305	0,49324	0,49343	0,49361
2,5	0,49379	0,49396	0,49413	0,49430	0,49446	0,49461	0,49477	0,49492	0,49506	0,49520
2,6	0,49534	0,49547	0,49560	0,49573	0,49585	0,49598	0,49609	0,49621	0,49632	0,49643
2,7	0,49653	0,49664	0,49674	0,49683	0,49693	0,49702	0,49711	0,49719	0,49728	0,49736
2,8	0,49744	0,49752	0,49760	0,49767	0,49774	0,49781	0,49788	0,49795	0,49801	0,49807
2,9	0,49813	0,49819	0,49825	0,49831	0,49836	0,49841	0,49846	0,49851	0,49856	0,49861
3,0	0,49865	0,49869	0,49874	0,49878	0,49882	0,49886	0,49889	0,49893	0,49897	0,49900
3,1	0,49903	0,49906	0,49910	0,49913	0,49916	0,49918	0,49921	0,49924	0,49926	0,49929
3,2	0,49931	0,49934	0,49936	0,49938	0,49940	0,49942	0,49944	0,49946	0,49948	0,49950
3,3	0,49952	0,49953	0,49955	0,49957	0,49958	0,49960	0,49961	0,49962	0,49964	0,49965
3,4	0,49966	0,49968	0,49969	0,49970	0,49971	0,49972	0,49973	0,49974	0,49975	0,49976
3,5	0,49977	0,49978	0,49978	0,49979	0,49980	0,49981	0,49982	0,49982	0,49983	0,49984
3,6	0,49984	0,49985	0,49985	0,49986	0,49986	0,49987	0,49987	0,49988	0,49988	0,49989
3,7	0,49989	0,49990	0,49990	0,49990	0,49990	0,49991	0,49991	0,49992	0,49992	0,49993
3,8	0,49993	0,49993	0,49993	0,49994	0,49994	0,49994	0,49994	0,49995	0,49995	0,49995
3,9	0,49995	0,49996	0,49996	0,49996	0,49996	0,49996	0,49996	0,49996	0,49997	0,49997

Табела 4.4. Таблица на функцијата $\Phi(x)$.

Од дефиницијата за граница следува дека теоремата ќе подобри приближни вредности при поголеми вредности на n .

Анализите, колку големи да бидат вредностите за да се добие точност на одреден број на децимали нема да бидат предмет на овој курс. Ќе напоменеме само дека при фиксно n теоремата дава подобри приближни вредности доколку p е поблиску до 0,5.

Задача 4.4.2. Од 1000 студенти на Градежниот факултет, околу 70 % редовно ја следат наставата. Колкава е веројатноста бројот на студенти кои редовно ја следат наставата да биде меѓу 672 и 721?

Решение. Прв начин. Веројатноста еден студент да ја следи наставата е $p = 0,7$. Оттука $q = 0,3$. Бидејќи бројот на експеримент $n = 1000$ е многу голем и p не е блиску до 0 или 1, може да ја примениме интегралната теорема на Моавр-Лаплас. Имаме:

$$x_b = \frac{b - np}{\sqrt{npq}} = \frac{721 - 1000 \cdot 0,7}{\sqrt{1000 \cdot 0,7 \cdot 0,3}} \approx 1,45 \text{ и}$$

$$x_a = \frac{a - np}{\sqrt{npq}} = \frac{672 - 1000 \cdot 0,7}{\sqrt{1000 \cdot 0,7 \cdot 0,3}} \approx -1,93,$$

Од таблицата за функцијата $\Phi(x)$ наоѓаме дека:

$$\Phi(x_b) \approx 0,42647 \text{ и } \Phi(x_a) \approx -0,47320.$$

Следува дека:

$$P_n(a \leq k \leq b) \approx \Phi(x_b) - \Phi(x_a) \approx 0,42647 - (-0,47320) \approx 0,900.$$

Втор начин. Ако ја користевме приближната формула (2), се добива:

$$x_{b+\frac{1}{2}} = 1,48, \quad \Phi\left(x_{b+\frac{1}{2}}\right) \approx 0,43056; \text{ и}$$

$$x_{a-\frac{1}{2}} = -1,97, \quad \Phi\left(x_{a-\frac{1}{2}}\right) \approx 0,47558,$$

од каде што:

$$P_n(a \leq k \leq b) \approx 0,90614 \approx 0,906. \blacksquare$$

Задача 4.4.3. Колку независни експерименти треба да се извршат за да веројатноста за појава на еден настан A повеќе од 4 пати изнесува $0,8$, ако веројатноста за појава на настанот A во секој од експериментите е $0,05$.

Решение. Потребно е да се одреди бројот на експерименти n таков што $P_n(5 \leq k \leq n) \approx 0,8$. Имаме:

$$p = 0,05, \quad q = 0,95, \quad a = 5 \text{ и } b = n.$$

Следува:

$$x_b = \frac{b - np}{\sqrt{npq}} = \frac{n(1 - p)}{\sqrt{npq}} = \frac{nq}{\sqrt{npq}} = \sqrt{\frac{nq}{p}} = \sqrt{\frac{0,95n}{0,05}} = \sqrt{19n} \approx 4,36\sqrt{n},$$

$$\Phi(x_b) \approx \frac{1}{2} \text{ и } x_a = \frac{a - np}{\sqrt{npq}} = \frac{5 - \frac{1}{20}n}{\sqrt{\frac{1}{20} \cdot \frac{19}{20}n}} = \frac{\frac{1}{20}(100 - n)}{\frac{1}{20}\sqrt{19n}} = \frac{100 - n}{\sqrt{19n}}.$$

Од друга страна:

$$P_n(5 \leq k \leq n) \approx \Phi(x_b) - \Phi(x_a) \Leftrightarrow 0,8 \approx 0,5 - \Phi(x_a) \Leftrightarrow$$

$$\Phi(x_a) \approx -0,3 \Leftrightarrow x_a \approx -0,84 = -\frac{84}{100} = -\frac{21}{25}.$$

Значи:

$$\begin{aligned}\frac{100-n}{\sqrt{19n}} &\approx -\frac{21}{25} \Leftrightarrow 2500-25n \approx -21\sqrt{19}\sqrt{n} \Leftrightarrow \\ 25n-21\sqrt{19}\sqrt{n}-2500 &\approx 0 \Leftrightarrow \\ \sqrt{n}_{1,2} &= \frac{21\sqrt{19} \pm \sqrt{441 \cdot 19 + 4 \cdot 25 \cdot 2500}}{50} \Leftrightarrow \\ \sqrt{n}_{1,2} &= \frac{91,54 \pm \sqrt{8379 + 250000}}{50} \approx \frac{91,54 \pm 508,31}{50}.\end{aligned}$$

Бидејќи коренот не е негативен ја избираме позитивната вредност:

$$\sqrt{n} \approx \frac{91,54 + 508,31}{50} = \frac{599,85}{50} = 11,997 \Leftrightarrow n \approx 143,928 \Leftrightarrow n \approx 144.$$

Следува дека потребно е да се извршат 144 експерименти. ■

4.5. ТЕОРЕМА НА ПУАСОН

Доколку p е блиску до 0 и 1, теоремите на Моавр-Лаплас не даваат во општ случај задоволителни резултати за вредностите на n кои имаат практична примена. Тогаш се користи теоремата на Пуасон што ќе ја дадеме во продолжение.

Теорема 4.5.1. (Теорема на Пуасон). Нека p , $0 < p < 1$ е веројатноста за појавување на настанот A во Бернулиеви шеми од n експерименти, $n = 1, 2, \dots$. Ако $\lim_{n \rightarrow \infty} np = \lambda$, тогаш:

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

Теоремата на Пуасон тврди дека за големи вредности на n и мали на p , $P_n(k) \approx \frac{\lambda^k}{k!} e^{-\lambda}$, $\lambda = np$, $k = 0, 1, \dots, n$.

Теоремата дава добри приближни вредности во таканаречените специјални случаи кога бројот на експерименти n е голем, додека веројатноста p за појава на настанот A мала. Анализите на оценките на грешките, покажуваат дека најчесто се

зема $n \geq 100$ и $np \leq 10$. Во останатите случаи теоремите на Моавр-Лаплас даваат подобри приближни вредности.

Ако вредноста p за појава на настанот A е близу до 1, тогаш наместо веројатноста настанот A да се појави k -пати во серијата од n експерименти, ја разгледуваме веројатноста настанот $\bar{A}$ со веројатност на појавување $1-p$ да се појави $n-k$ -пати. И во случај кога бројот на експерименти n е голем, ја применуваме теоремата на Пуасон.

Задача 4.5.2. Веројатноста еден телевизор да биде произведен со дефект е 0,002. Колкава е веројатноста од 500 случајно избрани телевизори да има: 2 телевизори со дефект.

Решение. Веројатноста еден телевизор да биде произведен со дефект е $p = 0,002$, додека бројот на телевизори е $n = 500$. Оттука $\lambda = np = 1$. Следува дека се исполнети претпоставките за примена на теоремата на Пуасон, $P_n(k) \approx \frac{\lambda^k}{k!} e^{-\lambda}$. Веројатноста за $k = 2$ производа да бидат со дефект е:

$$P_{500}(2) \approx \frac{1}{2} e^{-1} = \frac{1}{2e} \approx 0,184.$$

Задача 4.5.3. Веројатноста еден телевизор да биде произведен со дефект е 0,002. Колкава е веројатност од 500 случајно избрани телевизори да има барем 3 телевизори со дефект.

Решение. Имаме $p = 0,002$, $n = 500$, од каде што $\lambda = np = 1$. Ја применуваме теоремата на Пуасон. Настанот „барем 3 производа се со дефект“ е комплементарен со настанот „најмногу 2 производа се со дефект“. Следува дека неговата веројатност е:

$$P_{500}(k \geq 3) = 1 - P_{500}(0 \leq k \leq 2) = 1 - (P_{500}(0) + P_{500}(1) + P_{500}(2)) \approx$$

$$1 - \frac{1}{e} \left(\frac{1}{0!} + \frac{1}{1!} + \frac{1}{2!} \right) = \frac{5}{2e} \approx 1 - 0,920 = 0,080.$$

Пример 4.5.4*. Да ја разгледаме задачата 4.3.2. Веројатноста од 100 производа, 90 да се исправни е иста со веројатноста од 100 производа 10 да се неисправни. Притоа

веројатноста за неисправен производ е $p = 1 - 0,9 = 0,1$. Според теоремата на Пуасон би добиле:

$$P_{100}(10) \approx \frac{10^{10}}{10!} e^{-10} \approx 0,125.$$

Точната вредност заокружена на 3 децимали е 0,132. Според теоремата на Моавр-Лаплас добивме дека веројатноста е 0,133. Можеме да заклучиме дека во сите три случаи добиената веројатност заокружена на две децимали е иста, значи теоремите даваат добар приближен резултат, иако во конкретниот пример теоремата на Моавр-Лаплас дава подобра приближна вредност.

5. СЛУЧАЈНИ ПРОМЕНЛИВИ

5.1. СЛУЧАЈНИ ПРОМЕНЛИВИ

Често се случува да разгледуваме пресликување од множеството елементарни настани во множеството реални броеви.

Пример 5.1.1. Експеримент се состои од фрлање на коцка. Множеството елементарни настани е:

$$\Omega = \{\omega_1, \omega_2, \dots, \omega_6\},$$

каде што исходите ω_i , $i = 1, 2, \dots, 6$; соодветно означуваат дека на горната страна на фрлената коцка паднале i точки. Во овој случај имаме пресликување $X : \Omega \rightarrow \mathbb{R}$ зададено со:

$$X(\omega_i) \rightarrow i, \quad i = 1, 2, \dots, 6.$$

Пресликувањето е биекција и затоа елементарниот настан го поистоветуваме со сликата при пресликувањето т.е. велиме на коцката паднал бројот i , $i = 1, 2, \dots, 6$.

Нека (Ω, F, P) е простор на веројатност.

Дефиниција 5.1.2. Пресликувањето $X : \Omega \rightarrow \mathbb{R}$ во кое $\{\omega | X(\omega) < x\}$ е настан за секој $x \in \mathbb{R}$ се нарекува **случајна променлива**.

Ако Ω е конечно множество тогаш секое подмножество од Ω е настан, од каде што следува дека секоја функција од $X : \Omega \rightarrow \mathbb{R}$ е случајна променлива.

Случајната променлива X е реална функција со својство: за секој $x \in \mathbb{R}$ множеството $\{\omega | X(\omega) < x\}$ е случаен настан (настан) т.е. елемент на σ -алгебрата F . Настанот ќе го означиме со $(X < x)$ т.е.

$$(X < x) = \{\omega | X(\omega) < x\}.$$

За настанот $(X < x)$ е дефинирана веројатност $P((X < x))$ која ќе ја означуваме со:

$$P(X < x).$$

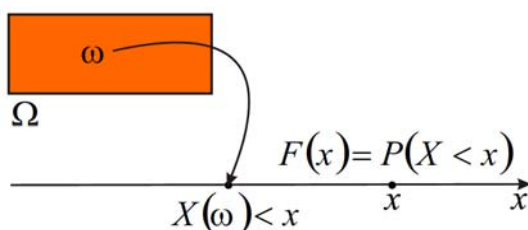
Затоа следниот поим е добро дефиниран.

Дефиниција 5.1.5. Функцијата $F: \mathbb{R} \rightarrow [0, 1]$ дефинирана со:

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

се нарекува **функција на распределба на случајната променлива X** .

Кога истовремено разгледуваме повеќе случајни променливи, нив ќе ги означуваме со големите латински букви $X, Y, Z, \dots$ додека соодветните функции на распределба со $F_X(x), F_Y(x), F_Z(x), \dots$



Цртеж 5.1.5. Шематски приказ на множество елементарни настани, случајна променлива и функција на распределба

Теорема 5.1.6. Нека F е функција на распределба на случајната променлива X . Тогаш:

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

Доказ. Бидејќи $(-\infty, b) = (-\infty, a) \cup [a, b)$, при што десната унија е дисјунктна, важи:

$$\begin{aligned} P(X < b) &= P(X < a) + P(X \in [a, b)) \Leftrightarrow \\ F(b) &= F(a) + P(a \leq X < b) \Leftrightarrow P(a \leq X < b) = F(b) - F(a). \blacksquare \end{aligned}$$

Теорема 5.1.7. Нека F е функција на распределба на случајната променлива X . Тогаш:

1) F е монотонно неопадачка т.е. за секои $x_1, x_2 \in \mathbb{R}$, такви што од $x_1 < x_2$ следува дека $F(x_1) \leq F(x_2)$;

2) $\lim_{x \rightarrow -\infty} F(x) = 0$ и $\lim_{x \rightarrow +\infty} F(x) = 1$;

3) За секое $x_0 \in \mathbb{R}$, F е непрекината од лево во x_0 , т. е.:

$$\lim_{x \rightarrow x_0^-} F(x) = F(x_0).$$

Доказ. 1) Нека $x_1 < x_2$. Тогаш $(X < x_1) \subseteq (X < x_2)$, од каде што следува дека:

$$F(x_1) = P(X < x_1) \leq P(X < x_2) = F(x_2).$$

2). а) Нека (x_n) е опаѓачка низа од реални броеви т. е.: $x_n > x_{n+1}$ за секое $n \in \mathbb{N}$ и $\lim_{n \rightarrow \infty} x_n = -\infty$.

Ги разгледуваме множествата $(X < x_n)$, $n \in \mathbb{N}$. Важи:

$$(X < x_1) \supseteq (X < x_2) \supseteq \dots \supseteq (X < x_n) \text{ и } \bigcap_{i=1}^n (X < x_i) = (X < x_n),$$

Од теоремата за непрекинатост на веројатноста следува дека:

$$P\left(\bigcap_{i=1}^{\infty} (X < x_i)\right) = \lim_{n \rightarrow \infty} P\left(\bigcap_{i=1}^n (X < x_i)\right) = \lim_{n \rightarrow \infty} P(X < x_n) = \lim_{n \rightarrow \infty} F(x_n) \quad (1).$$

Од друга страна:

$$\bigcap_{i=1}^{\infty} (X < x_i) = \emptyset, \text{ па } P\left(\bigcap_{i=1}^{\infty} (X < x_i)\right) = P(\emptyset) = 0 \quad (2).$$

Од (1) и (2) следува дека:

$$\lim_{x \rightarrow -\infty} F(x) = \lim_{n \rightarrow \infty} F(x_n) = 0.$$

б) Нека (x_n) растечка низа од реални броеви т.е. $x_n < x_{n+1}$ за секое $n \in \mathbb{N}$ и $\lim_{n \rightarrow \infty} x_n = +\infty$.

Ги разгледуваме множествата $(X < x_n)$, $n \in \mathbb{N}$. Важи:

$$(X < x_1) \subseteq (X < x_2) \subseteq \dots \subseteq (X < x_n)$$

и

$$\bigcup_{i=1}^n (X < x_i) = (X < x_n),$$

од каде што заради теоремата за непрекинатост на веројатноста важи:

$$P\left(\bigcup_{i=1}^{\infty} (X < x_i)\right) = \lim_{n \rightarrow \infty} P\left(\bigcup_{i=1}^n (X < x_i)\right) = \lim_{n \rightarrow \infty} P(X < x_n) = \lim_{n \rightarrow \infty} F(x_n).$$

Од друга страна:

$$\bigcup_{i=1}^{\infty} (X < x_i) = \Omega, \text{ па } P\left(\bigcup_{i=1}^{\infty} (X < x_i)\right) = P(\Omega) = 1.$$

Значи:

$$\lim_{x \rightarrow +\infty} F(x) = \lim_{n \rightarrow \infty} F(x_n) = 1.$$

3) Нека $x_0 \in \mathbb{R}$ и (x_n) растечка низа од реални броеви таква што $\lim_{n \rightarrow \infty} x_n = x_0$. Важи:

$$(X < x_1) \subseteq (X < x_2) \subseteq \dots \subseteq (X < x_n) \text{ и } \bigcup_{i=1}^n (X < x_i) = (X < x_n),$$

Оттука:

$$P\left(\bigcup_{i=1}^{\infty} (X < x_i)\right) = \lim_{n \rightarrow \infty} P\left(\bigcup_{i=1}^n (X < x_i)\right) = \lim_{n \rightarrow \infty} P(X < x_n) = \lim_{n \rightarrow \infty} F(x_n).$$

Од друга страна:

$$\bigcup_{i=1}^{\infty} (X < x_i) = (X < x_0),$$

па

$$P\left(\bigcup_{i=1}^{\infty} (X < x_i)\right) = P(X < x_0) = F(x_0).$$

Следува:

$$\lim_{x \rightarrow x_0^-} F(x) = \lim_{n \rightarrow \infty} F(x_n) = F(x_0). \blacksquare$$

Следнава теорема што ќе ја дадеме без доказ, покажува дека важи и обратното тврдење.

Теорема 5.1.8. Ако за функцијата $F : \Omega \rightarrow \mathbb{R}$ важи:

- 1) F е монотono неопаѓачка;
- 2) $\lim_{x \rightarrow -\infty} F(x) = 0$ и $\lim_{x \rightarrow +\infty} F(x) = 1$; и
- 3) За секое $x_0 \in \mathbb{R}$, $\lim_{x \rightarrow x_0^-} F(x) = F(x_0)$;

тогаш F е функција на распределба на случајната променлива X . ■

Едноелементното множество $x_0 \in \mathbb{R}$ може да се претстави како преброив пресек на настани. Имено:

$$\{x_0\} = \bigcap_{i=1}^{+\infty} \left[x_0, x_0 + \frac{1}{i} \right)$$

од каде што:

$$(X = x_0) = \bigcap_{i=1}^{+\infty} \left(X \in \left[x_0, x_0 + \frac{1}{i} \right) \right).$$

Бидејќи:

$$(X \in [x_0, x_0 + 1)) \supseteq \left(X \in \left[x_0, x_0 + \frac{1}{2} \right) \right) \supseteq \dots \supseteq \left(X \in \left[x_0, x_0 + \frac{1}{n} \right) \right)$$

и

$$\bigcap_{i=1}^n \left(X \in \left[x_0, x_0 + \frac{1}{i} \right) \right) = \left(X \in \left[x_0, x_0 + \frac{1}{n} \right) \right),$$

од теоремата за непрекинатост на веројатноста важи:

$$\begin{aligned} P(X = x_0) &= P\left(\bigcap_{i=1}^{+\infty} \left(X \in \left[x_0, x_0 + \frac{1}{i} \right) \right)\right) = \lim_{n \rightarrow \infty} P\left(\bigcap_{i=1}^n \left(X \in \left[x_0, x_0 + \frac{1}{i} \right) \right)\right) = \\ &= \lim_{n \rightarrow \infty} P\left(X \in \left[x_0, x_0 + \frac{1}{n} \right)\right) = \lim_{n \rightarrow \infty} \left(x_0 \leq X < x_0 + \frac{1}{n}\right) = \\ &= \lim_{n \rightarrow \infty} \left(F\left(x_0 + \frac{1}{n}\right) - F(x_0)\right) = \lim_{n \rightarrow \infty} F\left(x_0 + \frac{1}{n}\right) - F(x_0) = \\ &= \lim_{x \rightarrow x_0^+} F(x) - F(x_0). \end{aligned}$$

Следува дека е точна следнава теорема:

Теорема 5.1.9. Нека F е функција на распределба на случајната променлива X .

Ако F е непрекината во x_0 , тогаш: $P(X = x_0) = 0$;

Ако x_0 е точка на прекин на F , тогаш: $P(X = x_0) > 0$.

Доказ. Ако F е непрекината во x_0 тогаш $\lim_{x \rightarrow x_0} F(x) = F(x_0)$

и:

$$\lim_{x \rightarrow x_0^+} F(x) = F(x_0) \Leftrightarrow \lim_{x \rightarrow x_0^+} F(x) - F(x_0) = 0 \Leftrightarrow P(X = x_0) = 0.$$

Ако F има прекин во x_0 , тогаш не важи $\lim_{x \rightarrow x_0} F(x) = F(x_0)$ и бидејќи F е непрекината од лево во x_0 , F не е непрекината од десно. Од:

$$P(X = x_0) = \lim_{x \rightarrow x_0^+} F(x) - F(x_0),$$

следува дека десната граница постои и:

$$\lim_{x \rightarrow x_0^+} F(x) - F(x_0) > 0 \text{ т.е. } P(X = x_0) > 0.$$

Забелешка 5.1.10. Во некои литератури случајна променлива над просторот на веројатност (Ω, F, P) е пресликувањето $X: \Omega \rightarrow \mathbb{R}$, такво што секое $x \in \mathbb{R}$, $\{E | X(E) \leq x\} \in F$, т. е.: е настан. Во таков случај:

$$(X \leq x) = \{E | X(E) \leq x\},$$

и функцијата на распределба се дефинира со равенството:

$$F(x) = P(X \leq x).$$

За вака дефинираната функција на распределба важат својствата: 1) и 2), додека 3) добива облик: 3') е непрекината од десно во секоја точка $x_0 \in \mathbb{R}$.

5.2. СЛУЧАЈНИ ПРОМЕНЛИВИ ОД ДИСКРЕТЕН ТИП

Нека (Ω, F, P) е простор на веројатност.

Дефиниција 5.2.1. Случајната променлива X е од **дискретен тип** (дискретна) ако множеството вредности на функцијата V_X , е конечно или преброиво.

Во ова поглавје случајните променливи ќе бидат од дискретен тип.

Множеството вредности односно рангот, може да се запише во видот: $V_X = \{x_1, x_2, \dots, x_n\}$ ако е конечно или: $V_X = \{x_1, x_2, \dots\}$ во случај кога е преброиво. Ако означиме со $p_i = P(X = x_i)$, тогаш во првиот случај важи $\sum_{i=1}^n p_i = 1$ додека во вториот: $\sum_{i=1}^{\infty} p_i = 1$. Притоа, за секое $x \in \mathbb{R} \setminus \{V_X\}$, $P(X = x) = 0$. На тој начин е определено пресликувањето од множеството реални броеви во интервалот $[0, 1]$ кое на секој реален број x му ја придружува веројатноста $P(X = x)$.

Дефиниција 5.2.2. Закон на распределба на случајната променлива X е пресликување $X: \mathbb{R} \rightarrow [0, 1]$ определено со:

$$X(x) = p(X = x) \text{ за секое } x \in \mathbb{R}.$$

Ако множеството вредности на функцијата е конечно и со мал број на елементи, тогаш законот на распределба на X се запишува во видот:

$$X: \begin{pmatrix} x_1 & x_2 & \dots & x_n \\ p_1 & p_2 & \dots & p_n \end{pmatrix},$$

каде што во првиот ред се запишуваат вредностите што ги прима случајната променлива X , подредени по растечки редослед, а во вториот ред соодветните веројатности. Ако се знае аналитичкиот

запис на n -тиот член и неговата веројатност. На овој начин може да се запише и законот на распределба на случајната променлива и кога множеството вредности има конечно или преброиво многу елементи.

Ако е даден законот на распределба, може еднозначно да се определи функцијата на распределба.

Нека множеството вредности на случајната променлива е конечно, т. е. е $\{x_1, x_2, \dots, x_n\}$ каде што $x_1 < x_2 < \dots < x_n$ и нека соодветните веројатности се: $p_i, i = 1, 2, \dots, n$. Тогаш за:

$$x \leq x_1, F(x) = P(X < x) = P(\emptyset) = 0;$$

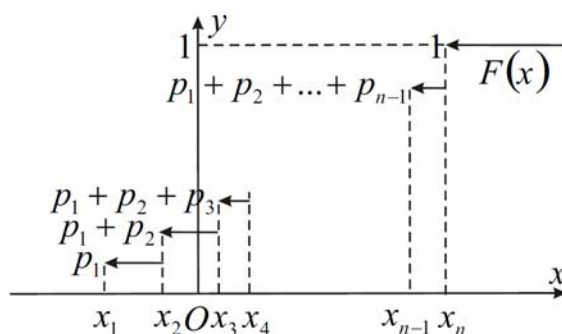
$$x_1 < x \leq x_2, F(x) = P(X < x) = P(x = x_1) = p_1;$$

$$x_2 < x \leq x_3, F(x) = P(X < x) = P(x \in \{x_1, x_2\}) = p_1 + p_2;$$

...

$$x_{n-1} < x \leq x_n, F(x) = P(X < x) = P(x \in \{x_1, x_2, \dots, x_{n-1}\}) = p_1 + p_2 + \dots + p_{n-1};$$

$$x > x_n, F(x) = P(X < x) = P(x \in \{x_1, x_2, \dots, x_n\}) = p_1 + p_2 + \dots + p_{n-1} + p_n.$$



Цртеж 5.2.2. Функција на распределба на случајна променлива од дискретен тип (конечен случај)

Следува дека функцијата на распределба е:

$$F(x) = \begin{cases} 0 & x \leq x_1 \\ p_1 & x_1 < x \leq x_2 \\ p_1 + p_2 & x_2 < x \leq x_3 \\ \dots & \dots \\ p_1 + p_2 + \dots + p_n, & x_{n-1} < x \leq x_n \\ 1 & x > x_n \end{cases}.$$

Функцијата на распределба може да се претстави и графички. Нејзиниот график има скалеста форма со скокови во точките x_i , $i = 1, 2, \dots, n$. Висината на скокот во точката x_i е p_i (види цртеж 5.2.2). Да забележиме дека подразбираме p_i да е позитивно, затоа го користиме поимот скок.

Аналогно се определува функцијата на распределба на преброивата случајна променлива.

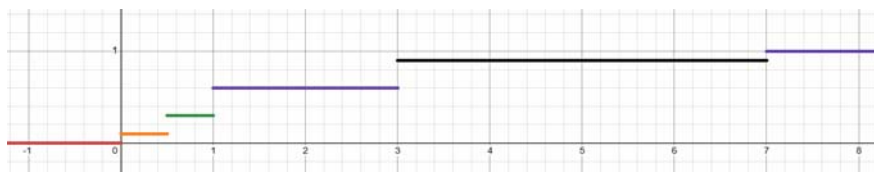
Задача 5.2.3. Случајната променлива X е зададена со законот на распределба:

$$X: \begin{pmatrix} 0 & 1 & 3 & 7 \\ 0,1 & 0,2 & 0,3 & 0,3 & 0,1 \end{pmatrix}.$$

Да се најде функцијата на распределба на случајната променлива X .

Решение. Функцијата на распределба на случајната променлива е:

$$F(x) = \begin{cases} 0; & x \leq 0 \\ 0,1; & 0 < x \leq 1 \\ 0,3; & 1 < x \leq 3 \\ 0,6; & 3 < x \leq 7 \\ 1; & x > 7 \end{cases}.$$



Цртеж 5.3.2. График на функцијата на распределба $F(x)$.

Аналогно се покажува и точноста на обратното тврдење: Функцијата на распределба на случајната променлива X еднозначно го определува законот на распределба на X .

Задача 5.2.4. Функцијата на распределба на случајната променлива X е дадена со:

$$F(x) = \begin{cases} 0; & x \leq -1 \\ 0,3; & -1 < x \leq 1 \\ 0,7; & 1 < x \leq 2 \\ 1; & x > 2 \end{cases}.$$

Да се најде законот на распределба на случајната променлива X .

Решение. Законот на распределба на случајната променлива X е:

$$X: \begin{pmatrix} -1 & 1 & 2 \\ 0,3 & 0,4 & 0,3 \end{pmatrix}.$$

5.3. СЛУЧАЈНИ ПРОМЕНЛИВИ ОД АПСОЛУТНО НЕПРЕКИНАТ ТИП

Нека множеството вредности на случајна променлива X е интервалот (a, b) , $a, b \in \overline{\mathbb{R}}$ и нека функцијата на распределба е непрекината на (a, b) . Тогаш за секое $x_0 \in (a, b)$,

$$P(X = x_0) = \lim_{x \rightarrow x_0^+} F(x) - F(x_0) = 0.$$

Значи поведението на случајната променлива не може да биде определено со помош на веројатностите кои ги прима во точно одредени вредности. Затоа треба да се најде друга карактеристика што ќе биде аналогна на законот на распределба на дискретната случајна променлива.

Да забележиме дека веројатноста $P(X = x_0) = 0$ не значи дека настанот $X = x_0$ никогаш нема да се реализира, туку дека веројатноста за неговата појава е многу мала. Тоа е разбирливо бидејќи случајната променлива може да прими која било вредност од интервалот (a, b) , а нив ги има бесконечно многу, па многу мали се шансите тоа да биде токму вредноста x_0 . Кај непрекинатите функции е невозможно на секој елемент од интервалот да му се додели позитивна веројатност, а да бидат исполнети останатите услови меѓу кои збирот на веројатностите од сите променливи да

биде единица. Затоа се зема $P(X = x) = 0$ во секоја точка $x \in (a, b)$ иако може $P(a < X < b) > 0$. На прв поглед ова е парадокс (точно тврдење што на прв поглед делува неточно), меѓутоа парадокси постојат во науката. На пример во геометријата ненулта отсечка има позитивна должина, а се состои од точки со должина нула.

Нека (Ω, F, P) е простор на веројатност.

Дефиниција 5.3.1. Нека $x, x + \Delta x \in (a, b)$. Ако постои границата (како конечен број):

$$p(x) \equiv \lim_{\Delta x \rightarrow 0} \frac{F(x + \Delta x) - F(x)}{\Delta x},$$

тогаш $p(x)$ се нарекува **густина на распределба на веројатностите на случајната променлива X** .

Густината на распределбата на веројатностите на случајната променлива X кратко се нарекува **густина на X** . Бидејќи:

$$p(x) \equiv F'(x) \text{ за } x \in (a, b),$$

ако $p(x)$ е интегрибилна функција, тогаш:

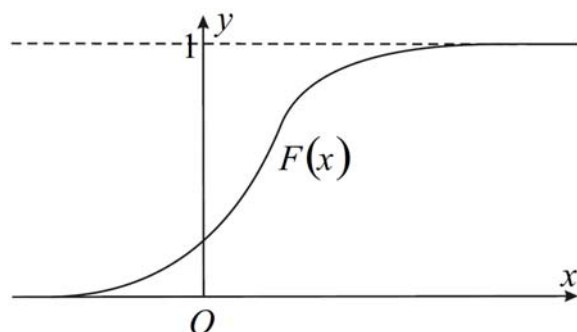
$$F(x) = \int_{-\infty}^x p(t) dt,$$

при што земаме дека $p(x) = 0$ за $x \leq a$ и $p(x) = 1$ за $x \geq b$.

Дефиниција 5.3.2. Ако постои интегрибилна функција $p(x)$ на (a, b) , таква што за функцијата на распределба на случајната променлива X , важи:

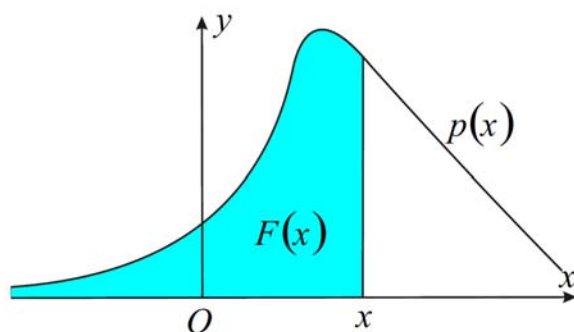
$$F(x) = \int_{-\infty}^x p(t) dt \text{ за секое } x \in (a, b),$$

тогаш велиме дека случајната променлива X е од апсолутно непрекинат тип.



Цртеж 5.3.1. Функција на распределба на случајна променлива од апсолутно непрекинат тип

Имајќи ја предвид геометриската интерпретација на определениот и несвојствениот интеграл, заклучуваме дека функцијата на распределба е плоштината на неограничениот криволиниски трапез во $y \leq x$ полурамнината, ограничен со x -оската, правата $y = x$ и графикот на функцијата $p(x)$ (види цртеж 5.3.2).



Цртеж 5.3.2. Густина на распределба

Следната теорема ги дава својствата на густината на случајната променлива од апсолутно непрекинат тип.

Теорема 5.3.3. Нека X е случајна променлива од апсолутно непрекинат тип со густина на распределба $p(x)$. Тогаш:

$$1) p(x) \geq 0 \text{ за секој } x \in \mathbb{R}; \quad 2) \int_{-\infty}^{+\infty} p(x) dx = 1;$$

$$3) P(a < X < b) = \int_a^b p(x) dx.$$

Доказ. 1) Бидејќи $p(x) \equiv F'(x)$ и $F'(x)$ е неопаѓачка функција т.е. $F'(x) \geq 0$ за секој $x \in \mathbb{R}$ следува дека

$$p(x) = F'(x) \geq 0 \text{ за секој } x \in \mathbb{R}.$$

2) Заради:

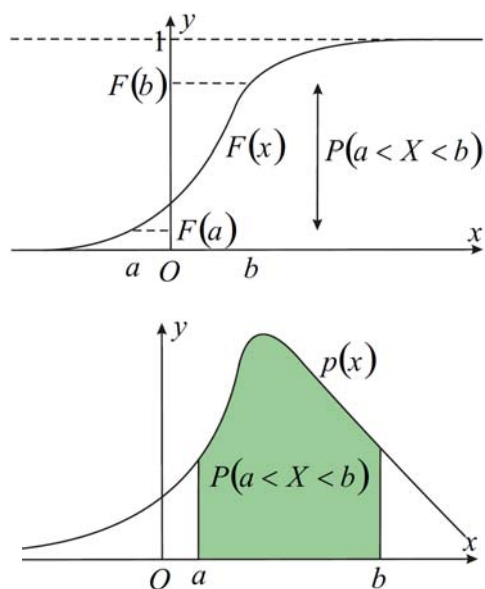
$$\lim_{x \rightarrow +\infty} F(x) = 1 \text{ и } \lim_{x \rightarrow +\infty} F(x) = \lim_{x \rightarrow +\infty} \int_{-\infty}^x p(t) dt = \int_{-\infty}^{+\infty} p(x) dx,$$

добиваме дека $\int_{-\infty}^{+\infty} p(x) dx = 1$.

3) Бидејќи случајната променлива X е од апсолутно непрекинат тип важи $P(X = a) = 0$, од каде што:

$$P(a < X < b) = P(a \leq X < b) = F(b) - F(a) = \int_{-\infty}^b p(x) dx - \int_{-\infty}^a p(x) dx = \int_a^b p(x) dx. \blacksquare$$

Густината овозможува да се пресмета веројатноста случајната променлива да се наоѓа во некој интервал (види цртеж 5.3.3 и 5.3.4.).



Цртеж 5.3.3-4.

Задача 5.3.4. Нека X е непрекината случајна променлива со густина:

$$p(x) = \begin{cases} C(2x - x^2), & x \in \left[0, \frac{5}{2}\right] \\ 0, & x \notin \left[0, \frac{5}{2}\right] \end{cases}.$$

Да се определи константата C , а потоа да се најде $P(X < 1)$.

Решение. а) Константата C ја определуваме од формулата за густина:

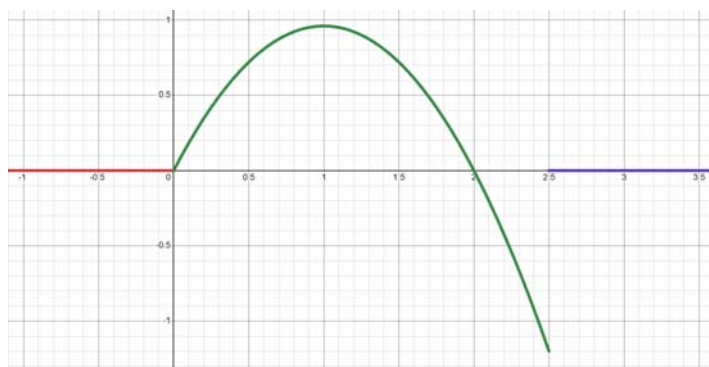
$$\int_{-\infty}^{+\infty} p(x) dx = 1 \Leftrightarrow \int_0^{\frac{5}{2}} C(2x - x^2) dx = 1.$$

Имаме:

$$\begin{aligned} C \int_0^{\frac{5}{2}} (2x - x^2) dx &= C \left(2 \frac{x^2}{2} - \frac{x^3}{3} \right) \bigg|_0^{\frac{5}{2}} = \\ C \left(\frac{25}{4} - \frac{1}{3} \frac{125}{8} \right) &= C \frac{150 - 125}{24} = C \frac{25}{24}, \end{aligned}$$

од каде што:

$$C \frac{25}{24} = 1 \Leftrightarrow C = \frac{24}{25}.$$



Цртеж 5.3.4. График на густината $p(x)$.

б) Според формулата за пресметување функција на распределба на непрекината случајна променлива, добиваме:

$$P(X < 1) = \int_{-\infty}^1 p(x) dx =$$

$$\frac{24}{25} \int_0^1 (2x - x^2) dx = \frac{24}{25} \left(x^2 - \frac{x^3}{3} \right) \Big|_0^1 = \frac{24}{25} \frac{2}{3} = \frac{16}{25}.$$

Задача 5.3.5. Непрекинатата случајна променлива X е зададена со нејзината густина на распределба:

$$F(x) = \begin{cases} 0, & x \leq 0 \\ C(x^2 - 3x), & x \in [0, 2] \\ 1, & x > 2 \end{cases}$$

Да се определат констатата и густината на распределба на случајната променлива X .

Решение. а) Имаме:

$$\lim_{x \rightarrow 0^+} F(x) = \lim_{x \rightarrow 0^+} C(x^2 - 3x) = 0, \quad F(0) = 0 \quad \text{и} \quad \lim_{x \rightarrow 0^-} F(x) = 0;$$

и

$$\lim_{x \rightarrow 2^-} F(x) = \lim_{x \rightarrow 2^-} C(x^2 - 3x) = C(4 - 6) = -2C,$$

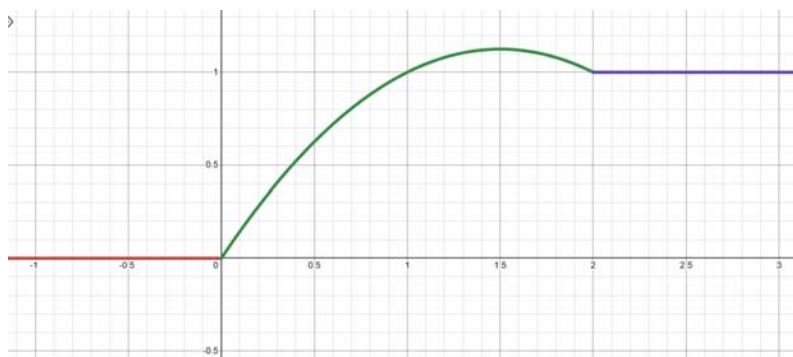
$$F(2) = -2C \quad \text{и} \quad \lim_{x \rightarrow 2^+} F(x) = 1.$$

Следува дека функцијата е непрекината во $x = 0$. За да е непрекината во $x = 2$ треба да важи:

$$-2C = 1 \Leftrightarrow C = -\frac{1}{2}.$$

Значи:

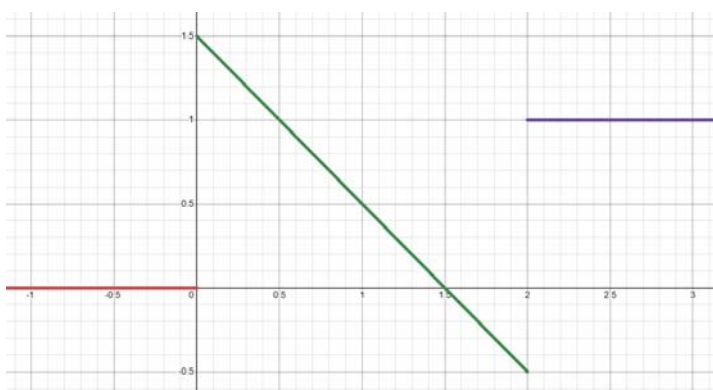
$$F(x) = \begin{cases} 0, & x \leq 0 \\ -\frac{1}{2}(x^2 - 3x), & x \in [0, 2] \\ 1, & x > 2 \end{cases}$$



Цртеж 5.3.5.1. График на функцијата $F(x)$.

Од равенството $p(x) = F'(x)$, добиваме дека:

$$p(x) = \begin{cases} 0, & x < 0 \\ -\frac{1}{2}(2x-3), & x \in (0, 2) \\ 1, & x > 2 \end{cases}$$



Цртеж 5.3.5.2. График на функцијата $p(x)$.

6. СРЕДНИ ВРЕДНОСТИ

Математичкото очекување и дисперзијата се бројни карактеристики на случајната променлива.

6.1. МАТЕМАТИЧКО ОЧЕКУВАЊЕ

Дефиниција 6.1.1. Математичко однесување на конечната случајна променлива X со множество вредности $V = \{x_1, x_2, \dots, x_n\}$, е бројот:

$$EX = x_1 p_1 + x_2 p_2 + \dots + x_n p_n = \sum_{i=1}^n x_i p_i,$$

каде што $p_i = P(X = x_i)$, $i = 1, 2, \dots, n$.

Математичкото очекување е мерка на тежинска средна вредност на случајната променлива.

Пример 6.1.2. Лотаријата пушта во промет 10000 лоза. Цената на еден лоз е 100 денари. Трошоците за изработка, продажба со плати за вработени и лиценци се 20 % од сумата, данокот на добивка е 10 % од добивките. Останатите 72 % од сумата се за награди. Добитните лозови се: 2000 по 100 денари, 1000 по 200 денари, 100 по 500 денари, 10 по 5000 и 1 носи 220000 денари.

Случајната променлива X : „Добивка од лозот“ е:

$$X : \begin{pmatrix} 0 & 100 & 200 & 500 & 5000 & 220000 \\ 0,6899 & 0,2 & 0,1 & 0,01 & 0,001 & 0,0001 \end{pmatrix}.$$

Математичкото очекување е:

$$\begin{aligned} EX &= x_1 p_1 + x_2 p_2 + \dots + x_6 p_6 = \\ 100 \cdot 0,2 + 200 \cdot 0,1 + 500 \cdot 0,01 + 5000 \cdot 0,001 + 220000 \cdot 0,0001 &= 72. \end{aligned}$$

Резултатот покажува дека ако се дадат 100 денари за лоз, очекуваната добивка е 72 денари, бидејќи 28 % од цената на лозот односно 28 денари се за трошоци на Лотаријата (20 % се трошци на лотаријата, додека 10 % од добивките кои се 80 % односно 8 % е данок на добивка. Ако играч со милион денари ги купи сите 10000

лоза, ќе има 720000 денари добивка, просечно по 72 денари од лоз. ■

Конечната случајна променлива може да се разгледа како случајна променлива од дискретен тип. Во продолжение ќе дадеме обопштување кое ќе важи за произволна случајна променлива од дискретен тип.

Дефиниција 6.1.3. Математичко однесување на дискретната случајна променлива X со множество вредности $V = \{x_1, x_2, \dots\}$ е бројот:

$$EX = x_1 p_1 + x_2 p_2 + \dots = \sum_{i=1}^{\infty} x_i p_i,$$

каде што $p_i = P(X = x_i)$, $i = 1, 2, \dots$.

Математичкото очекување на случајната променлива со конечно множество од n вредности се запишува со формулата од претходната дефиниција, така што $x_i = p_i = 0$ за сите $i > n$.

Пример 6.1.4. Математичкото очекување на случајната променлива X зададена со законот на распределба:

$$X: \begin{pmatrix} 1 & 2 & 3 & \dots & n & \dots \\ 0,9 & 0,045 & 0,003 & \dots & 9 \frac{10^{-n}}{n} & \dots \end{pmatrix},$$

е:

$$\begin{aligned} EX &= x_1 p_1 + x_2 p_2 + \dots = \\ &= 1 \cdot 0,9 + 2 \cdot 0,045 + 3 \cdot 0,003 + \dots + 9n \frac{10^{-n}}{n} + \dots = \\ &= 9(10^{-1} + 10^{-2} + \dots + 10^{-n} + \dots) = \\ &= 9 \cdot 10^{-1} (1 + 10^{-1} + 10^{-2} + \dots + 10^{-n-1} + \dots) = \\ &= 9 \cdot 10^{-1} \lim_{n \rightarrow \infty} (1 + 10^{-1} + 10^{-2} + \dots + 10^{-n-1} + \dots) = \\ &= 9 \frac{1}{10} \lim_{n \rightarrow \infty} \frac{1 - 10^{-n-1+1}}{1 - 10^{-1}} = \frac{9}{10} \lim_{n \rightarrow \infty} \frac{1 - \frac{1}{10^n}}{1 - \frac{1}{10}} = \frac{9}{10} \frac{1}{\frac{9}{10}} = 1. \end{aligned}$$

Дефиниција 6.1.5. Нека X е случајна променлива од апсолутно-непрекинат тип со густина на распределба $p(x)$. Математичко очекување на случајната променлива X е бројот:

$$EX = \int_{-\infty}^{+\infty} xp(x)dx,$$

под услов интегралот да е апсолутно конвергентен.

Пример 6.1.6. Најди го математичкото очекување EX на случајната променлива со густина на распределба:

$$p(x) = \begin{cases} x, & x \in (0, 1) \\ 1-x, & x \in [1, 2] \\ 0, & x \notin (0, 2) \end{cases}.$$

Колку е математичкото очекување EX^2 ?

Решение. Имаме:

$$\begin{aligned} EX &= \int_{-\infty}^{+\infty} xp(x)dx = \int_0^1 xxdx + \int_1^2 x(1-x)dx = \\ &= \int_0^1 x^2dx + \int_1^2 (x-x^2)dx = \left. \frac{x^3}{3} \right|_0^1 + \left(\left. \frac{x^2}{2} - \frac{x^3}{3} \right|_1^2 \right) = \\ &= \frac{1}{3} + \frac{1}{2}(4-1) - \frac{1}{3}(8-1) = \frac{1}{3} + \frac{3}{2} - \frac{7}{3} = -2 + \frac{3}{2} = -\frac{1}{2}; \end{aligned}$$

и

$$\begin{aligned} EX^2 &= \int_{-\infty}^{+\infty} x^2p(x)dx = \int_0^1 x^2xdx + \int_1^2 x^2(1-x)dx = \\ &= \int_0^1 x^3dx + \int_1^2 (x^2-x^3)dx = \left. \frac{x^4}{4} \right|_0^1 + \left(\left. \frac{x^3}{3} - \frac{x^4}{4} \right|_1^2 \right) = \\ &= \frac{1}{4} + \frac{1}{3}(8-1) - \frac{1}{4}(16-1) = \frac{1}{4} + \frac{7}{3} - \frac{15}{4} = -\frac{7}{2} + \frac{3}{2} = -2. \end{aligned}$$

Коментар. Забележуваме дека EX^2 (EX) се по договор скратени ознаки без загради за функции од случајните променливи $E(X^2)$, $(E(X))$, како што е случај и со други функции.

Теорема 6.1.7. (Својства на математичкото очекување). Нека X и Y се случајни променливи, и c е константа. Тогаш:

- 1) $Ec = c$;
- 2) $E(cX) = cEX$;
- 3) $E(X + Y) = EX + EY$;
- 4) $E(XY) = EX \cdot EY$,

каде што X и Y се независни случајни променливи;

- 5) $E|X| \leq E|X|$;
- 6) Ако $P(a \leq x \leq b) = 1$, тогаш $a \leq EX \leq b$;

при што во равенставата 2), 3) и 5) се подразбира постоење на EX и EY .

Доказ. Секоја случајна променлива од конечен тип може да се разгледа како случајна променлива од дискретен тип во која конечен број веројатности се ненулти, а сите други се нули. Затоа тврдењата што ќе ги покажеме за дискретен случај важат и за конечен случај.

1) Константата c може да се разгледа како дискретна случајна променлива која прима една вредност со веројатност 1, т. е.:

$$c: \begin{pmatrix} c \\ 1 \end{pmatrix}.$$

Тогаш:

$$Ec = c \cdot 1 = c.$$

2) Нека X е случајната променлива од дискретен тип:

$$X: \begin{pmatrix} x_1 & x_2 & \dots & x_n & \dots \\ p_1 & p_2 & \dots & p_n & \dots \end{pmatrix}.$$

Тогаш:

$$E(cX) = \sum_{i=1}^{\infty} cx_i p_i = c \sum_{i=1}^{\infty} x_i p_i = cEX.$$

3) Нека X и Y се дискретни случајни променливи такви што доменот на X е R_X , а на Y е R_Y . Тогаш случајната променлива

$X + Y$ има домен $R_{X+Y} = \{x_i + y_j \mid x_i \in R_X, y_j \in R_Y\}$ кој е исто така дискретно множество. Нека: $p_{ij} = P(X = x_i, Y = y_j)$. Важи:

$$p_i = P(X = x_i) = \sum_{y_j \in R_Y} p_{ij}, \quad x_i \in R_X \text{ и}$$

$$p_j = P(Y = y_j) = \sum_{x_i \in R_X} p_{ij}, \quad y_j \in R_Y.$$

Оттука:

$$E(X + Y) = EX + EY =$$

$$\sum_{x_i \in R_X} x_i p_i + \sum_{y_j \in R_Y} y_j p_j = \sum_{x_i \in R_X} x_i \sum_{y_j \in R_Y} p_{ij} + \sum_{y_j \in R_Y} y_j \sum_{x_i \in R_X} p_{ij} =$$

$$\sum_{x_i \in R_X} \sum_{y_j \in R_Y} (x_i + y_j) p_{ij} = E(X + Y).$$

4) Нека X и Y се независни случајни променливи од дискретен тип, зададени како во претходниот случај. Случајната променлива XY има домен $R_{XY} = \{x_i y_j \mid x_i \in R_X, y_j \in R_Y\}$. Нека $p_{ij} = P(X = x_i, Y = y_j)$. Бидејќи X и Y се независни случајни променливи, следува дека:

$$p_{ij} = P(X = x_i) P(Y = y_j) \text{ за сите } x_i \in R_X \text{ и } y_j \in R_Y.$$

Тогаш:

$$EX \cdot EY = \sum_{x_i \in R_X} x_i p_i \sum_{y_j \in R_Y} y_j p_j =$$

$$\sum_{x_i \in R_X} x_i y_i p_i p_j = \sum_{x_i \in R_X} x_i y_i p_{ij} = E(XY).$$

5) Имаме:

$$|EX| = \left| \sum_{x_i \in R_X} x_i p_i \right| \leq \sum_{x_i \in R_X} |x_i| p_i = E|X|.$$

6) Нека $P(a \leq X \leq b) = 1$. Тогаш:

$$EX = \sum_{x_i \in R_X} x_i p_i \geq \sum_{x_i \in R_X} a p_i = a \sum_{x_i \in R_X} p_i = a$$

и

$$EX = \sum_{x_i \in R_X} x_i p_i \leq \sum_{x_i \in R_X} b p_i = b \sum_{x_i \in R_X} p_i = b;$$

од каде што следува дека $a \leq EX \leq b$.

За случаите кога X е случајната променлива од апсолутно-непрекинат тип со густина $p(x)$ нема да дадеме докази. ■

Од 2) и 3) следува дека $E(cX + dY) = cEX + dEY$, каде што d е константа. Со помош на принципот на математичка индукција се заклучува дека за конечен број константи $c_1, c_2, \dots, c_n$ и математички очекувања $X_1, X_2, \dots, X_n$ важи:

$$E(c_1X_1 + c_2X_2 + \dots + c_nX_n) = c_1EX_1 + c_2EX_2 + \dots + c_nEX_n$$

и кога $X_1, X_2, \dots, X_n$ се независни случајни променливи:

$$E(X_1X_2 \dots X_n) = EX_1 \cdot EX_2 \cdot \dots \cdot EX_n.$$

6.2. ДИСПЕРЗИЈА

Дисперзијата е мерка за отстапување на вредностите на случајната променлива од нејзиното математичко очекување.

Дефиниција 6.2.1. (Дисперзија). Дисперзија (варијанса) на случајната променлива X , ознака DX , е бројот:

$$DX = E(X - EX)^2.$$

Теорема 6.2.2. (Критериум за пресметување дисперзија). Нека X е случајна променлива. Тогаш:

$$DX = EX^2 - (EX)^2.$$

Доказ. Со примена на дефиницијата за дисперзија и својствата на математичкото очекување, добиваме:

$$\begin{aligned} DX &= E(X - EX)^2 = E(X^2 - 2XEX + (EX)^2) = \\ &= EX^2 - 2EXEX + (EX)^2 = EX^2 - 2(EX)^2 + (EX)^2 = EX^2 - (EX)^2. \blacksquare \end{aligned}$$

Како кај договорот за означување степени на функции, изразот $(EX)^2$ ќе го запишуваме како E^2X . Значи $(EX)^2 \equiv E^2X$.

Критериумот ќе го користиме за пресметување на дисперзијата при решавањето задачи.

За случајни променливи од дискретен тип со множество вредности V и веројатности: $p_i = P(X = x_i)$, $i = 1, 2, \dots$, важи:

$$EX^2 = \sum_{i=1}^{\infty} x_i^2 p_i ; \text{ и } DX = \sum_{i=1}^{\infty} x_i^2 p_i - \left(\sum_{i=1}^{\infty} x_i p_i \right)^2 .$$

За случајни променливи од апсолутно-непрекинат тип со густина на распределба $p(x)$,

$$EX^2 = \int_{-\infty}^{+\infty} x^2 p(x) dx \text{ и } DX = \int_{-\infty}^{+\infty} x^2 p(x) dx - \left(\int_{-\infty}^{+\infty} xp(x) dx \right)^2 .$$

Теорема 6.2.3. (Својства на Дисперзија). Нека X и Y се случајни променливи, и c е константа. Тогаш:

- 1) $Dc = 0$,
- 2) $D(cX) = c^2 DX$,
- 3) $D(X + Y) = DX + DY$,

каде што X и Y се независни случајни променливи;

- 4) $D(X + c) = DX$,
- 5) $DX \geq 0$,

Доказ. 1) Константата c ја разгледаме како случајна променлива која прима вредност c со веројатност 1. Користејќи го критериумот за пресметување дисперзија $DX = EX^2 - E^2 X$, и својствата на математичкото очекување, добиваме:

$$Dc = Ec^2 - E^2 c = c^2 - c^2 = 0 .$$

2) Имаме:

$$\begin{aligned} D(cX) &= E(cX)^2 - E^2(cX) = E(c^2 X^2) - (cEX)^2 = \\ &= c^2 E(X^2) - c^2 (EX)^2 = c^2 (EX^2 - E^2 X) = c^2 DX . \end{aligned}$$

3) Бидејќи X и Y се независни случајни променливи т.е. $EXY = EXEY$, добиваме:

$$\begin{aligned} D(X + Y) &= E(X + Y)^2 - E^2(X + Y) = \\ &= E(X^2 + 2XY + Y^2) - (EX + EY)^2 = \end{aligned}$$

$$EX^2 + E2XY + EY^2 - ((EX)^2 + 2EXEY + (EY)^2) =$$

$$EX^2 + 2EXEY + EY^2 - E^2X - 2EXEY - E^2Y =$$

$$EX^2 - E^2X + EY^2 - E^2Y = DX + DY.$$

4) Имаме:

$$D(X + c) = DX + Dc = DX.$$

5) Од својството 6 за математичкото очекување следува дека ако за случајната променлива X важи $X \geq 0$, тогаш $EX \geq 0$.

Бидејќи секогаш $(X - EX)^2 \geq 0$ следува дека:

$$DX = E(X - EX)^2 \geq 0. \blacksquare$$

Во дискретен случај $DX = 0$ ако и само ако $P(X = c) = 1$ за некоја константа c .

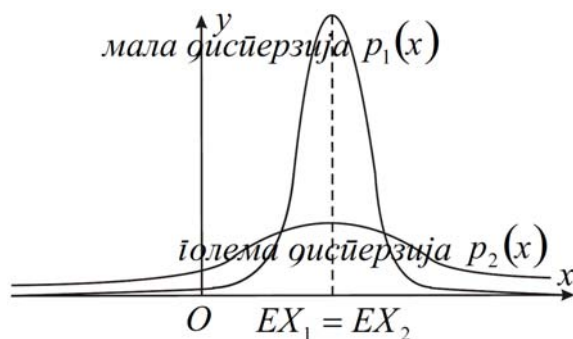
Имено нека $DX = E(X - EX)^2 = 0$. Бидејќи случајната променлива $(X - EX)^2$ е ненегативна, нејзиното математичко очекување е 0 само ако веројатноста $P((X - EX)^2 = 0) = 1$. Ако $P((X - EX)^2 = 0) < 1$, би постоела точка $(X - EX)^2 = x_{i_0} \neq 0$, за некое $i_0 \in \mathbb{N}$, со ненулта веројатност p_{i_0} . Но тогаш би добиле дека $E(X - EX)^2 \geq x_{i_0} p_{i_0} > 0$, што противречи на претпоставката.

Обратно ако $P(X = c) = 1$, тогаш $Ec = c$ и

$$Dc = Ec^2 - E^2c = c^2 - c^2 = 0.$$

Ако концентрацијата на вредностите на случајната променлива од дискретен тип околу математичкото очекување е голема, односно ако густината на случајната променлива од апсолутно непрекинат случај е голема, тогаш дисперзијата е мала. Во тој случај на вредностите кои лежат близу математичкото очекување им одговараат високи веројатности на случајната променлива, односно веројатноста за појава на случајната променлива на интервалите околу математичкото очекување е многу висока (види цртеж 6.2, график на $p_1(x)$ одговара на мала дисперзија). Ако вредностите на случајната променлива значајно се расејуваат околу математичкото очекување, тогаш дисперзијата е

голема (види цртеж 6.2, график на $p_2(x)$ одговара на голема дисперзија).



Цртеж 6.2.

6.3. ЗАДАЧИ СО СЛУЧАЈНИ ПРОМЕНЛИВИ

Задача 4.6.1. Да се определи функцијата на распределба, математичкото очекување и дисперзијата на случајната променлива:

$$X: \begin{pmatrix} -2 & -1 & 0 & 1 & 2 \\ 0,1 & 0,2 & 0,3 & 0,3 & 0,1 \end{pmatrix}.$$

Решение. Функцијата на распределба е:

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

Математичкото очекување е:

$$\begin{aligned} EX &= x_1 p_1 + x_2 p_2 + x_3 p_3 + x_4 p_4 + x_5 p_5 = \\ &= -2 \cdot 0,1 - 1 \cdot 0,2 + 0 + 1 \cdot 0,3 + 2 \cdot 0,1 = 0,1. \end{aligned}$$

За да ја најдеме дисперзијата прво го пресметуваме математичкото очекување:

$$EX^2 = x_1^2 p_1 + x_2^2 p_2 + x_3^2 p_3 + x_4^2 p_4 + x_5^2 p_5 =$$

$$4 \cdot 0,1 + 1 \cdot 0,2 + 0 + 1 \cdot 0,3 + 4 \cdot 0,1 = 1,3,$$

од каде што:

$$DX = EX^2 - (EX)^2 = 1,3 - 0,01 = 1,29.$$

Задача 4.6.2. Една монета е фрлена 7 пати во воздух. Да се определи функцијата на распределба, математичкото очекување и дисперзијата на случајната променлива: број на паѓања на глава.

Решение. Нека со X ја означиме случајната променлива: број на паѓања на глава. Случајната променлива е распределена според биомната распределба. Имаме $p = \frac{1}{2}$, $q = \frac{1}{2}$ и $n = 7$.

Веројатностите во биомната распределба се:

$$p_0 = P(X = 0) = C_7^0 p^0 q^7 = \binom{7}{0} \left(\frac{1}{2}\right)^0 \left(\frac{1}{2}\right)^7 = \frac{1}{128};$$

$$p_1 = P(X = 1) = C_7^1 p^1 q^6 = \binom{7}{1} \left(\frac{1}{2}\right)^1 \left(\frac{1}{2}\right)^6 = \frac{7}{128};$$

$$p_2 = P(X = 2) = C_7^2 p^2 q^5 = \binom{7}{2} \left(\frac{1}{2}\right)^2 \left(\frac{1}{2}\right)^5 = \frac{21}{128};$$

$$p_3 = P(X = 3) = C_7^3 p^3 q^4 = \binom{7}{3} \left(\frac{1}{2}\right)^3 \left(\frac{1}{2}\right)^4 = \frac{7 \cdot 6 \cdot 5}{2 \cdot 3} \frac{1}{128} = \frac{35}{128};$$

$$p_4 = P(X = 4) = P(X = 3) = \frac{35}{128}; \quad p_5 = P(X = 5) = P(X = 2) = \frac{21}{128};$$

$$p_6 = P(X = 6) = P(X = 1) = \frac{7}{128} \quad \text{и} \quad p_7 = P(X = 7) = P(X = 0) = \frac{1}{128}.$$

За проверка може да го искористиме својството збирот од сите веројантости да е 1. Навистина:

$$\frac{1}{128} + \frac{7}{128} + \frac{21}{128} + \frac{35}{128} + \frac{35}{128} + \frac{21}{128} + \frac{7}{128} + \frac{1}{128} = \frac{128}{128} = 1.$$

Го запишуваме законот на распределба:

$$X: \begin{pmatrix} 0 & 1 & 2 & 3 & 4 & 5 & 6 & 7 \\ \frac{1}{128} & \frac{7}{128} & \frac{21}{128} & \frac{35}{128} & \frac{35}{128} & \frac{21}{128} & \frac{7}{128} & \frac{1}{128} \end{pmatrix}.$$

Математичкото очекување е:

$$EX = x_0 p_0 + x_1 p_1 + x_2 p_2 + x_3 p_3 + x_4 p_4 + x_5 p_5 + x_6 p_6 + x_7 p_7 =$$

$$\frac{0 \cdot 1 + 1 \cdot 7 + 2 \cdot 21 + 3 \cdot 35 + 4 \cdot 35 + 5 \cdot 21 + 6 \cdot 7 + 7 \cdot 1}{128} =$$

$$\frac{7(1+7+21+35)}{128} = \frac{7 \cdot 64}{128} = \frac{7}{2}.$$

Математичкото очекување од случајната променлива X^2 е:

$$EX^2 = x_0^2 p_0 + x_1^2 p_1 + x_2^2 p_2 + x_3^2 p_3 + x_4^2 p_4 + x_5^2 p_5 + x_6^2 p_6 + x_7^2 p_7 =$$

$$\frac{0 \cdot 1 + 1 \cdot 7 + 4 \cdot 21 + 9 \cdot 35 + 16 \cdot 35 + 25 \cdot 21 + 36 \cdot 7 + 49 \cdot 1}{128} =$$

$$\frac{7(7+37+87+125)}{128} = \frac{7 \cdot 256}{128} = 14,$$

од каде што дисперзијата е:

$$DX = EX^2 - (EX)^2 = 14 - \frac{49}{4} = \frac{56-49}{4} = \frac{7}{4}.$$

Задача 4.6.3. Законите на распределба на независните случајни променливи X и Y , соодветно се:

$$X: \begin{pmatrix} 0 & 1 & 2 \\ 0,25 & 0,5 & 0,25 \end{pmatrix} \text{ и } Y: \begin{pmatrix} 0 & 1 & 2 & 3 \\ 0,4 & 0,3 & 0,2 & 0,1 \end{pmatrix}.$$

Да се определи законот на распределба на случајната променлива $X+Y$, а потоа да се пресметаат математичките очекувања и дисперзиите на секоја од променливите: X , Y и $X+Y$.

Решение. Прво го определуваме законот на распределба на случајната променлива $X+Y$. За полесно пресметување користиме помошни табели.

	X/Y	0	1	2	3		X/Y	0	1	2	3
$X+Y:$	0	0	1	2	3	и $P:$	0	0,1	0,075	0,05	0,025
	1	1	2	3	4		1	0,2	0,15	0,1	0,05
	2	2	3	4	5		2	0,1	0,075	0,05	0,025

Следува дека:

$$P(X+Y=0) = 0,1, \quad P(X+Y=1) = 0,2 + 0,075 = 0,275,$$

$$P(X + Y = 2) = 0,1 + 0,15 + 0,05 = 0,3,$$

$$P(X + Y = 3) = 0,075 + 0,1 + 0,025 = 0,2,$$

$$P(X + Y = 4) = 0,005 + 0,05 = 0,1 \text{ и } P(X + Y = 5) = 0,025;$$

од каде што законот на распределба е:

$$X + Y: \begin{pmatrix} 0 & 1 & 2 & 3 & 4 & 5 \\ 0,1 & 0,275 & 0,3 & 0,2 & 0,1 & 0,025 \end{pmatrix}.$$

За математичките очекувања и дисперзии добиваме:

$$EX = 0,5 + 2 \cdot 0,25 = 1,$$

$$EX^2 = 0,5 + 4 \cdot 0,25 = 1,5 \text{ и } DX = 1,5 - 1^2 = 0,5;$$

$$EY = 0,3 + 2 \cdot 0,2 + 3 \cdot 0,1 = 1,$$

$$EY^2 = 0,3 + 4 \cdot 0,2 + 9 \cdot 0,1 = 2 \text{ и } DY = 2 - 1^2 = 1;$$

и

$$E(X + Y) = 0,275 + 2 \cdot 0,3 + 3 \cdot 0,2 + 4 \cdot 0,1 + 5 \cdot 0,025 = 2,$$

$$E(X + Y)^2 = 0,275 + 4 \cdot 0,3 + 9 \cdot 0,2 + 16 \cdot 0,1 + 25 \cdot 0,025 = 0,275 + 1,2 + 1,8 + 1,6 + 0,625 = 5,5, \text{ и}$$

$$D(X + Y) = 5,5 - 2^2 = 1,5.$$

За проверка, математичкото очекување и дисперзијата на случајната променлива $X + Y$ може да се пресмета според својството:

$$E(X + Y) = EX + EY = 1 + 1 = 2 \text{ и}$$

$$D(X + Y) = DX + DY = 0,5 + 1 = 1,5.$$

Задача 4.6.3. Законите на распределба на независните случајни променливи X и Y соодветно се:

$$X: \begin{pmatrix} 0 & 1 & 2 \\ 0,25 & 0,5 & 0,25 \end{pmatrix} \text{ и } Y: \begin{pmatrix} 0 & 1 & 2 & 3 \\ 0,4 & 0,3 & 0,2 & 0,1 \end{pmatrix}.$$

Да се определи законот на распределба на случајната променлива XY , а потоа да се пресмета нејзиното математичко очекување и дисперзија.

Решение. Имаме:

6. Средни вредности

X/Y		0	1	2	3	X/Y		0	1	2	3
$XY:$	0	0	0	0	0	и $P:$	0	0,1	0,075	0,05	0,025
	1	0	1	2	3		1	0,2	0,15	0,1	0,05
	2	0	2	4	6		2	0,1	0,075	0,05	0,025

$$P(XY=0)=0,1+0,075+0,05+0,025+0,2+0,1=0,55;$$

$$P(XY=1)=0,15; \quad P(XY=2)=0,075+0,1=0,175;$$

$$P(XY=3)=0,05, \quad P(XY=4)=0,05 \quad \text{и} \quad P(XY=6)=0,025.$$

Оттука, законот на распределба е:

$$XY: \begin{pmatrix} 0 & 1 & 2 & 3 & 4 & 6 \\ 0,55 & 0,15 & 0,175 & 0,05 & 0,05 & 0,025 \end{pmatrix}.$$

За математичкото очекување и дисперзијата добиваме:

$$E(XY)=0,15+2 \cdot 0,175+3 \cdot 0,05+4 \cdot 0,05+6 \cdot 0,025= \\ 0,15+0,35+0,15+0,2+0,15=1,$$

$$E(XY)^2=0,15+4 \cdot 0,175+9 \cdot 0,05+16 \cdot 0,05+36 \cdot 0,025= \\ 0,15+0,7+0,45+0,8+0,9=3, \text{ и}$$

$$D(XY)=3-1^2=2.$$

Задача 4.6.4. Еден работник опслужува 4 машини. Веројатноста дека во еден час работа не е потребна интервенција на работникот при првата машина, изнесува 0,9. Соодветните веројатности за останатите три машини изнесуваат 0,8; 0,75 и 0,7. Да се пресмета математичкото очекување и дисперзијата на случајната променлива: Број на машини што во текот на еден час не побаруваат интервенција од работникот.

Решение. Да ги означиме со:

$$p_1=0,9, \quad p_2=0,8, \quad p_3=0,75 \quad \text{и} \quad p_4=0,7;$$

веројатностите на настаните од условот во задачата. Тогаш:

$$\bar{p}_1=0,1, \quad \bar{p}_2=0,2, \quad \bar{p}_3=0,25 \quad \text{и} \quad \bar{p}_4=0,3;$$

се веројатностите на соодветните комплементарни настани.

Нека со X ја означиме случајната променлива: Број на машини за кои во еден час не е потребна интервенција на работникот. Тогаш, X прима вредности од 0 до 4 со веројатности:

$$P(X=0) = \bar{p}_1 \bar{p}_2 \bar{p}_3 \bar{p}_4 = 0,1 \cdot 0,2 \cdot 0,25 \cdot 0,3 = 0,0015;$$

$$\begin{aligned} P(X=1) &= p_1 \bar{p}_2 \bar{p}_3 \bar{p}_4 + \bar{p}_1 p_2 \bar{p}_3 \bar{p}_4 + \bar{p}_1 \bar{p}_2 p_3 \bar{p}_4 + \bar{p}_1 \bar{p}_2 \bar{p}_3 p_4 = \\ &= 0,9 \cdot 0,2 \cdot 0,25 \cdot 0,3 + 0,1 \cdot 0,8 \cdot 0,25 \cdot 0,3 + 0,1 \cdot 0,2 \cdot 0,75 \cdot 0,3 + \\ &0,1 \cdot 0,2 \cdot 0,25 \cdot 0,7 = 0,0135 + 0,006 + 0,0045 + 0,0035 = 0,0275. \end{aligned}$$

$$P(X=2) =$$

$$\begin{aligned} &p_1 p_2 \bar{p}_3 \bar{p}_4 + p_1 \bar{p}_2 p_3 \bar{p}_4 + p_1 \bar{p}_2 \bar{p}_3 p_4 + \bar{p}_1 p_2 p_3 \bar{p}_4 + \bar{p}_1 p_2 \bar{p}_3 p_4 + \bar{p}_1 \bar{p}_2 p_3 p_4 = \\ &= 0,9 \cdot 0,8 \cdot 0,25 \cdot 0,3 + 0,9 \cdot 0,2 \cdot 0,75 \cdot 0,3 + 0,9 \cdot 0,2 \cdot 0,25 \cdot 0,7 + \\ &0,1 \cdot 0,8 \cdot 0,75 \cdot 0,3 + 0,1 \cdot 0,8 \cdot 0,25 \cdot 0,7 + 0,1 \cdot 0,2 \cdot 0,75 \cdot 0,7 = \\ &= 0,054 + 0,0405 + 0,0315 + 0,018 + 0,014 + 0,0105 = 0,1685; \end{aligned}$$

$$\begin{aligned} P(X=3) &= p_1 p_2 p_3 \bar{p}_4 + p_1 p_2 \bar{p}_3 p_4 + p_1 \bar{p}_2 p_3 p_4 + \bar{p}_1 p_2 p_3 p_4 = \\ &0,9 \cdot 0,8 \cdot 0,75 \cdot 0,3 + 0,9 \cdot 0,8 \cdot 0,25 \cdot 0,7 + 0,9 \cdot 0,2 \cdot 0,75 \cdot 0,7 + \\ &0,1 \cdot 0,8 \cdot 0,75 \cdot 0,7 = 0,162 + 0,126 + 0,0945 + 0,042 = 0,4245; \end{aligned}$$

$$P(X=4) = 0,9 \cdot 0,8 \cdot 0,75 \cdot 0,7 = 0,378.$$

Ја составуваме табелата што го опишува законот на распределба:

$$X: \begin{pmatrix} 0 & 1 & 2 & 3 & 4 \\ 0,0015 & 0,0275 & 0,1685 & 0,4245 & 0,378 \end{pmatrix}.$$

На крај ја спроведуваме постапката за определување на математичкото очекување и дисперзијата.

$$\begin{aligned} EX &= 0,0275 + 2 \cdot 0,1685 + 3 \cdot 0,4245 + 4 \cdot 0,378 = \\ &0,0275 + 0,337 + 1,2735 + 1,512 = 3,15, \end{aligned}$$

$$\begin{aligned} EX^2 &= 0,0275 + 4 \cdot 0,1685 + 9 \cdot 0,4245 + 16 \cdot 0,378 = \\ &0,0275 + 0,674 + 3,8205 + 6,048 = 10,57, \text{ и} \end{aligned}$$

$$DX = EX^2 - (EX)^2 = 10,57 - 3,15^2 = 10,57 - 9,9225 = 0,6475.$$

Коментар. Да забележиме дека кога се пресметува рачно полесно се пресметува ако извадиме 10^{-4} пред заграда:

$$P(X=0) = 10^{-4} \cdot 1 \cdot 2 \cdot 2,5 \cdot 3 = 10^{-4} \cdot 15 = 0,0015;$$

$$P(X=1) = 10^{-4} (9 \cdot 2 \cdot 2,5 \cdot 3 + 1 \cdot 8 \cdot 2,5 \cdot 3 + 1 \cdot 2 \cdot 7,5 \cdot 3 + 0,1 \cdot 2 \cdot 2,5 \cdot 7) =$$

$$10^{-4}(135 + 60 + 45 + 35) = 10^{-4} \cdot 275 = 0,0275.$$

$$P(X = 2) = 10^{-4}(9 \cdot 8 \cdot 2,5 \cdot 3 + 9 \cdot 2 \cdot 7,5 \cdot 3 + 9 \cdot 2 \cdot 2,5 \cdot 7 + 1 \cdot 8 \cdot 7,5 \cdot 3 + 1 \cdot 8 \cdot 2,5 \cdot 7 + 1 \cdot 2 \cdot 7,5 \cdot 7) = 10^{-4}(540 + 405 + 315 + 180 + 140 + 105) = 10^{-4} \cdot 1685 = 0,1685;$$

$$P(X = 3) = 10^{-4}(9 \cdot 8 \cdot 7,5 \cdot 3 + 9 \cdot 8 \cdot 2,5 \cdot 7 + 9 \cdot 2 \cdot 7,5 \cdot 7 + 1 \cdot 8 \cdot 7,5 \cdot 7) = 10^{-4}(1620 + 1260 + 945 + 420) = 10^{-4} \cdot 4245 = 0,4245; \text{ и}$$

$$P(X = 4) = 9 \cdot 8 \cdot 7,5 \cdot 7 = 10^{-4} \cdot 3780 = 0,378.$$

Вршине проверка на точноста на резултатите:

$$0,0015 + 0,0275 + 0,1685 + 0,4245 + 0,3780 = 1.$$

7. ПРИМЕРИ НА РАСПРЕДЕЛБИ

7.1. РАСПРЕДЕЛБИ ОД ДИСКРЕТЕН ТИП

7.1.1. Индикатор на настан A

Нека е даден настанот A . Случајната променлива I_A која означува број на појавувања на настанот A во експериментот се нарекува **индикатор на настанот A** .

Значи случајната променлива има вредност 0 ако не се појави настанот A и 1 ако се појави A .

Нека $P(A) = p$ и $q = 1 - p$. Тогаш законот на распределба е:

$$I_A: \begin{pmatrix} 0 & 1 \\ q & p \end{pmatrix}.$$

Од законот на распределба го гледаме множеството вредности на случајната променлива $I_A: V = \{0, 1\}$ и веројатноста на настанот $\bar{A}$: $P(\bar{A}) = 0$.

Математичкото очекување на индикаторот на настанот A е:

$$EI_A = 0 \cdot q + p = p.$$

За да ја определиме дисперзијата, прво го пресметуваме:

$$EI_A^2 = 0^2 \cdot q + 1^2 \cdot p = p$$

од каде што:

$$DI_A = EI_A^2 - (EI_A)^2 = p - p^2 = p(1 - p) = pq.$$

7.1.2. Биномна распределба

Нека случајната променлива X означува број на појавувања на настанот A во Бернулиева шема со n експерименти, каде што:

$$P(A) = p \text{ и } q = 1 - p.$$

Тогаш веројатноста настанот да се појави k пати е:

$$p_k = P(X = k) = P_n(k) = \binom{n}{k} p^k q^{n-k}, \quad k = 0, 1, \dots, n.$$

Во овој случај велиме дека случајната променлива има **биномна распределба**. Законот на распределба е:

$$X: \begin{pmatrix} 0 & 1 & \dots & k & \dots & n \\ q^n & npq & \dots & \binom{n}{k} p^k q^{n-k} & \dots & p^n \end{pmatrix}.$$

Случајната променлива може да се претстави како унија од n случајни променливи $X_i = I_A$, $i = 1, 2, \dots, n$ т. е.:

$$X = X_1 + X_2 + \dots + X_n,$$

кои се независни случајни променливи, бидејќи експериментите што ги определуваат се независни. Важи:

$$EX_i = EI_A = p \text{ и } DX_i = DI_A = pq, \quad i = 1, 2, \dots, n.$$

Затоа математичкото очекување на X е:

$$EX = E(X_1 + X_2 + \dots + X_n) = EX_1 + EX_2 + \dots + EX_n = np,$$

додека дисперзијата е:

$$DX = D(X_1 + X_2 + \dots + X_n) = DX_1 + DX_2 + \dots + DX_n = npq.$$

7.1.3. Рамномерна распределба

Случајната променлива која прима вредности од конечно множество:

$$V = \{x_1, x_2, \dots, x_n\},$$

со еднакви веројатности:

$$P(X = x_k) = \frac{1}{n},$$

велиме дека има **рамномерна распределба**.

Рамномерната распределба е определена со законот на распределба:

$$X: \begin{pmatrix} x_1 & x_2 & \dots & x_n \\ \frac{1}{n} & \frac{1}{n} & \dots & \frac{1}{n} \end{pmatrix}.$$

Математичкото очекување е:

$$EX = \frac{1}{n}x_1 + \frac{1}{n}x_2 + \dots + \frac{1}{n}x_n = \frac{x_1 + x_2 + \dots + x_n}{n} = \bar{x}_n,$$

додека дисперзијата е:

$$DX = E(X - EX)^2 = E(X - \bar{x}_n)^2 = \sum_{i=1}^n (x_i - \bar{x}_n)^2 = s_n^2.$$

Следува дека математичкото очекување и дисперзија на X се аритметичка средина и дисперзија на елементите од множеството вредности.

7.1.4. Геометриска распределба

Нека се изведува серија од независни експерименти со еднаква веројатност на појава на настанот A , $P(A) = p$. Серијата се изведува до експериментот во кој ќе се појави настанот A .

Нека A_i е настанот: Во i -тиот експеримент се појавил настанот A , $i = 1, 2, \dots$.

Настанот $(X = i)$ ќе се појави ако во првите $i-1$ експерименти не се појавил настанот A , а во i -тиот се појавил. Велме дека случајната променлива X има **геометриска распределба** со параметар p . Тогаш:

$$p_k = P(X = k) = P(\bar{A}_1 \bar{A}_2 \dots \bar{A}_{k-1} A_k) = P(\bar{A}_1)P(\bar{A}_2) \dots P(\bar{A}_{k-1})P(A_k) = q^{k-1}p.$$

Притоа, користејќи ја формулата за геометриска прогресија, добиваме навистина:

$$\sum_{k=1}^{\infty} q^{k-1}p = p \lim_{n \rightarrow \infty} \sum_{k=1}^n q^{k-1} = p \lim_{n \rightarrow \infty} \frac{1 - q^n}{1 - q} = 1 - \lim_{n \rightarrow \infty} q^n = 1.$$

Законот на распределба е:

$$X: \begin{pmatrix} 1 & 2 & \dots & n & \dots \\ p & qp & \dots & q^{n-1}p & \dots \end{pmatrix}.$$

Користејќи својства на редови се покажува дека математичкото очекување и дисперзијата на X се:

$$EX = \frac{1}{p} \text{ и } DX = \frac{q}{p^2}.$$

7.1.5. Пуасонова распределба

Случајната променлива X има **Пуасонова распределба** ако има закон на распределба:

$$X: \begin{pmatrix} 0 & 1 & \dots & n & \dots \\ e^{-\lambda} & \lambda e^{-\lambda} & \dots & \frac{\lambda^n}{n!} e^{-\lambda} & \dots \end{pmatrix}.$$

Значи, множеството вредности на случајна променлива со Пуасонова распределба е $V = \mathbb{N}_0$, а веројатностите се

$p_k = P(X = k) = \frac{\lambda^k}{k!} e^{-\lambda}$. Јасно од формулата за сумата на редот на e^x , следува дека:

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

Математичкото очекување е:

$$\begin{aligned} EX &= \lambda e^{-\lambda} + 2\lambda^2 \frac{e^{-\lambda}}{2!} + \dots + n \frac{\lambda^n}{n!} e^{-\lambda} = \\ &= \lambda e^{-\lambda} \left(1 + \lambda + \dots + \frac{\lambda^{n-1}}{(n-1)!} + \frac{\lambda^n}{n!} \dots \right) = \lambda e^{-\lambda} e^{\lambda} = \lambda. \end{aligned}$$

Го пресметуваме математичкото очекување на X^2 :

$$\begin{aligned} EX^2 &= \lambda e^{-\lambda} + 2^2 \lambda^2 \frac{e^{-\lambda}}{2!} + \dots + n^2 \frac{\lambda^n}{n!} e^{-\lambda} + \dots = \\ &= \lambda e^{-\lambda} \left(1 + 2\lambda + \dots + n \frac{\lambda^{n-1}}{(n-1)!} + (n+1) \frac{\lambda^n}{n!} + \dots \right) = \end{aligned}$$

$$\begin{aligned} \lambda e^{-\lambda} \left(\left(1 + \lambda + \dots + \frac{\lambda^n}{n!} + \dots \right) + \lambda + \dots + n \frac{\lambda^n}{n!} + \dots \right) &= \\ \lambda e^{-\lambda} \left(e^\lambda + \lambda \left(1 + \lambda + \dots + \frac{\lambda^{n-1}}{(n-1)!} + \frac{\lambda^n}{n!} + \dots \right) \right) &= \\ \lambda e^{-\lambda} (e^\lambda + \lambda e^\lambda) &= \lambda + \lambda^2. \end{aligned}$$

Притоа, групирањето во редот е можно бидејќи е ред со позитивни членови. Оттука:

$$DX = EX^2 - (EX)^2 = \lambda + \lambda^2 - \lambda^2 = \lambda.$$

Значи кај случајната променлива што има Пуасонова распределба, математичкото очекување и дисперзијата се еднакви.

7.2. РАСПРЕДЕЛБИ ОД АПСОЛУТНО-НЕПРЕКИНАТ ТИП

7.2.1. Рамномерна распределба

Случајната променлива X има **рамномерна распределба** на интервалот (a, b) , ако нејзината густина е константна на (a, b) и нула на $\mathbb{R} \setminus (a, b)$.

Следува дека: $p(x) = \begin{cases} C, & x \in (a, b) \\ 0, & x \notin (a, b) \end{cases}$. Константата C се

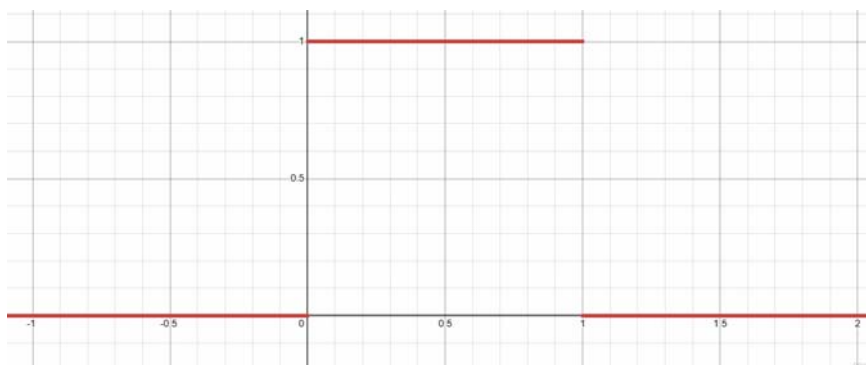
определува од условот:

$$\int_{-\infty}^{\infty} p(x) dx = 1 \Leftrightarrow \int_a^b C dx = 1 \Leftrightarrow C x \Big|_a^b = 1 \Leftrightarrow C(b-a) = 1 \Leftrightarrow C = \frac{1}{b-a}.$$

Значи:

$$p(x) = \begin{cases} \frac{1}{b-a}, & x \in (a, b) \\ 0, & x \notin (a, b) \end{cases}.$$

7. Примери на распределби



Цртеж 7.2.1.1. Густина на распределба на X за $b=1$ и $a=0$.



Цртеж 7.2.1.2. Функција на распределба на X за $b=1$ и $a=0$.

Ќе ја определиме функцијата на распределба на случајната променлива X . За:

$$x \in (-\infty, a], F(x) = \int_{-\infty}^x p(t) dt = \int_{-\infty}^x 0 dt = 0;$$

$$x \in (a, b], F(x) = \int_{-\infty}^x p(t) dt = \int_{-\infty}^a 0 dt + \int_a^x \frac{1}{b-a} dt = \frac{1}{b-a} x \Big|_a^x = \frac{x-a}{b-a};$$

$$x > b, F(x) = \int_{-\infty}^x p(t) dt = \int_a^b p(t) dt + \int_b^x 0 dt = \frac{x-a}{b-a} \Big|_a^b = 1.$$

Значи:

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

Математичкото очекување е:

$$EX = \int_{-\infty}^{+\infty} xp(x)dx = \int_a^b x \frac{1}{b-a} dx = \frac{1}{b-a} \frac{x^2}{2} \Big|_a^b =$$

$$\frac{b^2 - a^2}{2(b-a)} = \frac{(b-a)(b+a)}{2(b-a)} = \frac{a+b}{2}.$$

Математичкото очекување на EX^2 е:

$$EX^2 = \int_a^b x^2 \frac{1}{b-a} dx = \frac{1}{b-a} \frac{x^3}{3} \Big|_a^b =$$

$$\frac{b^3 - a^3}{3(b-a)} = \frac{(b-a)(b^2 + ab + a^2)}{3(b-a)} = \frac{a^2 + ab + b^2}{3},$$

од каде што дисперзијата е:

$$DX = EX^2 - (EX)^2 = \frac{a^2 + ab + b^2}{3} - \frac{(a+b)^2}{4} =$$

$$\frac{4a^2 + 4ab + 4b^2 - 3a^2 - 6ab - 3b^2}{12} = \frac{a^2 - 2ab + b^2}{12} = \frac{(a-b)^2}{12}.$$

7.2.2. Експоненцијална распределба

Случајната променлива X има **експоненцијална** распределба со параметар λ , ако нејзината густина е:

$$p(x) = \begin{cases} 0, & x < 0 \\ \lambda e^{-\lambda x}, & x \geq 0 \end{cases}.$$

Тогаш:

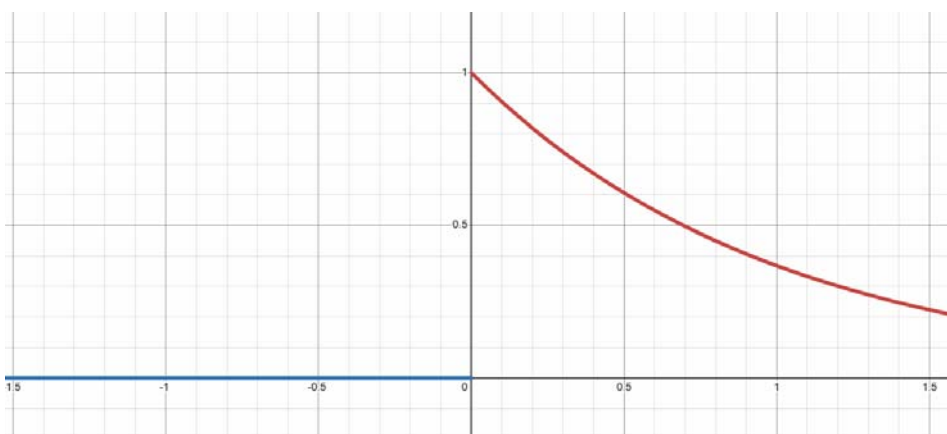
$$\text{за } x \leq 0, F(x) = 0;$$

$$\text{за } x > 0, F(x) = \int_0^x \lambda e^{-\lambda t} dt = \lambda \frac{e^{-\lambda t}}{-\lambda} \Big|_0^x = e^{-\lambda t} \Big|_x^0 = 1 - e^{-\lambda x}.$$

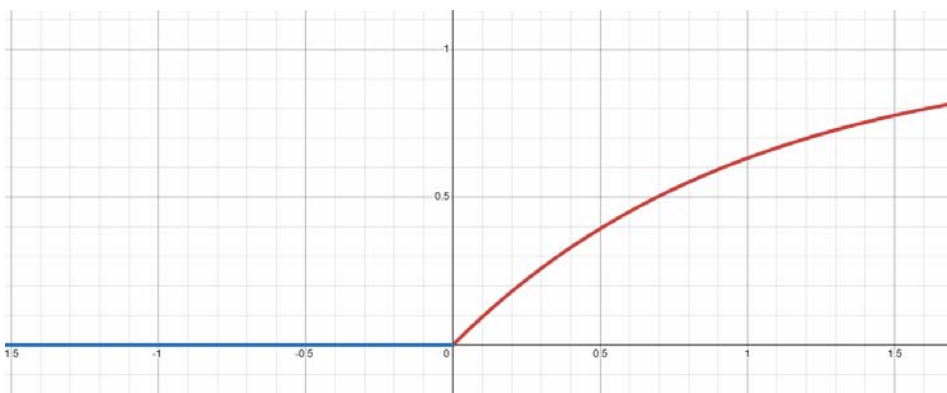
Значи:

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

7. Примери на распределби



Цртеж 7.2.2.1. Густина на експоненцијалната распределба со параметар $\lambda = 1$



Цртеж 7.2.2.2. Функција на распределба на експоненцијалната распределба со параметар $\lambda = 1$

Математичкото очекување е:

$$EX = \int_{-\infty}^{+\infty} xp(x)dx = \int_0^{+\infty} \lambda xe^{-\lambda x} dx = \lambda \lim_{x \rightarrow +\infty} \int_0^x te^{-\lambda t} dt,$$

каде што:

$$\lambda \int te^{-\lambda t} dt = \left(\begin{array}{ll} u = t & dv = e^{-\lambda t} dt \\ du = dt & v = \frac{e^{-\lambda t}}{-\lambda} \end{array} \right) = \lambda \left(-\frac{te^{-\lambda t}}{\lambda} + \frac{1}{\lambda} \int e^{-\lambda t} dt \right) = -te^{-\lambda t} + \frac{e^{-\lambda t}}{-\lambda},$$

$$\lambda \int_0^x t e^{-\lambda t} dt = -t e^{-\lambda t} + \frac{e^{-\lambda t}}{-\lambda} \Big|_0^x = -x e^{-\lambda x} - \frac{e^{-\lambda x}}{\lambda} + \frac{1}{\lambda}.$$

и

$$\lim_{x \rightarrow +\infty} x e^{-\lambda x} = \lim_{x \rightarrow +\infty} \frac{x}{e^{\lambda x}} \stackrel{LP}{=} \lim_{x \rightarrow +\infty} \frac{1}{\lambda e^{\lambda x}} = \frac{1}{+\infty} = 0, \quad \lim_{x \rightarrow +\infty} \frac{e^{-\lambda x}}{\lambda} = 0.$$

Следува:

$$EX = \frac{1}{\lambda}.$$

Сега ја пресметуваме дисперзијата.

$$EX^2 = \int_{-\infty}^{+\infty} x^2 p(x) dx = \int_0^{+\infty} \lambda x^2 e^{-\lambda x} dx = \lambda \lim_{x \rightarrow +\infty} \int_0^x t^2 e^{-\lambda t} dt,$$

каде што:

$$\begin{aligned} \lambda \int t^2 e^{-\lambda t} dt &= \left(\begin{array}{ll} u = t^2 & dv = e^{-\lambda t} dt \\ du = 2t dt & v = \frac{e^{-\lambda t}}{-\lambda} \end{array} \right) = \lambda \left(-\frac{t^2 e^{-\lambda t}}{\lambda} + \frac{2}{\lambda} \int t e^{-\lambda t} dt \right) = \\ &= -t^2 e^{-\lambda t} + 2 \left(-\frac{t e^{-\lambda t}}{\lambda} + \frac{1}{\lambda} \int e^{-\lambda t} dt \right) = -t^2 e^{-\lambda t} - \frac{2t e^{-\lambda t}}{\lambda} - \frac{2e^{-\lambda t}}{\lambda^2}, \\ \lambda \lim_{x \rightarrow +\infty} \int_0^x t^2 e^{-\lambda t} dt &= -t^2 e^{-\lambda t} - \frac{2t e^{-\lambda t}}{\lambda} - \frac{2e^{-\lambda t}}{\lambda^2} \Big|_0^x = \\ &= -x^2 e^{-\lambda x} - \frac{2x e^{-\lambda x}}{\lambda} - \frac{2(e^{-\lambda x} - 1)}{\lambda^2} = -x^2 e^{-\lambda x} - \frac{2x e^{-\lambda x}}{\lambda} - \frac{2e^{-\lambda x}}{\lambda^2} + \frac{2}{\lambda^2}. \end{aligned}$$

Аналогно, со Лопиталово правило, се покажува дека првите три граници на последниот израз се нула Кога $x \rightarrow +\infty$. Следува:

$$EX^2 = \frac{2}{\lambda^2},$$

поради што:

$$DX = EX^2 - (EX)^2 = \frac{2}{\lambda^2} - \frac{1}{\lambda^2} = \frac{1}{\lambda^2}.$$

7.2.3. Гама распределба

Случајната променлива X има **гама распределба** со параметри α и β , ако нејзината густина е:

$$p(x) = \begin{cases} 0, & x < 0 \\ \frac{x^{\alpha-1}}{\Gamma(\alpha)\beta^\alpha}, & x \geq 0 \end{cases}$$

каде што $\Gamma(\alpha) = \int_0^\infty x^{\alpha-1} e^{-x} dx$ е гама функција. Важи:

$$\Gamma(\alpha) = (\alpha-1)!, \quad \alpha \in \mathbb{N}.$$

Математичкото очекување на случајна променлива со Гама распределба е:

$$EX = \alpha\beta,$$

додека дисперзијата е:

$$DX = \alpha\beta^2.$$

(види литература [20]).

За $\alpha = 1$ и $\beta = \frac{1}{\lambda}$ добиваме:

$$p(x) = \begin{cases} 0, & x < 0 \\ \lambda e^{-\lambda x}, & x \geq 0 \end{cases}$$

Значи експоненцијалната распределба е специјален случај на гама распределбата.

7.2.4. Нормална (Гаусова) распределба

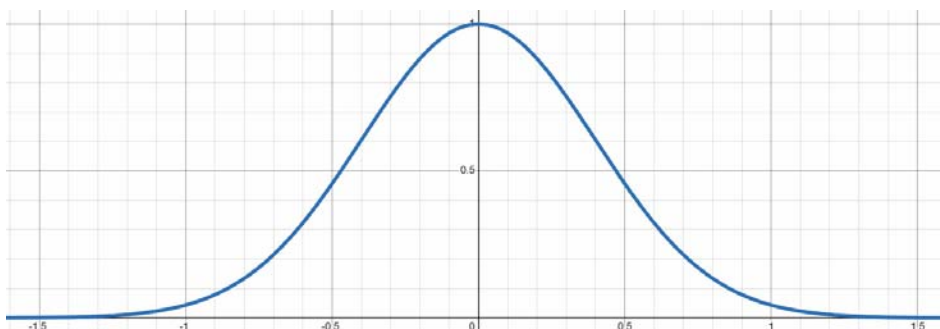
Случајната променлива X има **нормална распределба** со параметри a и σ^2 , ознака $N(a, \sigma^2)$, ако нејзината густина е:

$$p(x) = \frac{1}{\sqrt{2\pi}\sigma} e^{-\frac{(x-a)^2}{2\sigma^2}}, \quad a \in \mathbb{R}, \quad \sigma > 0.$$

Густината има максимална вредност ако и само ако $e^{-\frac{(x-a)^2}{2\sigma^2}}$ има максимум. Бидејќи $-\frac{(x-a)^2}{2\sigma^2} \leq 0$, важи $e^{-\frac{(x-a)^2}{2\sigma^2}} \leq 1$. Равенството се достигнува за:

$$-\frac{(x-a)^2}{2\sigma^2} = 0 \Leftrightarrow x-a=0 \Leftrightarrow x=a.$$

Затоа во $x = a$ густината $p(x)$ достигнува максимум.

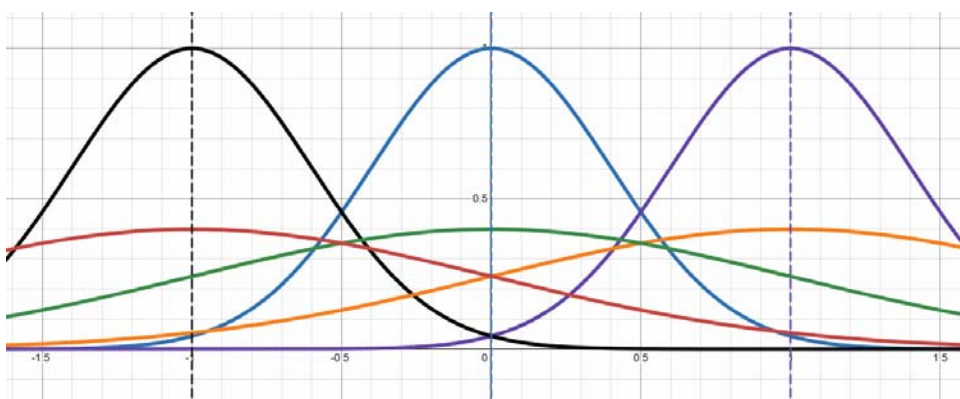


Цртеж 7.2.4. Густина на $N(a, \sigma^2)$ за $\sigma = \frac{1}{\sqrt{2\pi}}$ и $a = 0$

Важи:

$$p(a+t) = p(a-t) = \frac{1}{\sqrt{2\pi}\sigma} e^{-\frac{t^2}{2\sigma^2}}.$$

Следува дека $x = a$ е оска на симетрија на графикот на $p(x)$. За $x = 0$, оска на симетрија е y -оската. Значи првиот параметар го определува поместувањето на оската на симетрија во однос на y -оската. Вториот параметар σ го определува растегнувањето односно стеснувањето на графикот на $p(x)$, во однос на графикот на $p(x)$ за $\sigma = 1$; во правец на y -оската, односно висината на графикот.



Цртеж 7.2.4. Густини на нормални распределби за различни параметри a и σ

7. Примери на распределби

Виолетовиот, синиот и црниот график се добиваат за $\sigma = \frac{1}{\sqrt{2\pi}}$ и $a = -1$, $a = 0$ и $a = 1$ соодветно, тогаш $p = e^{-\pi(x-a)^2}$.

Останатите три графици се добиени за променето σ во $\sigma = 1$,

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

	0	0,01	0,02	0,03	0,04	0,05	0,06	0,07	0,08	0,09
0	0,5	0,504	0,508	0,512	0,516	0,519	0,523	0,527	0,531	0,535
0,1	0,539	0,543	0,547	0,551	0,555	0,559	0,563	0,567	0,571	0,575
0,2	0,579	0,583	0,587	0,591	0,594	0,598	0,602	0,606	0,61	0,614
0,3	0,617	0,621	0,625	0,629	0,633	0,636	0,64	0,644	0,648	0,651
0,4	0,655	0,659	0,662	0,666	0,67	0,673	0,677	0,68	0,684	0,687
0,5	0,691	0,695	0,698	0,701	0,705	0,708	0,712	0,715	0,719	0,722
0,6	0,725	0,729	0,732	0,735	0,738	0,742	0,745	0,748	0,751	0,754
0,7	0,758	0,761	0,764	0,767	0,77	0,773	0,776	0,779	0,782	0,785
0,8	0,788	0,791	0,793	0,796	0,799	0,802	0,805	0,807	0,81	0,813
0,9	0,815	0,818	0,821	0,823	0,826	0,828	0,831	0,834	0,836	0,838
1	0,841	0,843	0,846	0,848	0,85	0,853	0,855	0,857	0,859	0,862
1,1	0,864	0,866	0,868	0,87	0,872	0,874	0,877	0,879	0,881	0,883
1,2	0,884	0,886	0,888	0,89	0,892	0,894	0,896	0,898	0,899	0,901
1,3	0,903	0,904	0,906	0,908	0,909	0,911	0,913	0,914	0,916	0,917
1,4	0,919	0,92	0,922	0,923	0,925	0,926	0,927	0,929	0,93	0,931
1,5	0,933	0,934	0,935	0,937	0,938	0,939	0,94	0,941	0,942	0,944
1,6	0,945	0,946	0,947	0,948	0,949	0,95	0,951	0,952	0,953	0,954
1,7	0,955	0,956	0,957	0,958	0,959	0,959	0,96	0,961	0,962	0,963
1,8	0,964	0,964	0,965	0,966	0,967	0,967	0,968	0,969	0,969	0,97
1,9	0,971	0,971	0,972	0,973	0,973	0,974	0,975	0,975	0,976	0,976
2	0,977	0,977	0,978	0,978	0,979	0,979	0,98	0,98	0,981	0,981
2,1	0,982	0,982	0,983	0,983	0,983	0,984	0,984	0,985	0,985	0,985
2,2	0,986	0,986	0,986	0,987	0,987	0,987	0,988	0,988	0,988	0,989
2,3	0,989	0,989	0,989	0,99	0,99	0,99	0,99	0,991	0,991	0,991
2,4	0,991	0,992	0,992	0,992	0,992	0,992	0,993	0,993	0,993	0,993
2,5	0,993	0,994	0,994	0,994	0,994	0,994	0,994	0,994	0,995	0,995
2,6	0,995	0,995	0,995	0,995	0,995	0,996	0,996	0,996	0,996	0,996
2,7	0,996	0,996	0,996	0,996	0,996	0,997	0,997	0,997	0,997	0,997
2,8	0,997	0,997	0,997	0,997	0,997	0,997	0,997	0,997	0,998	0,998
2,9	0,998	0,998	0,998	0,998	0,998	0,998	0,998	0,998	0,998	0,998

7. Примери на распределби

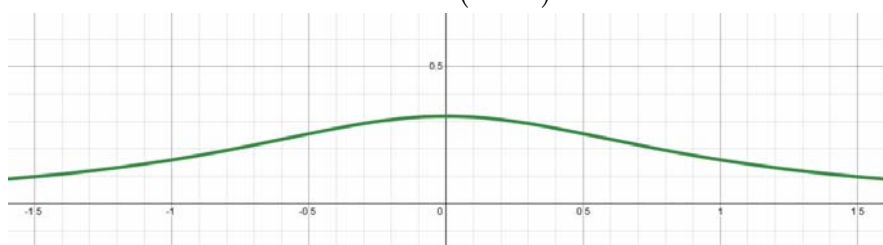
3	0,998	0,998	0,998	0,998	0,998	0,998	0,998	0,998	0,999	0,999
3,1	0,999	0,999	0,999	0,999	0,999	0,999	0,999	0,999	0,999	0,999
3,2	0,999	0,999	0,999	0,999	0,999	0,999	0,999	0,999	0,999	0,999

Табела 11.1.1. Вредности на нормалната распределба $N(0,1)$

7.2.5. Кошиева распределба

Случајната променлива X има **Кошиева распределба** ако нејзината густина е:

$$p(x) = \frac{1}{\pi(1+x^2)}.$$

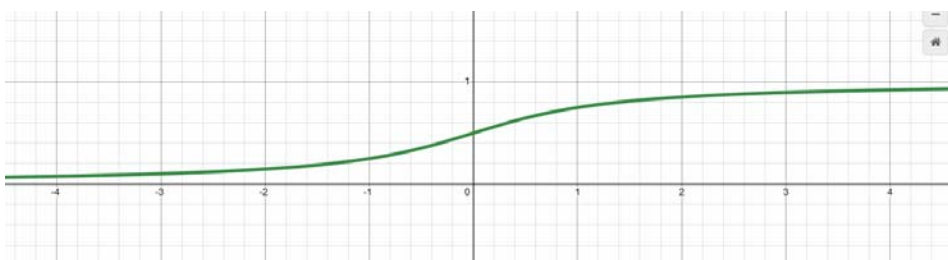


Цртеж 7.2.5.1. Густина на Кошиевата распределба

Функцијата на распределба е:

$$F(x) = \int_{-\infty}^x p(t) dt = \frac{1}{\pi} \lim_{a \rightarrow -\infty} \int_a^x \frac{1}{1+t^2} dt = \lim_{a \rightarrow -\infty} \arctg t \Big|_a^x =$$

$$\frac{1}{\pi} \left(\arctg x - \lim_{a \rightarrow -\infty} \arctg a \right) = \frac{1}{\pi} \left(\arctg x - \left(-\frac{\pi}{2} \right) \right) = \frac{\arctg x}{\pi} + \frac{1}{2}.$$



Цртеж 7.2.5.2. Функција на Кошиевата распределба

Математичкото очекување $EX = \frac{1}{\pi} \int_{-\infty}^{+\infty} \frac{x}{1+x^2} dx$ и дисперзија на случајните променливи со Кошиева распределба не постојат. Имено:

$$\int_0^{+\infty} \frac{x}{1+x^2} dx = \lim_{x \rightarrow +\infty} \int_0^x \frac{t}{1+t^2} dt = \lim_{x \rightarrow +\infty} \int_0^x \frac{d(1+t^2)}{2(1+t^2)} =$$

$$\lim_{x \rightarrow +\infty} \frac{1}{2} \ln(1+t^2) \Big|_0^x = \lim_{x \rightarrow +\infty} (\ln(1+x^2) - \ln 1) = +\infty.$$

Следува дека интегралот дивергира. Како последица и интегралот $\int_{-\infty}^{+\infty} \frac{|x|}{1+x^2} dx$ дивергира.

7.2.6. χ^2 -распределба

Нека претпоставиме дека случајната променлива X има стандардна нормална распределба $N(0,1)$. Да ја разгледаме случајната променлива $Y = X^2$. Оваа променлива е позитивна и нејзината функција на распределба изгледа поинаку од нормалната распределба. Се нарекува χ^2 (хи-квадрат) случајна променлива.

Нека $X_1, X_2, \dots, X_n$ се непрекинати случајни променливи со стандардна нормална распределба $N(0,1)$. Збирот на нивните квадрати $H = X_1^2 + X_2^2 + \dots + X_n^2$ дефинира случајна променлива која има χ^2 распределба. Оваа случајна променлива зависи од n параметри и затоа велиме дека случајната променлива има распределба со n степени на слобода. Степенот на слобода е бројот на линеарно независни случајни променливи $X_1, X_2, \dots, X_n$ во случајната променлива H . Ако е позната барем една врска помеѓу променливите, тогаш случајната променлива H ќе има $n-1$ степени на слобода.

Густината на χ^2 распределбата зависи само од степенот на слобода и таа е Ојлерова гама функција. Имено, случајната променлива X има χ^2 **распределба** со n степени на слобода, ако нејзината густина е:

$$p(x) = \begin{cases} 0, & x \leq 0 \\ \frac{x^{\frac{n}{2}-1} e^{-\frac{x}{2}}}{2^{\frac{n}{2}} \Gamma\left(\frac{n}{2}\right)}, & x > 0, \end{cases}$$

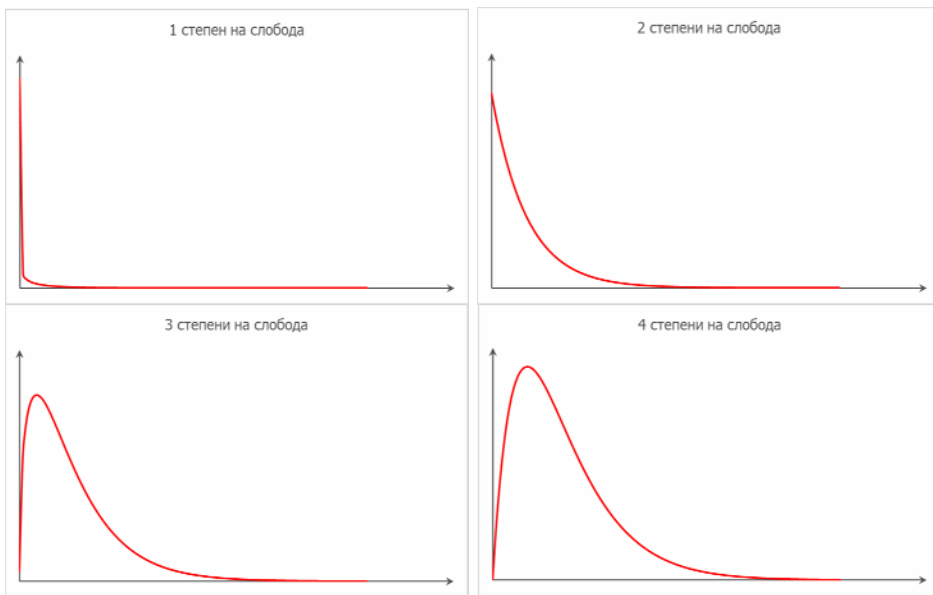
каде што $\Gamma\left(\frac{n}{2}\right)$ е гама функција.

Својства на функцијата на густина:

-Функцијата е дефинирана само за позитивни вредности на независната променлива.

-За негативни вредности, се зема дека вредноста на функцијата на густина е 0.

-Кога степенот на слобода е мал, функцијата на густина е асиметрична, а како што расте, графикот станува сè посиметричен. Кога степенот на слобода е голем број, распределбата асимптотски се приближува до нормалната распределба.



Цртеж 11.2.1-4. Графици на густината на χ^2 – распределба со 1, 2, 3 и 4 степени на слобода

Математичкото очекување на случајна променлива X која има χ^2 распределба со n степени на слобода е:

$$EX = n,$$

додека дисперзијата е:

$$DX = 2n.$$

(види литература [34]).

За пресметување на вредностите на распределбата се користат табели кои за даден степен на слобода, обично 1, 2, ..., 30 и даден број веројатност p , обично:

$$0,01; 0,05,$$

ги даваат вредностите $\chi^2_{n,p}$, такви што важи $P(\chi^2 < \chi^2_{n,p}) = 1 - p$.

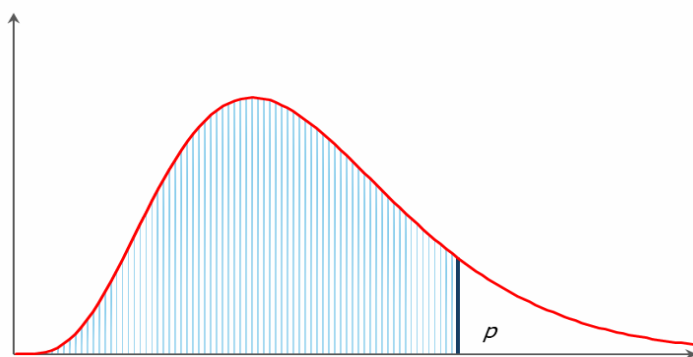
Една таква табела е дадена во прилог за максимум 10 степени на слобода.

степени на слобода	χ^2 -вредност										
1	0,004	0,02	0,06	0,15	0,46	1,07	1,64	2,71	3,84	6,64	10,83
2	0,1	0,21	0,45	0,71	1,39	2,41	3,22	4,6	5,99	9,21	13,82
3	0,35	0,58	1,01	1,42	2,37	3,66	4,64	6,25	7,82	11,34	16,27
4	0,71	1,06	1,65	2,2	3,36	4,88	5,99	7,78	9,49	13,28	18,47
5	1,14	1,61	2,34	3	4,35	6,06	7,29	9,24	11,07	15,09	20,52
6	1,63	2,2	3,07	3,83	5,35	7,23	8,56	10,64	12,59	16,81	22,46
7	2,17	2,83	3,82	4,67	6,35	8,38	9,8	12,02	14,07	18,48	24,32
8	2,73	3,49	4,59	5,53	7,34	9,52	11,03	13,36	15,51	20,09	26,12
9	3,32	4,17	5,38	6,39	8,34	10,66	12,24	14,68	16,92	21,67	27,88
10	3,94	4,87	6,18	7,27	9,34	11,78	13,44	15,99	18,31	23,21	29,59
p -вредност	0,95	0,9	0,8	0,7	0,5	0,3	0,2	0,1	0,05	0,01	0,001

Табела 11.2.2. Таблица $\chi^2_{n,p}$ во зависност од степените на слобода

Графикот на следниот цртеж 11.2.3. ја претставува функцијата на густина на χ^2 -распределбата, а веројатноста ја претставува исенчената површина на цртежот.

За $n \geq 30$ степени на слобода имајќи предвид дека математичкото очекување на χ^2 -распределбата е n , а дисперзијата е $2n$, χ^2 -распределбата се апроксимира со нормалната распределба $N(n, 2n)$.



Цртеж 11.2.2. График на густината на χ^2 – распределба и веројатност $1 - p$

7.2.7. Студентова распределба

Случајната променлива Z има стандардна нормална распределба и H има χ^2 распределба со k степени на слобода. Тогаш случајната променлива

$$T = \frac{Z}{\sqrt{\frac{H}{k}}}$$

има Студентова t -распределба со k степени на слобода.

Нека $X_1, X_2, \dots, X_n$ се непрекинати независни случајни променливи со нормална распределба $N(\mu, \sigma^2)$.

Нека $\bar{X} = \frac{X_1 + X_2 + \dots + X_n}{n}$ и $s^2 = \frac{1}{n-1} \sum_{i=1}^n (X_i - \bar{X})^2$. Тогаш

случајната променлива $H = (n-1) \frac{s^2}{\sigma^2}$ има χ^2 -распределба со $k = n-1$ степени на слобода [41].

Случајната променлива $Z = \frac{\bar{X}_n - \mu}{\sigma} \sqrt{n}$ има стандардна нормална распределба $N(0,1)$ бидејќи $\bar{X}_n$ има нормална

распределба $N\left(\mu, \frac{\sigma^2}{n}\right)$ и притоа случајните променливи H и T се независни.

Тогаш, случајната променлива:

$$T = \frac{Z}{\sqrt{\frac{H}{k}}} = \frac{(\bar{X}_n - \mu)\sqrt{n}}{s}$$

има Студентова распределба со $k = n - 1$ степени на слобода, [42].

Студентовата распределба (t -распределба) е една од најзначајните распределби во статистиката, особено кога се работи со помали примероци или кога стандардната девијација на популацијата е непозната. Студентовата распределба ја формулирал Вилијам Сили Госет во 1908 година додека работел за пиварницата Гинис во Даблин. Поради корпоративната политика, не можел да го користи своето име за академски публикации, па затоа го користел псевдонимот студент. Целта на неговата работа била да се анализира квалитетот на пивото со мали примероци, бидејќи било скапо да се работи со големи.

Густината на Студентовата t -распределба зависи само од степенот на слобода и таа е Ојлерова гама функција. Имено, случајната променлива X има Студентовата t -распределба со k степени на слобода, ако нејзината густина е:

$$p(x) = \frac{\Gamma\left(\frac{k+1}{2}\right)}{\sqrt{k\pi}\Gamma\left(\frac{k}{2}\right)} \left(1 + \frac{x^2}{k}\right)^{-\frac{k+1}{2}},$$

каде што $\Gamma\left(\frac{k+1}{2}\right)$ и $\Gamma\left(\frac{k}{2}\right)$ се гама функции.

Математичкото очекување на случајна променлива X која има Студентова t -распределба со k степени на слобода е:

$$EX = 0,$$

додека дисперзијата е:

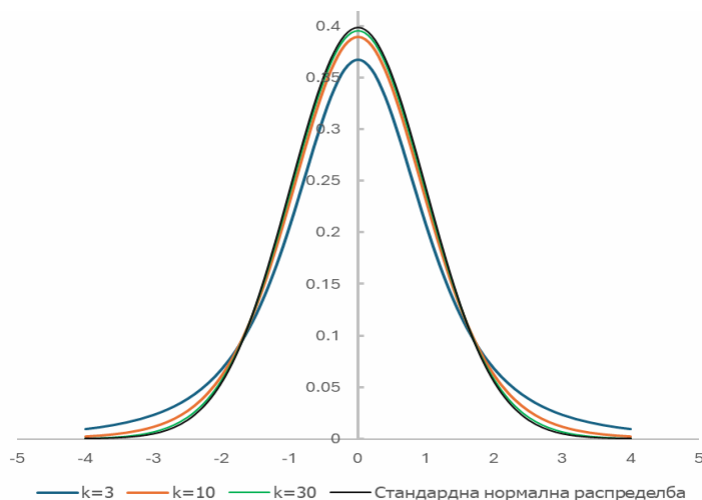
$$DX = \frac{k}{k-2}, k > 2$$

(види литература [42]).

Всушност, t -распределбата претставува фамилија од распределби, така што секоја од нив е одредена со параметарот k степени на слобода.

За да се разликуваат меѓу себе овие распределби, секоја t -распределба во заграда ја има вредноста на параметарот k . Графиците на густините на t -распределбите се симетрични во однос на нулата, исто како графикот на густината на стандардната нормална распределба. Меѓутоа, стандардната девијација на секоја t -распределбата е поголема од еден и зависи од вредноста на бројот на степени на слобода k , додека стандардната девијација од стандардната нормална распределба е еднаква на еден.

На Цртежот 11.3.1. се споредени графиците на густината на стандардната нормална распределба со графиците на густините на t -распределбите со 3, 10 и 30 степени на слобода. Општо земено, графиците на густините на t -распределбата имаат помала максимална висина и подебели опашки од графикот на густината на стандардната нормална распределба. А тоа ја прави t -распределбата покорисна при анализа на помали примероци, бидејќи тие опфаќаат поголема варијабилност.



Цртеж 11.3.1. Графици на густините на t -распределбите со 3, 10 и 30 степени на слобода и стандардната нормална распределба

Кривите на сите t -распределби се симетрични околу нула. Со зголемување на степените на слобода, t -распределбата се приближува до нормалната распределба. Над 30 степени на слобода, t -распределбата и нормалната распределба стануваат

7. Примери на распределби

слични. Од тие причини при примероци поголеми од 30 наместо статистиката одредена со t -распределбата се употребува статистика одредена со стандардната нормална распределба.

Студентовата распределба е основа за неколку различни t -тестови, што ќе ги опишеме во делот статистика. Табелата на Студентовата распределба за едностран и двостран тест за соодветните степени на слобода е дадена во прилог.

t - распределба	Ниво на интервална доверба (двостран)						
	80%	90%	95%	98%	99%	100%	99.90%
	α -ниво на значајност за двостран тест						
	0,2	0,1	0,05	0,02	0,01	0,002	0,001
степен на слобода	α -ниво на значајност за едностран тест						
	0,1	0,05	0,025	0,01	0,005	0,001	0,0005
1	3,078	6,314	12,706	31,821	63,657	318,310	636,619
2	1,886	2,920	4,303	6,965	9,925	22,327	31,599
3	1,638	2,353	3,182	4,541	5,841	10,215	12,924
4	1,533	2,132	2,776	3,747	4,604	7,173	8,610
5	1,476	2,015	2,571	3,365	4,032	5,893	6,869
6	1,440	1,943	2,447	3,143	3,707	5,208	5,959
7	1,415	1,895	2,365	2,998	3,499	4,785	5,408
8	1,397	1,860	2,306	2,896	3,355	4,501	5,041
9	1,383	1,833	2,262	2,821	3,250	4,297	4,781
10	1,372	1,812	2,228	2,764	3,169	4,144	4,587
11	1,363	1,796	2,201	2,718	3,106	4,025	4,437
12	1,356	1,782	2,179	2,681	3,055	3,930	4,318
13	1,350	1,771	2,160	2,650	3,012	3,852	4,221
14	1,345	1,761	2,145	2,624	2,977	3,787	4,140
15	1,341	1,753	2,131	2,602	2,947	3,733	4,073
16	1,337	1,746	2,120	2,583	2,921	3,686	4,015
17	1,333	1,740	2,110	2,567	2,898	3,646	3,965
18	1,330	1,734	2,101	2,552	2,878	3,610	3,922
19	1,328	1,729	2,093	2,539	2,861	3,579	3,883
20	1,325	1,725	2,086	2,528	2,845	3,552	3,850
21	1,323	1,721	2,080	2,518	2,831	3,527	3,819
22	1,321	1,717	2,074	2,508	2,819	3,505	3,792
23	1,319	1,714	2,069	2,500	2,807	3,485	3,768

7. Примери на распределби

24	1,318	1,711	2,064	2,492	2,797	3,467	3,745
25	1,316	1,708	2,060	2,485	2,787	3,450	3,725
26	1,315	1,706	2,056	2,479	2,779	3,435	3,707
27	1,314	1,703	2,052	2,473	2,771	3,421	3,690
28	1,313	1,701	2,048	2,467	2,763	3,408	3,674
29	1,311	1,699	2,045	2,462	2,756	3,396	3,659
30	1,310	1,697	2,042	2,457	2,750	3,385	3,646
40	1,303	1,684	2,021	2,423	2,704	3,307	3,551
50	1,299	1,676	2,009	2,403	2,678	3,261	3,496
60	1,296	1,671	2,000	2,390	2,660	3,232	3,460
80	1,292	1,664	1,990	2,374	2,639	3,195	3,416
100	1,290	1,660	1,984	2,364	2,626	3,174	3,390
120	1,289	1,658	1,980	2,358	2,617	3,160	3,373
1000	1,282	1,646	1,962	2,330	2,581	3,098	3,300
∞	1,282	1,645	1,960	2,326	2,576	3,090	3,291

Табела 11.3.2. Вредности на Студентовата распределба

8. ДВОДИМЕНЗИОНАЛНИ СЛУЧАЈНИ ПРОМЕНЛИВИ

8.1. ДВОДИМЕНЗИОНАЛНИ СЛУЧАЈНИ ПРОМЕНЛИВИ

Нека X и Y се произволни случајни променливи над ист простор на веројатност (Ω, F, P) и E е елементарен настан од множеството на елементарни настани F . Тогаш постои единствен вектор $(X(E), Y(E)) \in \mathbb{R}^2$. На овој начин се добива функција $(X, Y): \Omega \rightarrow \mathbb{R}^2$ која се нарекува двостепензионална случајна променлива или пар случајни променливи.

Дефиниција 8.1.1. Функција на распределба на двостепензионална случајна променлива (X, Y) е функцијата $F: \mathbb{R}^2 \rightarrow \mathbb{R}$ дефинирана со

$$F(x, y) = P(\{X \leq x\} \cap \{Y \leq y\}) = P\{X \leq x, Y \leq y\}, (x, y) \in \mathbb{R}^2.$$

Теорема 8.1.2. Ако $F(x, y)$ е функција на распределба на двостепензионалната случајна променлива (X, Y) , тогаш за $a < b$ и $c < d$, точно е равенството:

$$P\{a \leq X < b, c \leq Y < d\} = F(b, d) - F(b, c) - F(a, d) + F(a, c)$$

Доказ: Имаме,

$$\begin{aligned} P\{a < X \leq b, c < Y \leq d\} &= \\ P\{X \leq b, Y \leq d\} - P\{X \leq b, Y \leq c\} - P\{X \leq a, Y \leq d\} + P\{X \leq a, Y \leq c\} &= \\ F(b, d) - F(b, c) - F(a, d) + F(a, c) \end{aligned}$$

Теорема 8.1.3. Ако $F(x, y)$ е функција на распределба на двостепензионалната случајна променлива (X, Y) , тогаш F ги има следниве својства:

- 1) $F(x, y)$ е монотонно неопадачка по секоја од променливите;
- 2) $\lim_{x \rightarrow -\infty} F(x, y) = 0$, $\lim_{y \rightarrow -\infty} F(x, y) = 0$, $\lim_{(x, y) \rightarrow (+\infty, +\infty)} F(x, y) = 1$;
- 3) $F(x, y)$ е непрекината од десно по секоја од променливите.

Забелешка. Претходно наведените својства на функцијата на распределба на двостепенност на случајна променлива се потребни, но не и доволни за таа да биде функција на распределба на една двостепенност на случајна променлива (X, Y) . Тоа е илустрирано со следниот пример.

Пример 8.1.4. Функцијата:

$$F(x, y) = \begin{cases} 0, & x \leq 0 \text{ или } y \leq 0 \text{ или } x + y \leq 1 \\ 1, & \text{во останатите случаи} \end{cases},$$

ги задоволува условите од Теорема 8.1.3. но, според Теорема 8.1.2:

$$P\left\{\frac{1}{2} \leq X < 1, \frac{1}{2} \leq Y < 1\right\} = F(1, 1) - F\left(\frac{1}{2}, 1\right) - F\left(1, \frac{1}{2}\right) + F\left(\frac{1}{2}, \frac{1}{2}\right) = 1 - 1 - 1 + 0 = -1,$$

што е контрадикција доколку би претпоставиле дека функцијата $F(x, y)$ е функција на распределба на една двостепенност на случајна променлива (X, Y) , бидејќи веројатноста не може да биде негативна. Оттука, точна е следната теорема:

Теорема 8.1.5. Функцијата $F(x, y)$ е функција на распределба на двостепенност на случајна променлива (X, Y) ако и само ако се исполнети следниве четири услови:

- 1) $F(x, y)$ е монотонно неопадачка по секоја од променливите;
- 2) $\lim_{x \rightarrow -\infty} F(x, y) = 0$, $\lim_{y \rightarrow -\infty} F(x, y) = 0$, $\lim_{(x, y) \rightarrow (+\infty, +\infty)} F(x, y) = 1$;
- 3) $F(x, y)$ е непрекината од десно по секоја од променливите.
- 4) $F(b, d) - F(b, c) - F(a, d) + F(a, c) \geq 0$.

Исто како кај случајни променливи и кај двостепенностите случајни променливи разликуваме два вида: **дискретни двостепенност на случајни променливи** и **непрекинати двостепенност на случајни променливи**. Тие се дефинираат со обопштување на дефиницијата за дискретни, односно непрекинати случајни променливи.

8.2. ДИСКРЕТНИ ДВОСТЕПЕНЗИОНАЛНИ СЛУЧАЈНИ ПРОМЕНЛИВИ

Дефиниција 8.2.1. Двостепензионалната случајна променлива (X, Y) е дискретна, ако постои дискретно, односно конечно или преброиво множество $R_{(X,Y)} \subseteq \mathbb{R}^2$ така што:

$$P\{(X, Y) \in R_{(X,Y)}^c\} = 0.$$

Дискретната двостепензионална случајна променлива (X, Y) е целосно одредена со нејзиниот **закон на распределба**, односно со множеството вредности $R_{(X,Y)}$ и веројатностите

$$p_{ij} = P\{X = x_i, Y = y_j\} = P(x_i, y_j), \text{ за сите } (x_i, y_j) \in R_{(X,Y)},$$

при што $\sum_i \sum_j p_{ij} = 1$.

Ако множеството вредности $R_{(X,Y)}$ е конечно, односно:

$$R_{(X,Y)} = \{(x_i, y_j) \mid i = 1, \dots, m; j = 1, \dots, n\},$$

уште повеќе во случај кога бројот на елементи во ова множество е мал, законот на распределба на двостепензионалната случајна променлива (X, Y) да се претстави со табела од следниов облик:

$Y \setminus X$	x_1	x_2	$\dots$	x_m
y_1	p_{11}	p_{21}	$\dots$	p_{m1}
y_2	p_{12}	p_{22}	$\dots$	p_{m2}
$\vdots$	$\vdots$	$\vdots$	$\vdots$	$\vdots$
y_n	p_{1n}	p_{2n}	$\dots$	p_{mn}

Ако е познат законот на распределба на една дискретна двостепензионална случајна променлива исто како и кај дискретните случајни променливи, еднозначно може да се определи нејзината функција на распределба и обратно.

Постапката е следна: нека дискретна двостепензионална случајна променлива (X, Y) е зададена со нејзиниот закон на

распределба. Тогаш нејзината функција на распределба е одредена на следниов начин:

$$F(x, y) = P\{X \leq x, Y \leq y\} = \sum_{x_i \leq x} \sum_{y_j \leq y} p_{ij}, \text{ за сите } (x, y) \in \mathbb{R}^2.$$

Задача 8.2.1. Анализирано е образованието на секој член од група од 1000 луѓе и податоците се дадени во следнава табела:

Образование	Маж (0)	Жена (1)
основно (0)	60	40
средно (1)	240	200
дипломиран (2)	200	100
магистер (3)	100	60

Нека X е случајна променлива што прима вредност 0 ако членот на групата е маж, а 1 во спротивно, а Y е случајна променлива што доделува 0 ако членот на групата завршил само основно образование, 1 ако членот на групата највисокото образование што го стекнал е средно, 2 ако членот на групата е дипломиран и 3 ако членот на групата е магистер.

Одреди го законот на распределба на двостепенационалната случајна променлива (X, Y) и пресметај:

$$F(-1, 2), F(0, 2), F(2, 2), F(1, -2), F(-1, 2), F(2, 3), F(1, 4).$$

Решение. Множеството вредности на дадената двостепенационалната случајна променлива (X, Y) е:

$$R_{(X, Y)} = \{(i, j) \mid i = 0, 1, j = 0, 1, 2, 3\}.$$

Законот на распределба е даден со табелата:

$Y \setminus X$	0	1
0	$\frac{60}{1000} = 0,06$	$\frac{40}{1000} = 0,04$
1	$\frac{240}{1000} = 0,24$	$\frac{200}{1000} = 0,20$
2	$\frac{200}{1000} = 0,20$	$\frac{100}{1000} = 0,10$

3	$\frac{100}{1000} = 0,10$	$\frac{60}{1000} = 0,06$
---	---------------------------	--------------------------

Оттука,

$$F(-1, 2) = P(x \leq -1, y \leq 2) = 0.$$

$$F(0, 2) = P(x \leq 0, y \leq 2) =$$

$$P(x = 0, y = 0) + P(x = 0, y = 1) + P(x = 0, y = 2) =$$

$$0,06 + 0,24 + 0,2 = 0,5.$$

$$F(2, 2) = P(x \leq 2, y \leq 2) = P(x = 0, y = 0) + P(x = 0, y = 1) +$$

$$P(x = 0, y = 2) + P(x = 1, y = 0) + P(x = 1, y = 1) + P(x = 1, y = 2) =$$

$$0,06 + 0,24 + 0,2 + 0,04 + 0,2 + 0,1 = 0,84.$$

$$F(1, -2) = P(x \leq 1, y \leq -2) = 0.$$

$$F(2, 3) = P(x \leq 2, y \leq 3) = P(x = 0, y = 0) + P(x = 0, y = 1) +$$

$$P(x = 0, y = 2) + P(x = 0, y = 3) + P(x = 1, y = 0) + P(x = 1, y = 1) +$$

$$P(x = 1, y = 2) + P(x = 1, y = 3) =$$

$$0,06 + 0,24 + 0,2 + 0,1 + 0,04 + 0,2 + 0,1 + 0,06 = 1.$$

$$F(1, 4) = P(x \leq 1, y \leq 4) = P(x = 0, y = 0) + P(x = 0, y = 1) +$$

$$P(x = 0, y = 2) + P(x = 0, y = 3) + P(x = 1, y = 0) + P(x = 1, y = 1) +$$

$$P(x = 1, y = 2) + P(x = 1, y = 3) =$$

$$0,06 + 0,24 + 0,2 + 0,1 + 0,04 + 0,2 + 0,1 + 0,06 = 1.$$

8.3. НЕПРЕКИНАТИ ДВОСТИМЕНЗИОНАЛНИ СЛУЧАЈНИ ПРОМЕНЛИВИ

Дефиниција 8.3.1. Двостепенностен случаен променлива (X, Y) е непрекината, ако постои ненегативна интегрална функција $p(x, y), x, y \in \mathbb{R}$, таква што

$$F(x, y) = \int_{-\infty}^x \int_{-\infty}^y p(s, t) ds dt$$

Функцијата $p(x, y)$ се нарекува густина на распределба на двостепенностен случаен променлива (X, Y) . Бидејќи

$$\lim_{(x,y) \rightarrow (+\infty, +\infty)} F(x, y) = 1,$$

следува:

$$\int_{-\infty}^{+\infty} \int_{-\infty}^{+\infty} p(s, t) ds dt = \lim_{(x,y) \rightarrow (+\infty, +\infty)} \int_{-\infty}^x \int_{-\infty}^y p(s, t) ds dt = \lim_{(x,y) \rightarrow (+\infty, +\infty)} F(x, y) = 1$$

Да забележиме дека за функцијата на распределба на двостепенствоналната случајна променлива важи:

$$\frac{\partial^2 F(x, y)}{\partial x \partial y} = p(x, y),$$

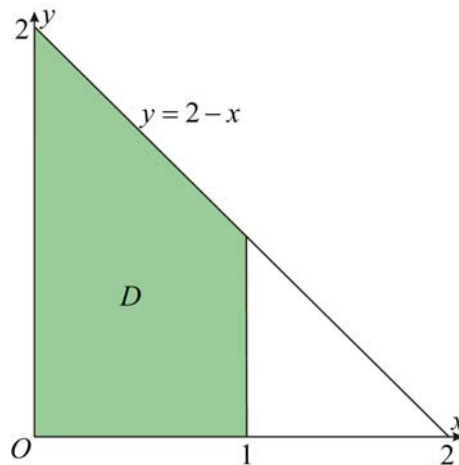
ако мешаниот извод на $F(x, y)$ постои во точката (x, y) .

Задача 8.3.1. Одреди ја константата ако функцијата:

$$p(x, y) = \begin{cases} c(x^2 + y^2), & (x, y) \in D : 0 \leq x \leq 1, y \geq 0, x + y \leq 2 \\ 0, & \text{инаку} \end{cases}$$

е густина на распределба на двостепенствоналната случајна променлива (X, Y) .

Решение. Областа D на која е дефинирана двостепенствоналната случајна променлива (X, Y) е прикажана на следниов цртеж 8.3.1.



Цртеж 8.3.1.

Бидејќи $p(x, y)$ е густина на распределба на случајната двостепенствонална случајна променлива (X, Y) точно е:

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

Оттука,

$$\begin{aligned} 1 &= \int_{-\infty}^{+\infty} \int_{-\infty}^{+\infty} p(x, y) dx dy = \iint_D c(x^2 + y^2) dx dy = \int_0^1 dx \int_0^{2-x} c(x^2 + y^2) dy = \\ &= c \int_0^1 dx \left(x^2 y + \frac{y^3}{3} \right) \Big|_0^{2-x} = c \int_0^1 \left(x^2(2-x) + \frac{(2-x)^3}{3} \right) dx = \\ &= c \int_0^1 \left(2x^2 - x^3 + \frac{8-12x+6x^2-x^3}{3} \right) dx = c \int_0^1 \left(\frac{8}{3} - 4x + 4x^2 - \frac{4}{3}x^3 \right) dx = \\ &= c \left(\frac{8}{3}x - 2x^2 + \frac{4}{3}x^3 - \frac{1}{3}x^4 \right) \Big|_0^1 = c \left(\frac{8}{3} - 2 + \frac{4}{3} - \frac{1}{3} \right) = \frac{5}{3}c. \end{aligned}$$

Затоа, $c = \frac{3}{5}$.

Задача 8.3.2. Одреди ја константата ако функцијата:

$$p(x, y) = \begin{cases} c(x+y), & (x, y) \in D : 0 < x < 2, 0 < y < 2 \\ 0, & \text{инаку} \end{cases}$$

е густина на распределба на двостепенствоналната случајна променлива (X, Y) .

Решение. Областа D на која е дефинирана двостепенствонална случајна променлива (X, Y) е прикажана на следниот цртеж 8.3.2.

Бидејќи $p(x, y)$ е густина на распределба на случајната двостепенствонална случајна променлива (X, Y) точно е:

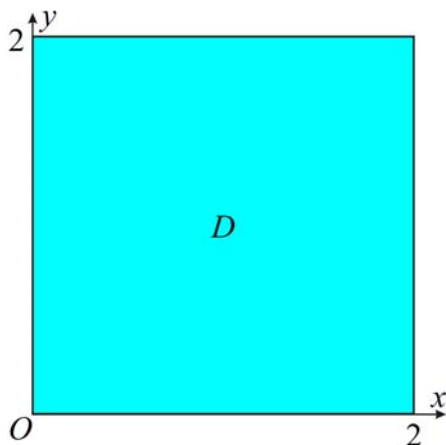
$$1 = \int_{-\infty}^{+\infty} \int_{-\infty}^{+\infty} p(x, y) dx dy.$$

Оттука,

$$1 = \int_{-\infty}^{+\infty} \int_{-\infty}^{+\infty} p(x, y) dx dy = \iint_D c(x+y) dx dy = \int_0^2 dx \int_0^2 c(x+y) dy =$$

$$c \int_0^2 dx \left(xy + \frac{y^2}{2} \right) \Big|_0^2 = c \int_0^2 (2x+2) dx = c(x^2 + 2x) \Big|_0^2 = 8c.$$

Затоа, $c = \frac{1}{8}$.



Цртеж 8.3.2.

8.4. МАРГИНАЛНИ РАСПРЕДЕЛБИ

Често пати е неопходно да се определат распределбите на секоја од променливите X и Y , поединечно ако е зададена распределбата на двостепенна случајна променлива (X, Y) .

Тие распределби се нарекуваат маргинални распределби на променливите X и Y , соодветно.

Имено, нека двостепенна случајна променлива (X, Y) е зададена со својата функција на распределба $F(x, y)$. Тогаш, маргиналните функции на распределбите $F_X(x)$ и $F_Y(y)$ на променливите X и Y , соодветно се одредени со следниве формули:

$$F_X(x) = P\{X \leq x\} = P\{X \leq x, Y < +\infty\} = \lim_{y \rightarrow +\infty} F(x, y);$$

$$F_Y(y) = P\{Y \leq y\} = P\{X < +\infty, Y \leq y\} = \lim_{x \rightarrow +\infty} F(x, y).$$

Нека (X, Y) е дискретна двостепензионална случајна променлива зададена со законот на распределба

$$p_{ij} = P\{X = x_i, Y = y_j\}, (x_i, y_j) \in R_{(X, Y)}.$$

Бидејќи $R_{(X, Y)} = R_X \times R_Y$ за сите вредности x_i од множеството R_X , маргиналниот закон на распределба на случајната променлива X е одреден со:

$$p_i = P\{X = x_i\} = P\{X = x_i, Y \in R_Y\} = \sum_{y_j \in R_Y} P\{X = x_i, Y = y_j\}.$$

На ист начин, маргиналниот закон на распределба на Y е:

$$q_j = P\{Y = y_j\} = P\{X \in R_X, Y = y_j\} = \sum_{x_i \in R_X} P\{X = x_i, Y = y_j\},$$

за секој $y_j \in R_Y$.

Задача 8.4.1. Да се одредат маргиналните распределби на дискретната двостепензионална случајна променлива (X, Y) од задачата 8.2.1, каде што X е случајна променлива што прима вредност 0 ако членот на групата е маж, 1 во спротивно, а Y е случајна променлива што доделува 0 ако членот на групата завршил само основно образование, 1 ако членот на групата највисокото образование што го стекнал е средно, 2 ако членот на групата е дипломиран и 3 ако членот на групата е магистер.

Решение. Маргиналните распределби на случајните променливи X и Y можат да се одредат директно со собирање на веројатностите по секоја колона и секоја редица, соодветно.

Имено,

$Y \setminus X$	0	1	$P(Y)$
0	0,06	0,04	0,1
1	0,24	0,2	0,44
2	0,2	0,1	0,3
3	0,1	0,06	0,16
$P(X)$	0,6	0,4	

Оттука, маргиналната распределба на случајната променлива X е дадена со табелата:

8. Двостепенни случајни променливи

X	0	1
$P(X)$	0,6	0,4

А маргиналната распределба на случајната променлива Y е дадена со табелата:

Y	0	1	2	3
$P(Y)$	0,1	0,44	0,3	0,16

Нека (X, Y) е непрекината двостепенна случајна променлива зададена со густината на распределба $p(x, y)$. Маргиналната густина на распределба на случајната променлива X ја одредуваме на следниот начин.

Имено, функцијата на распределба на X е дефинирана со:

$$F_X(x) = \lim_{y \rightarrow +\infty} F(x, y) = \lim_{y \rightarrow +\infty} \int_{-\infty}^x \int_{-\infty}^y p(s, t) dt ds = \int_{-\infty}^x \int_{-\infty}^{+\infty} p(s, t) dt ds ,$$

од друга страна:

$$F_X(x) = \int_{-\infty}^x p_X(s) ds .$$

Изедначувајќи ги десните страни на последните две равенства, имаме:

$$p_X(s) = \int_{-\infty}^{+\infty} p(s, t) dt ,$$

односно:

$$p_X(x) = \int_{-\infty}^{+\infty} p(x, y) dy .$$

На ист начин, за маргиналната густина на Y се добива:

$$p_Y(y) = \int_{-\infty}^{+\infty} p(x, y) dx .$$

Забележуваме дека ако се бара маргиналната густина на една од променливите, тогаш заедничката густина на двостепенната случајна променлива се интегрира по аргументот соодветен на другата променлива.

Задача 8.4.2. Да се одредат маргиналните густини на распределба на непрекинатата двостепенствителна случајна променлива (X, Y) од задачата 8.3.1.

Решение. Прво да забележиме според задачата 8.3.1 дека густината на распределба на дискретната двостепенствителна случајна променлива (X, Y) е:

$$p(x, y) = \begin{cases} \frac{3}{5}(x^2 + y^2), & (x, y) \in D: 0 \leq x \leq 1, y \geq 0, x + y \leq 2 \\ 0, & \text{инаку} \end{cases}$$

Тогаш, за да ја одредиме маргиналната густина на распределба на случајната променлива X прво земаме:

Ако $0 \leq x \leq 1$ бидејќи во D важи $y \leq 2 - x$

$$\begin{aligned} p_X(x) &= \int_0^{2-x} \frac{3}{5}(x^2 + y^2) dy = \frac{3}{5} \left(x^2 y + \frac{y^3}{3} \right) \Big|_0^{2-x} = \frac{3}{5} \left(x^2(2-x) + \frac{(2-x)^3}{3} \right) = \\ &= \frac{3}{5} \left(\frac{8}{3} - 4x + 4x^2 - \frac{4}{3}x^3 \right), \end{aligned}$$

тогаш, маргиналната густина на распределба на случајната променлива X е:

$$p_X(x) = \begin{cases} \frac{3}{5} \left(\frac{8}{3} - 4x + 4x^2 - \frac{4}{3}x^3 \right), & 0 \leq x \leq 1 \\ 0, & \text{инаку} \end{cases}$$

Во согласност со цртежот 8.3.1, ако $0 \leq y \leq 1$, тогаш $0 \leq x \leq 1$ и

$$p_Y(y) = \int_0^1 \frac{3}{5}(x^2 + y^2) dx = \frac{3}{5} \left(\frac{x^3}{3} + y^2 x \right) \Big|_0^1 = \frac{3}{5} \left(\frac{1}{3} + y^2 \right) = \frac{1+3y^2}{5};$$

ако $1 \leq y \leq 2$ тогаш $0 \leq x \leq 2 - y$ и

$$\begin{aligned} p_Y(y) &= \int_0^{2-y} \frac{3}{5}(x^2 + y^2) dx = \frac{3}{5} \left(\frac{x^3}{3} + y^2 x \right) \Big|_0^{2-y} = \frac{3}{5} \left(\frac{(2-y)^3}{3} + y^2(2-y) \right) = \\ &= \frac{3}{5} \left(\frac{8}{3} - 4y + 4y^2 - \frac{4}{3}y^3 \right). \end{aligned}$$

Тогел, маргелнелнел густелнел нел релспрелелбел нел слугелнелнел протенлвл Y е:

$$p_Y(y) = \begin{cases} \frac{1+3y^2}{5}, & 0 \leq y \leq 1 \\ \frac{3}{5} \left(\frac{8}{3} - 4y + 4y^2 - \frac{4}{3}y^3 \right), & 1 \leq y \leq 2 \\ 0, & \text{инаку} \end{cases}$$

Зелелел 8.4.3. Дел се одредел маргелнелнел густелнел нел релспрелелбел нел дскретнел двостепенствонелн слугелнелнел протенлвл (X, Y) од зелелелел 8.3.2.

Решелел. Прво дел забележеле спорел зелелел 8.3.2. делел густелнел нел релспрелелбел нел дскретнел двостепенствонелн слугелнелнел протенлвл (X, Y) е

$$p(x, y) = \begin{cases} \frac{1}{8}(x+y), & (x, y) \in D : 0 < x < 2, 0 < y < 2 \\ 0, & \text{инаку} \end{cases}$$

Тогел, дел дел одределе маргелнелнел густелнел нел релспрелелбел нел слугелнелнел протенлвл X прво зелелел:

Ако $0 \leq x \leq 2$ белејкел во D велел $0 \leq y \leq 2$

$$p_X(x) = \int_0^2 \frac{1}{8}(x+y) dy = \frac{1}{8} \left(xy + \frac{y^2}{2} \right) \Big|_0^2 = \frac{x+1}{4},$$

Тогел, маргелнелнел густелнел нел релспрелелбел нел слугелнелнел протенлвл X е:

$$p_X(x) = \begin{cases} \frac{x+1}{4}, & 0 \leq x \leq 2 \\ 0, & \text{инаку} \end{cases}$$

Во согласност со цртежот 8.3.2 ако $0 \leq y \leq 2$, велел $0 \leq x \leq 2$ и

$$p_Y(y) = \int_0^2 \frac{1}{8}(x+y) dx = \frac{1}{8} \left(\frac{x^2}{2} + xy \right) \Big|_0^2 = \frac{y+1}{4}.$$

Тогаш, маргиналната густина на распределба на случајната променлива Y е:

$$p_Y(y) = \begin{cases} \frac{y+1}{4}, & 0 \leq y \leq 2 \\ 0, & \text{инаку} \end{cases}.$$

8.5. УСЛОВНА РАСПРЕДЕЛБА

Нека настанот $\{Y \in S\}$ за некое $S \subseteq \mathbb{R}$ има позитивна веројатност, односно $P\{Y \in S\} > 0$. Тогаш функцијата на условна распределба на X при услов $\{Y \in S\}$ се дефинира на следниов начин:

$$F_X(x | \{Y \in S\}) = P(\{X \leq x\} | \{Y \in S\}) = \frac{P\{X \leq x, Y \in S\}}{P\{Y \in S\}}, \quad x \in \mathbb{R}.$$

Дефиниција 8.5.1. Нека (X, Y) е дискретна двостепенционална случајна променлива зададена со закон на распределба $p(x_i, y_j) = P\{X = x_i, Y = y_j\}$, за $(x_i, y_j) \in R_{(X,Y)}$ и нека маргиналните закони на распределба на X и на Y се означени со

$$p_X(x_i) = P\{X = x_i\}, x_i \in R_X \text{ и } p_Y(y_j) = P\{Y = y_j\}, y_j \in R_Y,$$

соодветно.

Законот на условна распределба на случајната променлива X при услов $\{Y = y_j\}$, каде што $P\{Y = y_j\} > 0$, се дефинира со:

$$p_X(x_i | y_j) = P(\{X = x_i\} | \{Y = y_j\}) = \frac{P\{X = x_i, Y = y_j\}}{P\{Y = y_j\}} = \frac{p(x_i, y_j)}{p_Y(y_j)}$$

за секој $x_i \in R_X$ и означуваме $X/Y = y_j$.

Законот на условна распределба на случајната променлива Y при услов $\{X = x_i\}$, каде што $P\{X = x_i\} > 0$, се дефинира со:

$$p_Y(y_j | x_i) = P(\{Y = y_j\} | \{X = x_i\}) = \frac{P\{Y = y_j, X = x_i\}}{P\{X = x_i\}} = \frac{p(x_i, y_j)}{p_X(x_i)},$$

за секој $y_j \in R_Y$ и означуваме $Y/X = x_i$.

Задача 8.5.2. Да се одреди условната распределба $Y/X = 0$ од дискретната двостепенционална случајна променлива (X, Y) во задачата 8.2.1, каде што X е случајна променлива што прима вредност 0 ако членот на групата е маж, 1 во спротивно, а Y е случајна променлива што доделува 0 ако членот на групата завршил само основно образование, 1 ако членот на групата највисокото образование што го стекнал е средно, 2 ако членот на групата е дипломиран и 3 ако членот на групата е магистер.

Решение. Бидејќи $R_{Y/X=0} = \{0, 1, 2, 3\}$

$$P(Y=0/X=0) = \frac{60}{600} = \frac{1}{10}, P(Y=1/X=0) = \frac{240}{600} = \frac{2}{5}$$

$$P(Y=2/X=0) = \frac{200}{600} = \frac{1}{3}, P(Y=3/X=0) = \frac{100}{600} = \frac{1}{6}$$

условната распределба $Y/X = 0$ е зададена во табелата:

$Y/X=0$	0	1	2	3
$P(Y/X=0)$	$\frac{1}{10}$	$\frac{2}{5}$	$\frac{1}{3}$	$\frac{1}{6}$

Да забележиме дека кога (X, Y) е непрекината двостепенционална случајна променлива, условната распределба на X при услов $\{Y=y\}$ не може да се дефинира со претходната дефиниција, бидејќи $P\{Y=y\} = 0$.

Доколку сакаме да дефинираме густина на условна распределба на X при услов $\{Y=y\}$, претходно ќе ја разгледаме следната дискусија.

Нека (X, Y) е непрекината двостепенционална случајна променлива зададена со густина $p(x, y)$ и нека y е вредност во која маргиналната густина на распределба на Y е позитивна, односно $p_Y(y) > 0$.

Претпоставуваме дека $p(x, y)$ и $p_Y(y)$ се непрекинати секаде каде што тоа е потребно при математичкото изведување.

За условната функција на распределба на X при услов $\{y \leq Y < y+h\}$, за $h > 0$, имаме:

$$F_X(x | \{y \leq Y < y+h\}) = \frac{P\{X < x, y \leq Y < y+h\}}{P\{y \leq Y < y+h\}} = \frac{\int_{-\infty}^x \int_y^{y+h} p(s,t) dt ds}{\int_y^{y+h} p_Y(t) dt} = \int_{-\infty}^x \left(\frac{\int_y^{y+h} p(s,t) dt ds}{\int_y^{y+h} p_Y(t) dt} \right) ds.$$

За двата интеграла кои се наоѓаат во заграда во последниот интергал ја применувам теоремата за средна вредност на интегралното сметање, која гласи:

Нека f е непрекината функција на интервал $[a, b]$. Тогаш

$$\text{постои } c \in [a, b], \text{ така што } \int_a^b f(x) dx = f(c)(b-a),$$

или еквивалентното тврдење кое гласи: постои $0 \leq d \leq 1$, така што

$$\int_a^b f(x) dx = f(a + d(b-a))(b-a).$$

Ако се примени оваа теорема интегралот $\int_y^{y+h} p(s,t) dt$,

разгледувајќи ја функцијата $p(s,t)$ како функција од t (s ја сметаме за константа), тогаш постои c_1 , ($0 < c_1 < 1$) така што:

$$y + c_1 h \in [y, y+h], \text{ па } \int_y^{y+h} p(s,t) dt = p(s, y + c_1 h) h.$$

Исто ако се примени оваа теорема интегралот $\int_y^{y+h} p_Y(t) dt$, од

теоремата за средна вредност на интегралното сметање постои c_2 ($0 < c_2 < 1$), така што $y + c_2 h \in [y, y+h]$ и

$$\int_y^{y+h} p_Y(t) dt = p_Y(y + c_2 h) h \int_y^{y+h} p_Y(v) dv = p_Y(y + c_2 h) h.$$

Последните два интеграла ги заменуваме во претходниот израз за условната функција на распределба на X при услов $\{y \leq Y < y+h\}$ и добиваме:

$$F_X(x | \{y \leq Y < y+h\}) = \int_{-\infty}^x \frac{p(s, y+c_1h)}{p_Y(y+c_2h)} ds.$$

Кога $h \rightarrow 0^+$, ја добиваме дефиницијата на функција на условна распределба на X при услов $\{Y = y\}$, за $p_Y(y) > 0$:

$$F_X(x | y) = \int_{-\infty}^x \frac{p(s, y)}{p_Y(y)} ds, \quad x \in \mathbb{R}$$

Оттука ја добиваме следнава дефиниција:

Дефиниција 8.5.3. Условна густина на непрекинатата случајна променлива X , при услов $\{Y = y\}$, за $p_Y(y) > 0$, се функцијата:

$$p_X(x | y) = \frac{p(x, y)}{p_Y(y)}, \quad x \in \mathbb{R}.$$

По дефиниција $p_X(x | y) \geq 0$, и $F_X(x | y) = \int_{-\infty}^x p_X(s | y) ds, x \in \mathbb{R}$.

Исто така,

$$\int_{-\infty}^{+\infty} p_X(x | y) dx = \int_{-\infty}^{+\infty} \frac{p(x, y)}{p_Y(y)} dx = \frac{\int_{-\infty}^{+\infty} p(x, y) dx}{p_Y(y)} = \frac{p_Y(y)}{p_Y(y)} = 1$$

Значи, функцијата $p_X(x | y) = \frac{p(x, y)}{p_Y(y)}, x \in \mathbb{R}$ е навистина густина на распределба.

На ист начин се дефинира условна густина на Y при услов $\{X = x\}$, за $p_X(x) > 0$, со равенството:

$$p_Y(y | x) = \frac{p(x, y)}{p_X(x)}, \quad y \in \mathbb{R}.$$

Задача 8.5.4. Да се одредат условната густина на распределба на случајната променлива $X/Y = 0,5$ од

непрекинатата двостепенност случајна променлива (X, Y) од задачата 8.3.1.

Решение. Бидејќи $p_{X/Y}(x) = \frac{p(x, y)}{p_Y(y)}$, следува:

$$p_{X/Y \in [0,1]}(x) = \frac{\frac{3}{5}(x^2 + y^2)}{\frac{1+3y^2}{5}} = \frac{3x^2 + 3y^2}{1+3y^2}.$$

И затоа што $0,5 \in [0,1]$, условната густина на $X/Y = 0,5$ е

$$p_{X/Y=0,5}(x) = \frac{3x^2 + 3 \cdot 0,5^2}{1 + 3 \cdot 0,5^2} = \frac{3x^2 + 0,75}{1,75}.$$

8.6. Независност на случајни променливи

Независност на случајни променливи се дефинира со помош на независноста на случајни настани.

Дефиниција 8.6.1. Случајните променливи $X_1, X_2, \dots, X_n$ се независни, ако за произволни $S_i \subseteq \mathbb{R}, i=1, 2, \dots, n$, случајните настани $\{X_1 \in S_1\}, \{X_2 \in S_2\}, \dots, \{X_n \in S_n\}$ се независни, односно:

$$P(\{X_1 \in S_1\} \cap \{X_2 \in S_2\} \cap \dots \cap \{X_n \in S_n\}) = \prod_{i=1}^n P\{X_i \in S_i\}.$$

Ако разгледуваме две случајни променливи, тогаш:

Дефиниција 8.6.2. Случајните променливи X и Y се независни, ако за произволни $S_1 \subseteq \mathbb{R}$ и $S_2 \subseteq \mathbb{R}$, случајните настани $\{X \in S_1\}, \{Y \in S_2\}$ се независни, односно

$$P(\{X \in S_1\} \cap \{Y \in S_2\}) = P\{X \in S_1\} \cdot P\{Y \in S_2\}$$

Ако $S_1 = (-\infty, x]$, тогаш настанот $\{X \in S_1\}$ може да се запише во следниов обликот:

$$\{X \in S_1\} = \{X \in (-\infty, x]\} = \{X \leq x\}.$$

Аналогно, ако $S_2 = (-\infty, y]$, тогаш $\{Y \in S_2\} = \{Y \leq y\}$.

За вака избрани S_1 и S_2 , равенството од дефиницијата 8.6.2. може да се запише во облик:

$$P\{X \leq x, Y \leq y\} = P\{X \leq x\} \cdot P\{Y \leq y\}, \text{ за секој } (x, y) \in \mathbb{R}^2$$

Односно

$$F(x, y) = F_X(x)F_Y(y)$$

од каде што заклучуваме дека ако X и Y се независни случајни променливи, тогаш заедничката функција на распределба на двостепензионалната случајна променлива (X, Y) може да се претстави како производ на маргиналните функции на распределба на случајните променливи X и Y .

Од друга страна, важи и обратното тврдење, односно ако за да двостепензионалната случајна променлива (X, Y) може да се претстави како производ на маргиналните функции на распределба на случајните променливи X и Y , тогаш X и Y се независни случајни променливи.

Теорема 8.6.3. Случајните променливи X и Y се независни ако и само ако двостепензионалната случајна променлива (X, Y) може да се претстави како производ на маргиналните функции на распределба на случајните променливи X и Y , односно

$$F(x, y) = F_X(x)F_Y(y)$$

Ако (X, Y) е дискретна двостепензионална случајна променлива, тогаш потребниот и доволен услов за независност на случајните променливи X и Y е од следниов облик:

$$P\{X = x, Y = y\} = P\{X = x\}P\{Y = y\}, \text{ за сите } (x, y) \in R_{(X, Y)}.$$

Ако (X, Y) е непрекината двостепензионална случајна променлива зададена со густина на распределба $p(x, y)$, тогаш потребниот и доволен услов за независност на случајните променливи X и Y е од следниот облик:

$$p(x, y) = p_X(x)p_Y(y), \quad x, y \in \mathbb{R}.$$

Од практични причини кога се проверува независност на две случајни променливи се користи контрапозицијата на тврдењето од следнава теорема

Теорема 8.6.4. Ако случајните променливи X и Y се независни тогаш $E(XY) = E(X)E(Y)$, при што:

-Кога X и Y се дискретни случајни променливи

$$E(XY) = \sum_x \sum_y xy p(x, y),$$

каде што $p(x, y)$ е распределбата на двостепенностен случаен променлива (X, Y) ;

-Кога X и Y се непрекинати случајни променливи

$$E(XY) = \int_{-\infty}^{+\infty} \int_{-\infty}^{+\infty} xy p(x, y) dx dy,$$

каде што $p(x, y)$ е густина на распределба на двостепенностен случаен променлива (X, Y) .

Имено, ако важи дека $E(XY) \neq E(X)E(Y)$ тогаш случајните променливи X и Y не се независни.

Задача 8.6.5. Да се одреди дали случајните променливи X и Y од дискретната двостепенностен случаен променлива (X, Y) во задачата 8.2.1 се независни.

Решение. Законите за распределби на случајните променливи X и Y и дискретната двостепенностен случаен променлива (X, Y) се дадени во следнава табела:

$Y \setminus X$	0	1	$P(Y)$
0	0,06	0,04	0,1
1	0,24	0,2	0,44
2	0,2	0,1	0,3
3	0,1	0,06	0,16
$P(X)$	0,6	0,4	

Оттука,

$$E(XY) = \sum_{x=0}^1 \sum_{y=0}^3 xy p(x, y) = 1 \cdot 0,2 + 2 \cdot 0,1 + 3 \cdot 0,06 = 0,58;$$

$$E(X) = \sum_{x=0}^1 xp_X(x) = 1 \cdot 0,4 = 0,4;$$

$$E(Y) = \sum_{y=0}^3 yp_Y(y) = 1 \cdot 0,44 + 2 \cdot 0,3 + 3 \cdot 0,16 = 1,52.$$

Бидејќи $E(XY) \neq E(X)E(Y)$ тогаш случајните променливи X и Y не се независни.

Задача 8.6.6. Да се одреди дали случајните променливи X и Y од непрекинатата двостепенна случајна променлива (X, Y) од задачата 8.3.1 се независни.

Решение. Од задача 8.3.1 густината на веројатност на двостепенна случајна променлива (X, Y) е:

$$p(x, y) = \begin{cases} \frac{3}{5}(x^2 + y^2), & (x, y) \in D: 0 \leq x \leq 1, y \geq 0, x + y \leq 2 \\ 0, & \text{инаку} \end{cases},$$

додека од задача 8.4.2, маргиналните густини на веројатност на случајните променливи X и Y , соодветно се:

$$p_X(x) = \begin{cases} \frac{3}{5} \left(\frac{8}{3} - 4x + 4x^2 - \frac{4}{3}x^3 \right), & 0 \leq x \leq 1 \\ 0, & \text{инаку} \end{cases}$$

$$p_Y(y) = \begin{cases} \frac{1+3y^2}{5}, & 0 \leq y \leq 1 \\ \frac{3}{5} \left(\frac{8}{3} - 4y + 4y^2 - \frac{4}{3}y^3 \right), & 1 \leq y \leq 2 \\ 0, & \text{инаку} \end{cases}$$

Оттука,

$$E(XY) = \int_{-\infty}^{+\infty} \int_{-\infty}^{+\infty} p(x, y) dx dy = \iint_D \frac{3}{5} xy(x^2 + y^2) dx dy =$$

$$\begin{aligned}
 \int_0^1 x dx \int_0^{2-x} \frac{3}{5} y (x^2 + y^2) dy &= \frac{3}{5} \int_0^1 x dx \left(\frac{x^2 y^2}{2} + \frac{y^4}{4} \right) \Big|_0^{2-x} = \\
 &= \frac{3}{5} \int_0^1 \left(x^3 (2-x)^2 + \frac{x(2-x)^4}{4} \right) dx = \\
 &= \frac{3}{5} \int_0^1 \left(4x^3 - 4x^4 + x^5 + \frac{16x - 32x^2 + 24x^3 - 8x^4 + x^5}{4} \right) dx = \\
 &= \frac{3}{5} \int_0^1 \left(10x^3 - 6x^4 + \frac{5}{4}x^5 + 4x - 8x^2 \right) dx = \\
 &= \frac{3}{5} \left(\frac{10}{4}x^4 - \frac{6}{5}x^5 + \frac{5}{24}x^6 + 2x^2 - \frac{8}{3}x^3 \right) \Big|_0^1 = \frac{3}{5} \left(\frac{10}{4} - \frac{6}{5} + \frac{5}{24} + 2 - \frac{8}{3} \right) = \frac{101}{200} \\
 E(X) &= \int_{-\infty}^{+\infty} x \cdot p_X(x) dx = \frac{3}{5} \int_0^1 x \left(\frac{8}{3} - 4x + 4x^2 - \frac{4}{3}x^3 \right) dx = \\
 &= \frac{3}{5} \int_0^1 \left(\frac{8}{3}x - 4x^2 + 4x^3 - \frac{4}{3}x^4 \right) dx = \frac{3}{5} \left(\frac{4}{3}x^2 - \frac{4}{3}x^3 + x^4 - \frac{4}{15}x^5 \right) \Big|_0^1 = \\
 &= \frac{3}{5} \left(\frac{4}{3} - \frac{4}{3} + 1 - \frac{4}{15} \right) = \frac{11}{25}, \\
 E(Y) &= \int_{-\infty}^{+\infty} y \cdot p_Y(y) dy = \\
 &= \frac{3}{5} \int_0^1 y \left(\frac{1+3y^2}{5} \right) dy + \frac{3}{5} \int_1^2 y \left(\frac{8}{3} - 4y + 4y^2 - \frac{4}{3}y^3 \right) dy = \\
 &= \frac{3}{5} \int_0^1 \left(\frac{y+3y^3}{5} \right) dy + \frac{3}{5} \int_1^2 \left(\frac{8}{3}y - 4y^2 + 4y^3 - \frac{4}{3}y^4 \right) dy = \frac{3}{5} \left(\frac{1}{10}y^2 + \frac{3}{20}y^4 \right) \Big|_0^1 + \\
 &= \frac{3}{5} \int_0^1 \left(\frac{y+3y^3}{5} \right) dy + \frac{3}{5} \int_1^2 \left(\frac{8}{3}y - 4y^2 + 4y^3 - \frac{4}{3}y^4 \right) dy = \frac{3}{5} \left(\frac{1}{10}y^2 + \frac{3}{20}y^4 \right) \Big|_0^1 + \\
 &= \frac{3}{5} \left(\frac{4}{3}y^2 - \frac{4}{3}y^3 + y^4 - \frac{4}{15}y^5 \right) \Big|_1^2 = \frac{3}{5} \left(\frac{1}{10} + \frac{3}{20} \right) + \\
 &= \frac{3}{5} \left(\frac{4}{3} \cdot 4 - \frac{4}{3} \cdot 8 + 16 - \frac{4}{15} \cdot 32 \right) - \frac{11}{25} = \frac{99}{100}.
 \end{aligned}$$

Бидејќи $E(XY) \neq E(X)E(Y)$ тогаш случајните променливи X и Y не се независни.

Нека X и Y се независни дискретни случајни променливи од тогаш за условниот закон на распределба на X при услов $\{Y = y_j\}$ се добива:

$$p_X(x_i | y_j) = \frac{p(x_i, y_j)}{p_Y(y_j)} = \frac{p_X(x_i) p_Y(y_j)}{p_Y(y_j)} = p_X(x_i)$$

за сите $x_i \in R_X$, односно условната распределба на X при услов $\{Y = y_j\}$ се совпаѓа со безусловната распределба на X .

На ист начин, за условниот закон на распределба на Y при услов $\{X = x_i\}$ се добива:

$$p_Y(y_j | x_i) = \frac{p(x_i, y_j)}{p_X(x_i)} = \frac{p_X(x_i) p_Y(y_j)}{p_X(x_i)} = p_Y(y_j),$$

за сите $y_j \in R_Y$, односно

условната распределба на Y при услов $\{X = x_i\}$ се совпаѓа со безусловната распределба на Y .

Ако пак, X и Y се независни непрекинати случајни, тогаш за условната густина на X при услов $\{Y = y\}$ се добива:

$$p_X(x | y) = \frac{p(x, y)}{p_Y(y)} = \frac{p_X(x) p_Y(y)}{p_Y(y)} = p_X(x)$$

одноно условната густина на X при услов $\{Y = y\}$ се совпаѓа со безусловната густина на X .

Исто, за условната густина на Y при услов $\{X = x\}$ се добива:

$$p_Y(y | x) = \frac{p(x, y)}{p_X(x)} = \frac{p_X(x) p_Y(y)}{p_X(x)} = p_Y(y)$$

одноно условната густина на Y при услов $\{X = x\}$ се совпаѓа со безусловната густина на Y .

8.7. КОВАРИЈАНСА И КОЕФИЦИЕНТ НА КОРЕЛАЦИЈА

Дефиниција 8.7.1. Нека (X, Y) е непрекинатата двостепензионална случајна променлива. Коваријанса на случајните променливи X и Y е бројот дефиниран со:

$$\text{cov}(X, Y) = E((X - E(X))(Y - E(Y)))$$

или

$$\text{cov}(X, Y) = E(XY) - E(X)E(Y).$$

Ако $\text{cov}(X, Y) = 0$ тогаш велиме дека случајните променливи X и Y се некорелирани, и притоа важи

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

Коваријансата покажува **како двете променливи се менуваат една во однос на друга во делот каде што се линеарно зависни**. Имено, ако коваријансата е позитивна тогаш двете променливи се правопрпорционални. Имено, ако коваријансата е негативна, тогаш двете променливи се обратнопрпорционални

Ако коваријанса е близу нула или нула, тогаш нема линеарна врска помеѓу променливите.

Забелешка 8.7.1. Две променливи што се независни се и некорелирани, но не важи обратното. Имено, постојат случајни променливи кои не се линеарно поврзани, односно се некорелирани, но можат да се независни.

Коефициентот на корелација е нормализирана вредност на коваријансата и тој не зависи од мерните единици на обележјата (случајните променливи) за кои проверуваме линеарна врска.

Дефиниција 8.7.2. Нека (X, Y) е непрекинатата двостепензионална случајна променлива. Коефициентот на корелација на случајните променливи X и Y кои имаат ненулти стандарни девијации е бројот дефиниран со:

$$\rho(X, Y) = \frac{\text{cov}(X, Y)}{\sigma_X \sigma_Y}.$$

За коефициентот на корелација важат следниве својства:

- Коефициентот на корелација може да има вредности од -1 до +1.
- Ако коефициентот на корелација има позитивни вредности, тогаш корелацијата помеѓу променливите е право пропорционална.
- Ако коефициентот на корелација има негативни вредности, тогаш врската е обратна или обратно пропорционална.
- Ако коефициентот на корелација е еднаков на 1 или -1, тогаш корелацијата помеѓу променливите е линеарно совршена. Колку повеќе апсолутната вредност на коефициентот на корелација е поблиску до 1, толку е посилна корелацијата помеѓу променливите. Спротивно, колку е коефициентот на корелација поблиску до 0, толку е послаба корелацијата помеѓу променливите.

Задача 8.7.3. Разгледуваме случаен експеримент на фрлање на една хомогена паричка двапати. Нека случајната променлива X и го означува бројот на добиени писма, а случајната променлива Y го означува бројот на добиени глави.

(а) Пронајдете ја коваријансата помеѓу X и Y .

(б) Пронајдете го коефициентот на корелација. Што можете да заклучите?

Решение. Да се одреди дали случајните променливи X и Y од Законот за распределба на двостепензионална случајна променлива (X, Y) е даден во следната табела:

$Y \backslash X$	0	1	2	$P(Y)$
0	0	0	0.25	0.25
1	0	0,5	0	0,5
2	0,25	0	0	0,25
$P(X)$	0,25	0,5	0,25	

Маргиналната распределба за X е:

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

Маргиналната распределба за Y е:

Y	0	1	2
$P(Y)$	0,25	0,5	0,25

а) Коваријанса на случајните променливи X и Y е:

$$\text{cov}(X, Y) = E(XY) - E(X)E(Y)$$

$$E(XY) = \sum_{x \in R_X} \sum_{y \in R_Y} xyP((X, Y) = (x, y))$$

$$E(XY) = 0 \cdot 0 \cdot 0 + 0 \cdot 1 \cdot 0 + 0 \cdot 2 \cdot 0,25 + 1 \cdot 0 \cdot 0 + 1 \cdot 1 \cdot 0,5 + 1 \cdot 2 \cdot 0 + 2 \cdot 0 \cdot 0,25 + 2 \cdot 1 \cdot 0 + 2 \cdot 2 \cdot 0 = 0,5;$$

$$E(X) = 0 \cdot 0,25 + 1 \cdot 0,5 + 2 \cdot 0,25 = 1;$$

$$E(Y) = 0 \cdot 0,25 + 1 \cdot 0,5 + 2 \cdot 0,25 = 1; \text{ и}$$

$$\text{cov}(X, Y) = 0,5 - 1 \cdot 1 = -0,5$$

б) Коефициент на корелација на случајните променливи X и Y кои имаат ненулни стандарни девијации е бројот дефиниран со:

$$\rho(X, Y) = \frac{\text{cov}(X, Y)}{\sigma_X \sigma_Y}.$$

$$E(X^2) = 0^2 \cdot 0,25 + 1^2 \cdot 0,5 + 2^2 \cdot 0,25 = 1,5;$$

$$E(Y^2) = 0^2 \cdot 0,25 + 1^2 \cdot 0,5 + 2^2 \cdot 0,25 = 1,5;$$

$$\sigma_X = E(X^2) - (E(X))^2 = 1,5 - 1^2 = 0,5;$$

$$\text{Var}(Y) = E(Y^2) - (E(Y))^2 = 1,5 - 1^2 = 0,5; \text{ и}$$

$$\rho(X, Y) = \frac{\text{cov}(X, Y)}{\sigma_X \sigma_Y} = \frac{-0,5}{0,5} = -1.$$

Бидејќи коефициентот на корелација е -1 овие две случајни променливи се совршено линеарно зависни и уште повеќе нивната зависност е обратно пропорционална.

9. ФУНКЦИИ ОД СЛУЧАЈНИ ПРОМЕНЛИВИ

9.1. Функции од една случајна променлива

Нека X е дадена случајна променлива со позната распределба и нека $f: \mathbb{R} \rightarrow \mathbb{R}$ е Борелова функција. Тогаш и функцијата $Y = f(X)$ е случајна променлива. Следно потребно е да се одреди нејзината распределба. Нејзината распределба зависи од распределбата на случајната променлива X .

Прво, нека X е дискретна случајна променлива зададена со множество вредности $R_X = \{x_1, x_2, \dots\}$ (тоа може да е конечно или преброиво) и веројатности $p_i = P\{X = x_i\}$, за $x_i \in R_X$.

Тогаш, случајната променлива Y прима вредности од $R_Y = \{y_1, y_2, \dots\}$, каде што $y_i = f(x_i)$, а веројатностите со кои Y ги прима одделните вредности се определуваат според тоа кој од следниве два случаја е исполнет:

- Ако f е инјективна функција, односно

$$x_i \neq x_j \Rightarrow f(x_i) \neq f(x_j),$$

тогаш за секој $y_i \in R_Y$, постои единствен $x_i \in R_X$, така што:

$$f(x_i) = y_i, \text{ па } P\{Y = y_i\} = P\{X = x_i\} = p_i, \quad i = 1, 2, \dots$$

- Нека f не е инјективна функција. Тогаш за $y_i \in R_Y$ може да постојат повеќе $x_{i_r} \in R_X$ кои со f се пресликуваат во y_i , односно

$$f(x_{i_1}) = y_i, f(x_{i_2}) = y_i, \dots, f(x_{i_k}) = y_i;$$

Оттука следува дека настанот $\{Y = y_i\}$ ќе се појави ако и само ако се појави еден од настаните $\{X = x_{i_r}\}$, $r = 1, \dots, k$; кои се дисјунктни помеѓу себе, па затоа:

$$P\{Y = y_i\} = \sum_{r=1}^k P\{X = x_{i_r}\} = \sum_{r=1}^k p_{i_r}.$$

Задача 9.1.1. Нека X е дискретна случајна променлива зададена со законот на распределба:

$$X: \begin{pmatrix} -2 & -1 & 1 & 2 \\ 0,4 & 0,2 & 0,1 & 0,3 \end{pmatrix}.$$

Да ги одредиме законите на распределба на следните случајни променливи:

а) $Y = 2X + 1$; б) $Y = X^2$; в) $Y = X^3 + X - 2$.

Решение. Во согласност со дискусијата погоре, имаме:

$$а) R_Y = \{2x+1 | x \in R_X\} = \{-3, -1, 3, 5\};$$

$$P(Y = -3) = P(2X + 1 = -3) = P(X = -2) = 0,4;$$

$$P(Y = -1) = P(2X + 1 = -1) = P(X = -1) = 0,2;$$

$$P(Y = 3) = P(2X + 1 = 3) = P(X = 1) = 0,1;$$

$$P(Y = 5) = P(2X + 1 = 5) = P(X = 2) = 0,3.$$

Оттука, законот на распределба на променливата $Y = 2X + 1$ е:

$$Y: \begin{pmatrix} -3 & -1 & 3 & 5 \\ 0,4 & 0,2 & 0,1 & 0,3 \end{pmatrix}.$$

$$б) R_Y = \{x^2 | x \in R_X\} = \{1, 4\},$$

$$P(Y = 1) = P(X^2 = 1) = P(X = -1) + P(X = 1) = 0,2 + 0,1 = 0,3,$$

$$P(Y = 4) = P(X^2 = 4) = P(X = -2) + P(X = 2) = 0,4 + 0,3 = 0,7.$$

Оттука, законот на распределба на променливата $Y = X^2$ е:

$$Y: \begin{pmatrix} 1 & 4 \\ 0,3 & 0,7 \end{pmatrix}.$$

$$в) R_Y = \{x^3 + x - 2 | x \in R_X\} = \{-12, -4, 0, 8\};$$

$$P(Y = -12) = P(X^3 + X - 2 = -12) = P(X = -2) = 0,4;$$

$$P(Y = -4) = P(X^3 + X - 2 = -4) = P(X = -1) = 0,2;$$

$$P(Y = 0) = P(X^3 + X - 2 = 0) = P(X = 1) = 0,1;$$

$$P(Y=8)=P(X^3+X-2=8)=P(X=2)=0,3.$$

Оттука, законот на распределба на променливата $Y = X^3 + X - 2$ е:

$$Y: \begin{pmatrix} -12 & -4 & 0 & 8 \\ 0,4 & 0,2 & 0,1 & 0,3 \end{pmatrix}.$$

Нека, сега, X е непрекината случајна променлива зададена со густина на веројатност $p_X(x)$.

Теорема 9.1. Нека X е непрекината случајна променлива која прима вредности од интервалот (a,b) (може $a = -\infty$ или $b = +\infty$ или и двете). Ако f е строго монотона функција и има непрекинат извод на (a,b) , тогаш густината на распределба на $Y = f(X)$ се определува со:

$$p_Y(x) = p_X(f^{-1}(x)) \left| (f^{-1})'(x) \right|$$

Доказ. Прво, нека f е строго монотонно растечка на (a,b) . Тогаш и f^{-1} е строго монотонно растечка на интервалот $(f(a), f(b))$, па $(f^{-1})'(x) > 0$, односно:

$$\left| (f^{-1})'(x) \right| = (f^{-1})'(x), \text{ за } x \in (f(a), f(b)).$$

Тогаш, функцијата на распределба на случајната променлива Y е:

$$F_Y(x) = P(Y < x) = P(f(X) < x) = P(X < f^{-1}(x)) = F_X(f^{-1}(x)).$$

Со диференцирање на последниот израз ќе ја добиеме густината на распределба на случајната променлива Y . Имено,

$$p_Y(x) = F_Y'(x) = p_X(f^{-1}(x)) (f^{-1})'(x) = p_X(f^{-1}(x)) \left| (f^{-1})'(x) \right|$$

при што во последното равенство е искористено дека изводот на $(f^{-1})'$ е позитивен на $(f(a), f(b))$.

Ако f е строго монотono опаѓачка на (a, b) тогаш и f^{-1} е строго монотono опаѓачка на $(f(b), f(a))$. Оттука,

$$(f^{-1})'(x) < 0, \text{ односно } \left| (f^{-1})'(x) \right| = -(f^{-1})'(x), \text{ за } x \in (f(b), f(a)).$$

Користејќи го претходното, за функцијата на распределба на Y добиваме:

$$F_Y(x) = P\{Y < x\} = P(f(X) < x) = P(X > f^{-1}(x)) = 1 - F_X(f^{-1}(x)).$$

Сега, со диференцирање за густината на распределба на Y се добива:

$$p_Y(x) = F_Y'(x) = -p_X(f^{-1}(x))(f^{-1})'(x) = p_X(f^{-1}(x)) \left| (f^{-1})'(x) \right|,$$

бидејќи изводот на f^{-1} е негативен на $(f(b), f(a))$.

Задача 9.1.2. Нека X е непрекината случајна променлива со функција на распределба $F_X(x)$ и густина на распределба $p_X(x)$. Нека $Y = aX + b$, каде што a и b се реални константи и $a \neq 0$. Да се одредат функцијата на распределба $F_Y(y)$ и густина на распределба $p_Y(y)$ на $Y = aX + b$.

Решение. а) Ако $a > 0$, тогаш:

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

Ако $a < 0$, тогаш:

$$\begin{aligned} F_Y(y) &= P(Y \leq y) = P(aX + b \leq y) = P(aX \leq y - b) = P\left(X \geq \frac{y-b}{a}\right) = \\ &= 1 - P\left(X < \frac{y-b}{a}\right) = 1 - F_X\left(\frac{y-b}{a}\right) + P\left(X = \frac{y-b}{a}\right) = \\ &= 1 - F_X\left(\frac{y-b}{a}\right) + 0 = 1 - F_X\left(\frac{y-b}{a}\right). \end{aligned}$$

б) Бидејќи густината на распределба $p_Y(y)$ на една непрекината случајна променлива е извод на функцијата на распределба $F_Y(y)$, ја диференцираме $F_Y(y)$ и добиваме:

Ако $a > 0$, тогаш:

$$p_Y(y) = (F_Y(y))' = \left(F_X\left(\frac{y-b}{a}\right) \right)' = F_X'\left(\frac{y-b}{a}\right) \left(\frac{y-b}{a}\right)' = \frac{1}{a} F_X'\left(\frac{y-b}{a}\right) = \frac{1}{a} p_X\left(\frac{y-b}{a}\right).$$

Ако $a < 0$, тогаш:

$$p_Y(y) = (F_Y(y))' = \left(1 - F_X\left(\frac{y-b}{a}\right) \right)' = - \left(F_X\left(\frac{y-b}{a}\right) \right)' = -F_X'\left(\frac{y-b}{a}\right) \left(\frac{y-b}{a}\right)' = -\frac{1}{a} F_X'\left(\frac{y-b}{a}\right) = -\frac{1}{a} p_X\left(\frac{y-b}{a}\right).$$

Оттука, густината на распределба на Y е:

$$p_Y(y) = \frac{1}{|a|} p_X\left(\frac{y-b}{a}\right).$$

Задача 9.1.3. Да се одреди густината на распределба на случајната променлива $Y = aX + b$, каде што X е рамномерна распределба на $(0,1)$.

Решение. Бидејќи X е рамномерна распределба на $(0,1)$,

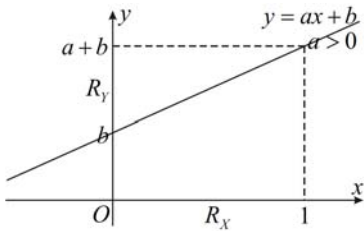
$$p_X(x) = \begin{cases} 1, & x \in (0,1) \\ 0, & \text{инаку} \end{cases}.$$

Од задачата 9.1.2., густината на распределба на $Y = aX + b$ е:

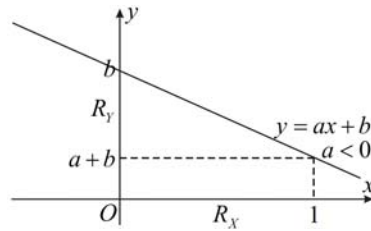
$$p_Y(y) = \frac{1}{|a|} p_X\left(\frac{y-b}{a}\right), \quad \frac{y-b}{a} \in R_X.$$

Оттука,

$$p_Y(y) = \begin{cases} \frac{1}{|a|} p_X\left(\frac{y-b}{a}\right), & \frac{y-b}{a} \in R_X \\ 0, & \text{инаку} \end{cases} = \begin{cases} \frac{1}{|a|} p_X\left(\frac{y-b}{a}\right), & y \in R_Y \\ 0, & \text{инаку} \end{cases}.$$



Цртеж 9.1.1.



Цртеж 9.1.2.

Во согласност со цртежите погоре: Ако $a > 0$, тогаш:

$$p_Y(y) = \begin{cases} \frac{1}{a} p_X\left(\frac{y-b}{a}\right), & y \in (b, a+b) \\ 0, & \text{инаку} \end{cases}.$$

Ако $a < 0$, тогаш:

$$p_Y(y) = \begin{cases} -\frac{1}{a} p_X\left(\frac{y-b}{a}\right), & y \in (a+b, b) \\ 0, & \text{инаку} \end{cases}.$$

Задача 9.1.4. Да се одреди густината на распределба на случајната променлива $Y = X^2$, каде што X е непрекината случајна променлива.

Решение. Настанот $(Y \leq y)$ во R_Y е еквивалентен со настанот $(X^2 \leq y) = (-\sqrt{y} \leq X \leq \sqrt{y})$ во R_X . Оттука, ако $y \leq 0$ тогаш $F_Y(y) = P(Y \leq y) = 0$ и ако $y > 0$ тогаш функцијата на распределба на $Y = X^2$ е:

$$F_Y(y) = P(Y \leq y) = P(-\sqrt{y} \leq X \leq \sqrt{y}) = F_X(\sqrt{y}) - F_X(-\sqrt{y}).$$

Со нејзино диференцирање, ја добиваме густината на распределба на $Y = X^2$. Ако $y \leq 0$, $p_Y(y) = 0$ и ако $y > 0$:

$$\begin{aligned} p_Y(y) &= (F_Y(y))' = (F_X(\sqrt{y}) - F_X(-\sqrt{y}))' = (F_X(\sqrt{y}))' - (F_X(-\sqrt{y}))' = \\ &= F_X'(\sqrt{y})(\sqrt{y})' - F_X'(-\sqrt{y})(-\sqrt{y})' = \frac{1}{2\sqrt{y}} p_X(\sqrt{y}) - \left(-\frac{1}{2\sqrt{y}}\right) p_X(-\sqrt{y}) = \end{aligned}$$

$$\frac{1}{2\sqrt{y}} \left(p_X(\sqrt{y}) + p_X(-\sqrt{y}) \right),$$

или,

$$p_Y(y) = \begin{cases} \frac{1}{2\sqrt{y}} \left(p_X(\sqrt{y}) + p_X(-\sqrt{y}) \right), & y \geq 0 \\ 0, & y < 0 \end{cases}.$$

Задача 9.1.5. Да се одреди густината на распределба на случајната променлива $Y = X^2$, каде што X е рамномерна распределба на $(-1, 2)$.

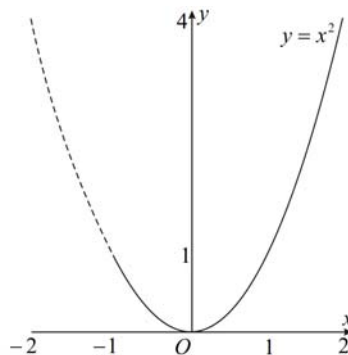
Решение. Бидејќи X е рамномерна распределба на $(-1, 2)$:

$$p_X(x) = \begin{cases} \frac{1}{2 - (-1)}, & x \in (-1, 2) \\ 0, & \text{инаку} \end{cases} = \begin{cases} \frac{1}{3}, & x \in (-1, 2) \\ 0, & \text{инаку} \end{cases}.$$

Од задачата 9.1.4, густината на распределба на случајната променлива $Y = X^2$ е:

$$p_Y(y) = \begin{cases} \frac{1}{2\sqrt{y}} \left(p_X(\sqrt{y}) + p_X(-\sqrt{y}) \right), & y \in R_Y \\ 0, & y < 0 \end{cases}.$$

Бидејќи $R_X = (-1, 2)$ и $Y = X^2$, од цртежот подолу имаме $R_Y = (0, 4)$



Цртеж 9.1.3.

и ако $y \in (0, 1)$ тогаш $\sqrt{y}, -\sqrt{y} \in (-1, 1) \subseteq (-1, 2)$, затоа:

$$p_X(\sqrt{y}) = \frac{1}{3}, \quad p_X(-\sqrt{y}) = \frac{1}{3};$$

ако $y \in (1, 4)$ тогаш:

$$\sqrt{y} \in (1, 2) \subseteq (-1, 2), \quad -\sqrt{y} \in (-2, -1) \not\subseteq (-1, 2),$$

па

$$p_X(\sqrt{y}) = \frac{1}{3}, \quad p_X(-\sqrt{y}) = 0.$$

Оттука, густината на распределба на случајната променлива $Y = X^2$ е:

$$p_Y(y) = \begin{cases} \frac{1}{2\sqrt{y}} \left(\frac{1}{3} + \frac{1}{3} \right), & y \in (0, 1) \\ \frac{1}{2\sqrt{y}} \left(\frac{1}{3} + 0 \right), & y \in (1, 4) \\ 0, & \text{инаку} \end{cases} = \begin{cases} \frac{1}{2\sqrt{y}} \cdot \frac{2}{3}, & y \in (0, 1) \\ \frac{1}{2\sqrt{y}} \cdot \frac{1}{3}, & y \in (1, 4) \\ 0, & \text{инаку} \end{cases}$$

$$\begin{cases} \frac{1}{3\sqrt{y}}, & y \in (0, 1) \\ \frac{1}{6\sqrt{y}}, & y \in (1, 4) \\ 0, & \text{инаку} \end{cases}$$

Задача 9.1.6. Да се одреди густината на распределба на случајната променлива $Y = e^X$, каде што X е рамномерна распределба на $(0, 1)$.

Решение. Бидејќи X е рамномерна распределба на $(0, 1)$ нејзината густина на распределба е:

$$p_X(x) = \begin{cases} 1, & x \in (0, 1) \\ 0, & \text{инаку} \end{cases} = \begin{cases} 1, & x \in (0, 1) \\ 0, & \text{инаку} \end{cases}.$$

Тогаш функцијата на распределба на $Y = e^X$ е:

$$F_Y(y) = P(Y \leq y) = P(e^X \leq y) = P(X \leq \ln y) = \int_{-\infty}^{\ln y} p_X(x) dx.$$

Ако $0 < \ln y < 1$, односно $1 < y < e$ тогаш:

$$F_Y(y) = \int_{-\infty}^{\ln y} p_X(x) dx = \int_{-\infty}^0 0 dx + \int_0^{\ln y} 1 dx = x \Big|_0^{\ln y} = \ln y$$

Затоа, $p_Y(y) = (F_Y(y))' = (\ln y)' = \frac{1}{y}$ ако $1 < y < e$, инаку

$p_Y(y) = 0$ или:

$$p_Y(y) = \begin{cases} \frac{1}{y}, & y \in (1, e) \\ 0, & \text{инаку} \end{cases}.$$

9.2. ФУНКЦИИ ОД ДВЕ СЛУЧАЈНИ ПРОМЕНЛИВИ

Нека (X, Y) е дводимензионална случајна променлива и нека $f: \mathbb{R}^2 \rightarrow \mathbb{R}$. Тогаш $Z = f(X, Y)$ е случајна променлива. Велиме дека случајната променлива Z е функција од дводимензионална случајна променлива, односно функција од две случајни променливи X и Y .

Цел е да се определи распределбата на случајната променлива $Z = f(X, Y)$. Оваа распределба зависи од видот на распределба на дводимензионална случајна променлива (X, Y) .

Најпрво, нека (X, Y) е дискретна дводимензионална случајна променлива зададена со нејзиниот закон на распределба

$$p_{ij} = P\{X = x_i, Y = y_j\}, (x_i, y_j) \in R_{(X, Y)}.$$

За законот на распределба на Z се добива следново:

$$P\{Z = z\} = P\{f(X, Y) = z\} = \sum_{\substack{(x_i, y_j) \in R_{(X, Y)} \\ f(x_i, y_j) = z}} P\{X = x_i, Y = y_j\} = \sum_{\substack{(x_i, y_j) \in R_{(X, Y)} \\ f(x_i, y_j) = z}} p_{ij}.$$

Задача 9.2.1. Нека X и Y се две дискретни случајни променливи, чиешто закони за распределба се зададени со следниве табели:

$$X: \begin{pmatrix} -2 & -1 & 1 & 2 \\ 0,4 & 0,2 & 0,1 & 0,3 \end{pmatrix} \text{ и } Y: \begin{pmatrix} 0 & 2 & 4 \\ 0,3 & 0,2 & 0,5 \end{pmatrix}.$$

Да се одредат законите за распределба на случајната променлива $Z = XY$.

Решение. Бидејќи:

$$R_Z = \{xy \mid x \in R_X, y \in R_Y\} = \{-8, -4, -2, 0, 2, 4, 8\};$$

$$P(Z = -8) = P(XY = -8) =$$

$$P(X = -2)P(Y = 4) = 0,4 \cdot 0,5 = 0,2;$$

$$P(Z = -4) = P(XY = -4) =$$

$$P(X = -2)P(Y = 2) + P(X = -1)P(Y = 4) = \\ 0,4 \cdot 0,2 + 0,2 \cdot 0,5 = 0,18;$$

$$P(Z = -2) = P(XY = -2) =$$

$$P(X = -1)P(Y = 2) = 0,2 \cdot 0,2 = 0,04;$$

$$P(Z = 0) = P(XY = 0) =$$

$$P(X = -2)P(Y = 0) + P(X = -1)P(Y = 0) +$$

$$P(X = 1)P(Y = 0) + P(X = 2)P(Y = 0) =$$

$$0,4 \cdot 0,3 + 0,2 \cdot 0,3 + 0,1 \cdot 0,3 + 0,3 \cdot 0,3 = 0,3;$$

$$P(Z = 2) = P(XY = 2) = P(X = 1)P(Y = 2) = 0,1 \cdot 0,2 = 0,02;$$

$$P(Z = 4) = P(XY = 4) = P(X = 1)P(Y = 4) + P(X = 2)P(Y = 2) = \\ 0,1 \cdot 0,5 + 0,3 \cdot 0,2 = 0,11;$$

$$P(Z = 4) = P(XY = 4) = P(X = 1)P(Y = 4) + P(X = 2)P(Y = 2) = \\ 0,1 \cdot 0,5 + 0,3 \cdot 0,2 = 0,11;$$

$$P(Z = 8) = P(XY = 8) = P(X = 2)P(Y = 4) = 0,3 \cdot 0,5 = 0,15.$$

Оттука, законот за распределба на $Z = X + 2Y$ е:

$$Z: \begin{pmatrix} -2 & -1 & 1 & 2 & 3 & 5 & 6 & 7 & 9 & 10 \\ 0,12 & 0,06 & 0,03 & 0,17 & 0,04 & 0,02 & 0,26 & 0,1 & 0,05 & 0,15 \end{pmatrix}.$$

Нека (X, Y) е непрекината дводимензионална случајна променлива зададена со густина на веројатност $p(x, y)$ и нека $Z = f(X, Y)$. Густината на веројатност Z се определува вака.

Прво, се воведува помошна случајна променлива U која наједноставно е да биде идентична случајна променлива. Значи, $U = X$ или $U = Y$. Без губење на општоста, нека $U = X$. На тој начин, добиваме систем од две случајни променливи Z и U што се функции од случајните променливи X и Y . Во теоријата на обични променливи, трансформацијата со која од (X, Y) се добива векторот (Z, U) е зададена со системот равенства:

$$\begin{cases} z = f(x, y) \\ u = x \end{cases} \quad (8.1).$$

Инверзната трансформација на претходно дадената се добива ако x и y се изразат преку z и u , се добива:

$$\begin{cases} x = g_1(z, u) \\ y = g_2(z, u) \end{cases} \quad (8.2),$$

за некои функции g_1 и g_2 од $\mathbb{R}^2$ во $\mathbb{R}$.

Нека $S_1 \subseteq \mathbb{R}^2$ со трансформацијата зададена со 8.1 се пресликува во множество $S_2 \subseteq \mathbb{R}^2$, односно:

$$S_2 = \{(z, u) \mid z = f(x, y), u = x, (x, y) \in S_1\}.$$

Јакобијанот на инверзната трансформација 8.2 е:

$$J = \begin{vmatrix} \frac{\partial x}{\partial z} & \frac{\partial x}{\partial u} \\ \frac{\partial y}{\partial z} & \frac{\partial y}{\partial u} \end{vmatrix}.$$

Сега, во согласност со правилото за смена во двоен интеграл, имаме:

$$P\{(Z, U) \in S_2\} = P\{(X, Y) \in S_1\} = \iint_{S_1} p(x, y) dx dy =$$

$$\iint_{S_2} p(g_1(z, u), g_2(z, u)) |J| dz du .$$

Оттука, густината на веројатност на дводимензионаланта случајна променлива (Z, U) е:

$$p_{ZU}(z, u) = p(g_1(z, u), g_2(z, u)) |J| .$$

На крај, распределбата на случајната променлива Z се наоѓа како маргинална распределба од распределбата на случајниот вектор (Z, U) .

Задача 9.2.2. Нека X и Y се случајни променливи со густина на распределба $f_{XY}(x, y)$. Да се одреди густината на распределба на $Z = XY$.

Решение. Нека $Z = XY$ и $W = X$. Трансформацијата $z = xy$, $w = x$ има инверзна трансформација $x = w$, $y = \frac{z}{w}$, и

$$J(x, y) = \begin{vmatrix} \frac{\partial z}{\partial x} & \frac{\partial z}{\partial y} \\ \frac{\partial w}{\partial x} & \frac{\partial w}{\partial y} \end{vmatrix} = \begin{vmatrix} y & x \\ 1 & 0 \end{vmatrix} = -x$$

Заедничката густина на случајните променливи Z и W е дадена со:

$$f_{ZW}(z, w) = f_{XY}(x, y) |J(x, y)|^{-1} = \frac{1}{|x|} f_{XY}(x, y) = \frac{1}{|w|} f_{XY}\left(w, \frac{z}{w}\right)$$

Густината на распределба на Z е маргинална од Заедничката густина на случајните променливи Z и W :

$$f_Z(z) = \int_{-\infty}^{+\infty} f_{ZW}(z, w) dw = \int_{-\infty}^{+\infty} \frac{1}{|w|} f_{XY}\left(w, \frac{z}{w}\right) dw .$$

Задача 9.2.3. Нека X и Y се независни рамномерно распределени случајни променливи на $(0, 1)$. Одреди густина на распределба на $Z = XY$.

Решение. Бидејќи X и Y се независни рамномерно распределени случајни променливи на $(0, 1)$:

$$f_{XY}(x, y) = \begin{cases} 1, & 0 < x < 1, 0 < y < 1 \\ 0, & \text{инаку} \end{cases}.$$

Од задачата 9.2.2. густината на распределба на случајната променлива е:

$$f_Z(z) = \int_{-\infty}^{+\infty} f_{ZW}(z, w) dw = \int_{-\infty}^{+\infty} \frac{1}{|w|} f_{XY}\left(w, \frac{z}{w}\right) dw.$$

Таа е дефинирана на интервалот $(0, 1)$. Тогаш:

$$f_{XY}\left(w, \frac{z}{w}\right) = \begin{cases} 1, & 0 < w < 1, 0 < \frac{z}{w} < 1 \\ 0, & \text{инаку} \end{cases} = \begin{cases} 1, & 0 < z < w < 1 \\ 0, & \text{инаку} \end{cases},$$

па,

$$f_Z(z) = \int_{-\infty}^{+\infty} \frac{1}{|w|} f_{XY}\left(w, \frac{z}{w}\right) dw = \int_z^1 \frac{1}{w} dw = \ln w \Big|_z^1 = -\ln z, \text{ ако } 0 < z < 1.$$

Задача 9.2.4. Нека X и Y се случајни променливи со густина на распределба $f_{XY}(x, y)$. Да се одреди густината на распределба на $Z = \frac{X}{Y}$.

Решение. Нека $Z = \frac{X}{Y}$ и $W = Y$. За трансформацијата $z = \frac{x}{y}$

и $w = y$ е точно:

$$J(x, y) = \begin{vmatrix} \frac{\partial z}{\partial x} & \frac{\partial z}{\partial y} \\ \frac{\partial w}{\partial x} & \frac{\partial w}{\partial y} \end{vmatrix} = \begin{vmatrix} \frac{1}{y} & -\frac{x}{y^2} \\ 0 & 1 \end{vmatrix} = \frac{1}{y}.$$

Тогаш, заедничката густина на случајните променливи Z и W е дадена со:

$$f_{ZW}(z, w) = f_{XY}(x, y) |J(x, y)|^{-1} = |y| f_{XY}(x, y) = |w| f_{XY}(zw, w).$$

Густината на распределба на Z може да се најде како маргинална на заедничката густина на случајните променливи Z и W :

$$f_Z(z) = \int_{-\infty}^{+\infty} f_{ZW}(z, w) dw = \int_{-\infty}^{+\infty} |w| f_{XY}(zw, w) dw.$$

Задача 9.2.5. Нека X и Y се независни случајни променливи со нормална распределба. Да се одреди густината на распределба на $Z = \frac{X}{Y}$.

Решение. Нека X и Y се независни случајни променливи со нормална распределба, тогаш густината на распределба на (X, Y) е:

$$f_{XY}(x, y) = \frac{1}{\sqrt{2\pi}} e^{-\frac{x^2}{2}} \frac{1}{\sqrt{2\pi}} e^{-\frac{y^2}{2}} = \frac{1}{2\pi} e^{-\frac{(x^2+y^2)}{2}}$$

Од задачата 9.2.4:

$$\begin{aligned} f_Z(z) &= \int_{-\infty}^{+\infty} f_{ZW}(z, w) dw = \int_{-\infty}^{+\infty} |w| f_{XY}(zw, w) dw = \frac{1}{2\pi} \int_{-\infty}^{+\infty} |w| e^{-\frac{(zw)^2 + w^2}{2}} dw = \\ &= \frac{1}{2\pi} \int_{-\infty}^{+\infty} |w| e^{-\frac{(z^2+1)w^2}{2}} dw = -\frac{1}{2\pi} \int_{-\infty}^0 w e^{-\frac{(z^2+1)w^2}{2}} dw + \frac{1}{2\pi} \int_0^{+\infty} w e^{-\frac{(z^2+1)w^2}{2}} dw = \\ &= \left(\begin{array}{l} t = \frac{(z^2+1)w^2}{2} \quad dt = w(z^2+1)dw \\ \frac{dt}{(z^2+1)} = wdw \end{array} \quad \begin{array}{l} w=0 \Rightarrow t=0 \\ w=\pm\infty \Rightarrow t=\pm\infty \end{array} \right) = \\ &= -\frac{1}{2(z^2+1)\pi} \int_{+\infty}^0 e^{-t} dt + \frac{1}{2(z^2+1)\pi} \int_0^{+\infty} e^{-t} dt = \frac{1}{(z^2+1)\pi} \int_0^{+\infty} e^{-t} dt = \\ &= -\frac{1}{(z^2+1)\pi} e^{-t} \Big|_0^{+\infty} = -\left(\frac{1}{(z^2+1)\pi} \lim_{t \rightarrow +\infty} e^{-t} - \frac{e^{-0}}{(z^2+1)\pi} \right) = \\ &= -\left(0 - \frac{1}{(z^2+1)\pi} \right) = \frac{1}{(z^2+1)\pi}, \text{ за сите } z \in (-\infty, +\infty). \end{aligned}$$

10. ЗАКОН НА ГОЛЕМИТЕ БРОЕВИ. ЦЕНТРАЛНА ГРАНИЧНА ТЕОРЕМА

10.1. Закон на големите броеви

Пред да преминеме на законот на големите броеви ќе докажеме две теореми во кои се дадени две значајни неравенства во теоријата на веројатност и статистиката.

Теорема 10.1.1. (Основно неравенство на Чебишев).

Нека X е ненегативна случајна променлива. Тогаш за кој било позитивен реален број $a > 0$ важи неравенството:

$$P\{X \geq a\} \leq \frac{EX}{a}$$

Доказ: Прво, нека X е ненегативна случајна променлива од дискретен тип со множество вредности R_X и со веројатности:

$$p_n = P\{X = x_n\}, x_n \in R_X.$$

Од ненегативноста на X следува дека $x_n \geq 0$, за секој n . Тогаш,

$$EX = \sum_{x_n \in R_X} x_n p_n \geq \sum_{x_n, x_n \geq a} x_n p_n \geq \sum_{x_n, x_n \geq a} a p_n = a \sum_{x_n, x_n \geq a} p_n = a P\{X \geq a\}.$$

Ако, пак, X е ненегативна случајна променлива од апсолутно-непрекинат тип зададена со густина $p(x)$, тогаш се добива:

$$EX = \int_0^{+\infty} xp(x)dx \geq \int_a^{+\infty} xp(x)dx \geq \int_a^{+\infty} ap(x)dx = a \int_a^{+\infty} p(x)dx = a P\{X \geq a\}$$

Теорема 10.1.2. (Класично неравенство на Чебишев).

За која било случајна променлива X со конечна дисперзија DX и за кој било реален број $\varepsilon > 0$, точно е следново неравенство:

$$P\{|X - EX| \geq \varepsilon\} \leq \frac{DX}{\varepsilon^2}.$$

Доказ: Во согласност со основното неравенство на Чебишев за ненегативна случајна променлива, добиваме:

$$P\{|X - EX| \geq \varepsilon\} = P\{(X - EX)^2 \geq \varepsilon^2\} \leq \frac{E(X - EX)^2}{\varepsilon^2} = \frac{DX}{\varepsilon^2}.$$

Нека $X_1, X_2, \dots$ е низа независни случајни променливи. Се разгледува низата аритметички средини:

$$\bar{X}_1, \bar{X}_2 = \frac{X_1 + X_2}{2}, \bar{X}_3 = \frac{X_1 + X_2 + X_3}{3}, \dots, \bar{X}_n = \frac{1}{n} \sum_{i=1}^n X_i, \dots$$

Теорема 10.1.3. Нека $X_1, X_2, \dots$ е низа независни случајни променливи со еднаква распределба на веројатности со математичко очекување μ и дисперзија σ^2 . Тогаш:

$$E\bar{X}_n = \mu \text{ и } D\bar{X}_n = \frac{\sigma^2}{n}.$$

Доказ. Во согласност со теорема 4.4.7:

$$E\bar{X}_n = E\left(\frac{1}{n} \sum_{i=1}^n \bar{X}_n\right) = \frac{1}{n} E \sum_{i=1}^n \bar{X}_n = \frac{1}{n} \sum_{i=1}^n E\bar{X}_n = \frac{1}{n} \sum_{i=1}^n \mu = \frac{1}{n} \cdot n\mu = \mu.$$

Во согласност со теоремите 4.3.4 и 4.3.5 од [34]:

$$D\bar{X}_n = D\left(\frac{1}{n} \sum_{i=1}^n \bar{X}_n\right) = \frac{1}{n^2} D \sum_{i=1}^n \bar{X}_n = \frac{1}{n^2} \sum_{i=1}^n D\bar{X}_n = \frac{1}{n^2} \sum_{i=1}^n \sigma^2 = \frac{1}{n^2} \cdot n\sigma^2 = \frac{\sigma^2}{n}$$

Дефиниција 10.1.4. Нека $X_1, X_2, \dots$ е низа од случајни променливи и L е некој реален број. Велиме дека дадената низа случајни променливи конвергира по веројатност кон бројот L ако за секој реален број $\varepsilon > 0$,

$$\lim_{n \rightarrow \infty} P\{|X_n - L| < \varepsilon\} = 1.$$

Конвергенцијата по веројатност ја означуваме со:

$$X_n \xrightarrow{P} L, n \rightarrow \infty$$

и читаме X_n конвергира кон L по веројатност.

Нека $X_1, X_2, \dots$ е низа од независни случајни променливи. Се проверува дали оваа низа случајни променливи конвергира по веројатност кон бројот $\frac{1}{n} \sum_{i=1}^n EX_i$. Имено, се проверува дали:

$$\frac{1}{n} \sum_{i=1}^n X_i \xrightarrow{P} \frac{1}{n} \sum_{i=1}^n EX_i.$$

Ако тоа е точно, тогаш велиме дека за низата независни случајни променливи $X_1, X_2, \dots$ важи слабиот закон на големите броеви.

Теорема 10.1.5. (Теорема на Чебишев). Нека $X_1, X_2, \dots$ се независни случајни променливи кои имаат конечни дисперзии ограничени од горе со некоја константа c . Тогаш за секој реален број $\varepsilon > 0$ важи слабиот закон на големите броеви, односно

$$\lim_{n \rightarrow +\infty} P \left\{ \left| \frac{1}{n} \sum_{i=1}^n X_i - \frac{1}{n} \sum_{i=1}^n EX_i \right| < \varepsilon \right\} = 1$$

Доказ. Нека $Y_n = \frac{1}{n} \sum_{i=1}^n X_i$. Тогаш:

$$EY_n = \frac{1}{n} \sum_{i=1}^n EX_i$$

а од независноста на $X_i, i=1, 2, \dots$ и ограниченоста од горе на нивните дисперзии со константата c , следува дека:

$$DY_n = D \left(\frac{1}{n} \sum_{i=1}^n X_i \right) = \frac{1}{n^2} \sum_{i=1}^n DX_i \leq \frac{1}{n^2} \sum_{i=1}^n c = \frac{1}{n^2} \cdot nc = \frac{c}{n}$$

Согласно Теорема 10.1.2:

$$P \left\{ \left| \frac{1}{n} \sum_{i=1}^n X_i - \frac{1}{n} \sum_{i=1}^n EX_i \right| < \varepsilon \right\} = P \{ |Y_n - EY_n| < \varepsilon \} = 1 - P \{ |Y_n - EY_n| \geq \varepsilon \} \geq 1 - \frac{DY_n}{\varepsilon^2} \geq 1 - \frac{c}{n\varepsilon^2} \rightarrow 1, n \rightarrow \infty$$

Значи,

$$\lim_{n \rightarrow +\infty} P \left\{ \left| \frac{1}{n} \sum_{i=1}^n X_i - \frac{1}{n} \sum_{i=1}^n EX_i \right| < \varepsilon \right\} = 1.$$

Теорема 10.1.6. Ако $X_1, X_2, \dots$ се независни и еднакво распределени случајни променливи со математичко очекување

$$EX_i = \mu \text{ и } DX_i = \sigma^2, i = 1, 2, \dots,$$

тогаш за нив важи слабиот закон на големите броеви, односно

$$\lim_{n \rightarrow +\infty} P \left\{ \left| \frac{1}{n} \sum_{i=1}^n X_i - \mu \right| < \varepsilon \right\} = 1.$$

Доказ. Во согласност со теорема 10.1.3:

$$\frac{1}{n} \sum_{i=1}^n EX_i = \frac{1}{n} \sum_{i=1}^n \mu = \frac{1}{n} n\mu = \mu$$

па тврдењето следува директно од теоремата 10.1.5.

Последната теорема има големо значење во статистиката. Имено, често пати кога не е позната распределбата на некоја случајна променлива, односно некое квантитативно обележје на дадена популација, според претходната теорема $\frac{1}{n} \sum_{i=1}^n X_i \rightarrow \mu = EX$, односно оценка за очекуваната вредност на квантитативното обележје може да се земе вредноста пресметана со $m = \frac{1}{n} \sum_{i=1}^n x_i$, каде што x_i се измерените вредности на квантитативното обележје. Да забележиме дека колку n е поголемо, таа оценка ќе биде поточна.

Задача 10.1.7. Случајната променлива X има математичко очекување 1 и дисперзија 0,04. Со помош на неравенството на Чебишев да се одреди веројатноста да важи неравенството: $0,5 < X < 1,5$.

Решение. Од неравенството на Чебишев:

$$P(|X - E(X)| \geq \varepsilon) \geq \frac{DX}{\varepsilon^2},$$

оценувајќи го спротивниот настан од бараниот,

$$P(|X - E(X)| < \varepsilon) > 1 - \frac{DX}{\varepsilon^2}$$

добиваме:

$$\begin{aligned} P(0,5 < X < 1,5) &= P(|X - E(X)| < 0,5) > \\ 1 - \frac{DX}{0,5^2} &= 1 - \frac{0,04}{0,25} = 1 - 0,16 = 0,84. \end{aligned}$$

Задача 10.1.8. Веројатноста за изработка на квалитетни предмети изнесува 0,9. Изработени се 18000 предмети. Колкава е веројатноста: бројот на квалитетни производи да се разликува по апсолутна вредност од математичкото очекување за помалку од 200 предмети.

Решение. Случајната променлива има биномна распределба. Притоа: $p = 0,9$, од каде што $q = 0,1$ и дисперзијата е:

$$DX = npq = 18000 \cdot 0,9 \cdot 0,1 = 1620.$$

Земајќи го спротивниот настан од настанот во неравенството на Чебишев, добиваме:

$$P(|X - E(X)| < 200) > 1 - \frac{1620}{200^2} = 1 - \frac{1620}{40000} = 1 - \frac{405}{10000} = 1 - 0,0405 = 0,9595.$$

Задача 10.1.9. Дисперзијата на секоја од 800 независни случајни променливи, не е поголема од 9. Колкава треба да биде горната граница на апсолутната вредност од отстапувањата на аритметичката средина на случајните големини од аритметичката средина на нивните математички очекувања, за веројатноста на тоа отстапување да биде поголема од 0,977?

Решение. Според теоремата на Чебишев:

$$P\left(\left|\frac{1}{n} \sum_{i=1}^n X_i - \frac{1}{n} \sum_{i=1}^n MX_i\right| < \varepsilon\right) > 1 - \delta,$$

добиваме:

$$P\left(\left|\frac{1}{800} \sum_{i=1}^{800} X_i - \frac{1}{800} \sum_{i=1}^{800} MX_i\right| < \varepsilon\right) > 1 - \frac{9}{800\varepsilon^2} > 0,977.$$

Притоа:

$$1 - \frac{9}{800\varepsilon^2} > 0,977 \Leftrightarrow 0,023 > \frac{9}{800\varepsilon^2} \Leftrightarrow \varepsilon^2 > \frac{9}{800 \cdot 0,023} \approx 0,489 \Leftrightarrow \varepsilon > \approx 0,7.$$

10.2. ЦЕНТРАЛНА ГРАНИЧНА ТЕОРЕМА

Нека $X_1, X_2, \dots$ е низа независни случајни променливи со еднаква распределба на веројатности со математичко очекување μ и дисперзија σ^2 . Тогаш според Теоремата 10.1.3. нумеричките карактеристики на случајната променлива $\bar{X} = \frac{1}{n} \sum_{i=1}^n X_i$ изнесуваат:

$$E\bar{X}_n = \mu \text{ и } D\bar{X}_n = \frac{\sigma^2}{n},$$

додека централната гранична теорема овозможува да се одреди распределбата на веројатностите на случајната променлива

$$\bar{X} = \frac{1}{n} \sum_{i=1}^n X_i, \text{ за доволно големо } n.$$

Теорема 10.2.1. (Централна гранична теорема). Нека $X_1, X_2, \dots$ се независни и еднакво распределени случајни променливи со математичко очекување μ и дисперзија σ^2 . Тогаш

$$P\left\{\frac{\sum_{i=1}^n X_i - n\mu}{\sigma\sqrt{n}} \leq x\right\} \xrightarrow{P} \frac{1}{\sqrt{2\pi}} \int_{-\infty}^x e^{-t^2/2} dt = \Phi(x), n \rightarrow \infty$$

Доказ. Нека $Z_n = \sum_{i=1}^n X_i$. Тогаш:

$$\bar{Z}_n = \frac{\sum_{i=1}^n X_i - E\left(\sum_{i=1}^n X_i\right)}{\sqrt{D\left(\sum_{i=1}^n X_i\right)}} = \frac{\sum_{i=1}^n X_i - n\mu}{\sigma\sqrt{n}} = \frac{1}{\sqrt{n}} \sum_{i=1}^n \frac{X_i - \mu}{\sigma}.$$

Нека

$$Y_i = \frac{X_i - \mu}{\sigma}, \text{ за } i = 1, 2, \dots$$

Тогаш:

$$EY_i = E\left(\frac{X_i - \mu}{\sigma}\right) = \frac{EX_i - \mu}{\sigma} = 0, \quad DY_i = D\left(\frac{X_i - \mu}{\sigma}\right) = \frac{DX_i}{\sigma^2} = 1,$$

и Y_i се независни случајни променливи како функции од независните случајни променливи X_i . Нека $\varphi_{Y_i}(t) = E(e^{itY_i})$ е карактеристичната функција на Y_i , за секое $i = 1, 2, \dots, n$. Во согласност со теоремата која ја дава врската помеѓу изводите на карактеристичната функција и моментите $E(Y_i^k)$, $k = 1, 2, 3, \dots$ на случајната променлива Y_i , за секое $i = 1, 2, \dots, n$, односно $\varphi_{Y_i}^{(k)}(0) = i^k EY_i^k$, (види [39]) наоѓаме:

$$\varphi_{Y_i}'(0) = i \cdot EY_i = 0, \quad \varphi_{Y_i}''(0) = i^2 EY_i^2 = -DY_i = -1$$

Функцијата $\varphi_{Y_i}(t)$ за секое $i = 1, 2, \dots, n$, може да се развие во Маклоренов ред од следниов облик:

$$\varphi_{Y_i}(t) = \varphi_{Y_i}(0) + \varphi_{Y_i}'(0)t + \frac{\varphi_{Y_i}''(0)}{2!}t^2 + o(t^2) \text{ за секое } i = 1, 2, \dots, n,$$

па со замена на претходно добиените вредности на изводите, добиваме дека:

$$\varphi_{Y_i}(t) = 1 - \frac{t^2}{2} + o(t^2) \text{ за секое } i = 1, 2, \dots, n.$$

Сега, бидејќи $\bar{Z}_n$ може да се запише во обликот:

$$\bar{Z}_n = \frac{1}{\sqrt{n}} \sum_{i=1}^n \frac{X_i - \mu}{\sigma} = \frac{1}{\sqrt{n}} \sum_{i=1}^n Y_i = \sum_{i=1}^n \frac{Y_i}{\sqrt{n}},$$

па за карактеристичната функција на $\bar{Z}_n$ се добива:

$$\varphi_{\bar{Z}_n}(t) = E \left(e^{it \sum_{r=1}^n \frac{Y_r}{\sqrt{n}}} \right) = \prod_{r=1}^n e^{it \frac{Y_r}{\sqrt{n}}} = \prod_{r=1}^n \varphi_{Y_r}(t) = \left[1 - \frac{t^2}{2n} + o\left(\frac{t^2}{n}\right) \right]^n \rightarrow e^{-\frac{t^2}{2}}.$$

Значи, низата карактеристични функции $\{\varphi_{\bar{Z}_n}(t)\}$ конвергира кон функцијата $\varphi_N(t) = e^{-t^2/2}$ која е карактеристична функција на $N(0,1)$ распределба и $\varphi_N(t)$ е непрекината во $t = 0$. Според обратната гранична теорема од [40], следува дека низата функции на распределба $\{F_{\bar{Z}_n}(x)\}$ конвергира кон функција на нормална нормирана распределба, односно:

$$\{F_{\bar{Z}_n}(x)\} \rightarrow \Phi(x), n \rightarrow \infty$$

или

$$P\{\bar{Z}_n(x) \leq x\} \rightarrow \frac{1}{\sqrt{2\pi}} \int_{-\infty}^x e^{-t^2/2} dt, n \rightarrow \infty,$$

$$P\left\{\frac{\sum_{i=1}^n X_i - n\mu}{\sigma\sqrt{n}} \leq x\right\} = P\left\{\frac{1}{\sqrt{n}} \sum_{i=1}^n \frac{X_i - \mu}{\sigma} \leq x\right\} = P\{\bar{Z}_n(x) \leq x\} \rightarrow$$

$$\frac{1}{\sqrt{2\pi}} \int_{-\infty}^x e^{-t^2/2} dt, n \rightarrow \infty,$$

ОДНОСНО

$$P\left\{\frac{\sum_{i=1}^n X_i - n\mu}{\sigma\sqrt{n}} \leq x\right\} \xrightarrow{P} \frac{1}{\sqrt{2\pi}} \int_{-\infty}^x e^{-t^2/2} dt = \Phi(x), n \rightarrow \infty.$$

Во согласност со централната гранична теорема заклучуваме дека распределбата на нормирана сума од независни и еднакво распределени случајни променливи со произволна распределба се стреми кон $N(0,1)$ распределба, за доволно големо n , во пракса поголемо од 30.

Да забележиме дека ако $X_1, X_2, \dots$ е низа независни случајни променливи со еднаква нормална распределба на веројатности со математичко очекување μ и дисперзија σ^2 тогаш за секое n важи централната гранична теорема.

Централната гранична теорема има големо теориско и практично значење бидејќи дава теориско објаснување на фактот дека нормалната распределба е најчеста распределба на реалните големини, и во пракса посебно во статистиката овозможува распределбата на нормирана сума на случајни променливи, при дури и општи услови, да се апроксимира со нормална нормирана распределба.

Задача 10.2.1. Нека тежината на популација од студенти претставува нормална распределба со просечна тежина 70 kg, и стандардна девијација 8 kg. Да претпоставиме дека 16 студенти се во еден лифт, и дека максималната носивост на лифтот е 1200 kg.

Најдете ја веројатноста на настанот дека групата од 16 студенти ја надминува максималната носивост на лифтот.

Решение. Разгледуваме $n = 16$ независни случајни променливи со еднаква нормална распределба на веројатности со математичко очекување $\mu = 70$ и стандардна девијација $\sigma = 80$.

Треба да ја пресметаме веројатноста на настанот дека групата од 16 студенти ја надминува максималната носивост на лифтот, односно $P\left(\sum_{i=1}^{16} X_i > 1200\right)$.

Според централната гранична теорема:

$$\begin{aligned} P\left(\sum_{i=1}^{16} X_i > 1200\right) &= 1 - P\left(\sum_{i=1}^{16} X_i \leq 1200\right) = \\ 1 - P\left(\frac{\sum_{i=1}^{16} X_i - n\mu}{\sigma\sqrt{n}} \leq \frac{1200 - 16 \cdot 70}{8\sqrt{16}}\right) &= 1 - P\left(\frac{\sum_{i=1}^{16} X_i - n\mu}{\sigma\sqrt{n}} \leq 2,5\right) = \\ 1 - \Phi(2,5) &= 1 - 0,9938 = 0,0062. \end{aligned}$$

Значи, веројатноста на настанот дека групата од 16 студенти ја надминува максималната носивост на лифтот е 0,0062.

Задача 10.2.2. Размислете за производствена линија на даден механички елемент. Номиналната должина на елементот е 50 mm. Во производниот процес има стандардно отстапување од 4 mm. Точната распределба на должината на елементот не е позната. Примерок од 36 различни елементи е избран по случаен избор од многу голем број елементи произведени на оваа производствена линија, како податок за дневна проверка на квалитетот на производство (QA). Притоа, се мери должината на секој елемент од примерокот и доколку просекот од овие должини отстапува од номиналната вредност од 50mm за повеќе од 2 mm, тогаш производната линија веднаш се прекинува.

Го дефинираме настанот на „лажна тревога“, односно производствената линија е запрена поради „QA аларм“ кој се дефинира како неуспех во квалитетот на производството, иако должината на елементите од популацијата е во точната просечна вредност од 50 mm. Пронајдете ја веројатноста за настанот „лажна тревога“.

Решение. Разгледуваме $n = 36$ независни случајни променливи со еднаква нормална распределба на веројатности со математичко очекување $\mu = 50$ и стандардна девијација $\sigma = 4$.

Треба да ја пресметаме веројатноста на настанот „лажна тревога“, односно:

$$P\left(\left|\frac{\sum_{i=1}^{40} X_i}{40} - 50\right| > 2\right).$$

Според централната гранична теорема

$$P\left(\left|\frac{\sum_{i=1}^{40} X_i}{40} - 50\right| > 2\right) = 1 - P\left(\left|\frac{\sum_{i=1}^{40} X_i}{40} - 50\right| \leq 2\right).$$

Бидејќи,

$$\begin{aligned} P\left(\left|\frac{\sum_{i=1}^{40} X_i}{40} - 50\right| \leq 2\right) &= P\left(-2 \leq \frac{\sum_{i=1}^{40} X_i}{40} - 50 \leq 2\right) = P\left(48 \leq \frac{\sum_{i=1}^{40} X_i}{40} \leq 52\right) = \\ &= P\left(\frac{48-50}{\frac{4}{\sqrt{36}}} \leq \frac{\frac{\sum_{i=1}^{40} X_i}{40} - \mu}{\frac{\sigma}{\sqrt{36}}} \leq \frac{52-50}{\frac{4}{\sqrt{36}}}\right) = P\left(\frac{-2}{0,667} \leq \frac{\frac{\sum_{i=1}^{40} X_i}{40} - \mu}{\frac{\sigma}{\sqrt{36}}} \leq \frac{2}{0,667}\right) = \end{aligned}$$

$$\Phi(3) - \Phi(-3) = \Phi(3) - (1 - \Phi(3)) = 2\Phi(3) - 1,$$

$$P\left(\left|\frac{\sum_{i=1}^{40} X_i}{40} - 50\right| > 2\right) = 1 - (2\Phi(3) - 1) = 2 - 2\Phi(3) = 2 - 2 \cdot 0,998 = 0,004.$$

Значи, веројатноста на настанот „лажна тревога“ е 0,004.

Задача 10.2.3. Еден диск има слободен простор од 330 мегабајти. Има 300 независни слики и секоја слика има очекувана големина од 1 мегабајт со стандардна девијација од 0,5 мегабајти. Дефинирајте го настанот дека дискот е доволен за 300 независни слики. Дали е можно да се пресмета точната веројатност на настанот дека дискот е доволен за 300 независни слики? Објаснете дали „ДА“ или „НЕ“. Најдете добра апроксимација на веројатноста на настанот дека дискот е доволен за 300 независни слики. Определете ја употребата на централната гранична теорема и пресметајте ја веројатноста.

Решение. Нека $X_1, X_2, \dots, X_{300}$ се случајни променливи со кои се обележуваат големините на 300 слики. Тие се независни и еднакво распределени случајни променливи со математичко очекување $\mu = 1$ и стандардна девијација $\sigma = 0,5$. Вкупната големина на сликите изнесува $\sum_{i=1}^{300} X_i$.

Настанот дека дискот е доволен за сите 300 слики е кога вкупната големина на сликите не го надминува достапниот слободен простор на дискот, односно $\sum_{i=1}^{300} X_i \leq 330$.

Точната веројатност на настанот: „дискот е доволен за 300 независни слики“ не може да се пресмета бидејќи не се знае точната распределба на големините на сликите.

Бидејќи $n = 300$ е поголемо од 30 може да се примени централната гранична теорема и

$$P\left(\sum_{i=1}^{300} X_i \leq 330\right) = P\left(\frac{\sum_{i=1}^{300} X_i - n\mu}{\sigma\sqrt{n}} \leq \frac{330 - 300 \cdot 1}{0,5\sqrt{300}}\right) = \Phi(3,46) = 0,999.$$

Значи, веројатноста на настанот дека дискот е доволен за сите 300 слики е 0,999.

II Дел: СТАТИСТИКА

1. СТАТИСТИКА

Статистиката е научна дисциплина која се занимава со собирање, анализирање, толкување и презентирање податоци. Нејзината примена опфаќа широк спектар на области. Преку статистичките методи, ние можеме да разбереме скриени трендови, да предвидиме идни случувања и да донесуваме одлуки засновани на податоци.

Во 17-от век, статистиката почнува да се развива како посебна дисциплина благодарение на работата на научници како Џон Граунт и Вилијам Петти, кои го воведуваат концептот на демографски истражувања и економски статистики. Во текот на 18-от и 19-от век, статистиката се проширува со нови методи и апликации, особено во областа на веројатноста, благодарение на научниците како Пјер-Симон Лаплас и Карл Фридрих Гаус. Развојот на математичката статистика овозможува подлабоко разбирање на случајноста и несигурноста, што го поттикнува нејзиниот раст како важна научна алатка. Во 20-от и 21-от век, со развојот на компјутерската технологија, статистиката значително еволуира. Станува неразделен дел од истражувањата во различни дисциплини и добива примена во нови области како машинско учење, анализа на големи податоци и вештачка интелигенција. Денес, статистиката е клучна алатка во донесувањето одлуки базирани на податоци, како во приватниот сектор, така и во јавната администрација.

Статистиката е поделена на две главни гранки: **дескриптивна** (описна) статистика и **инференцијална** (изводна) статистика.

1. Дескриптивната статистика се занимава со собирање, организирање, сумирање и презентирање податоци. Таа ги претставува податоците на разбирлив начин, користејќи графички прикази (табели, графикони, хистограми, полигони, пати (кругови) и др.) и параметри како: аритметичка средина (просек), медијана, мода, ранг, варијанса и стандардна девијација. Описната статистика е корисна за добивање општ увид во податоците, но не овозможува донесување заклучоци или предвидувања.

2. Инференцијалната статистика оди чекор понатаму, користејќи примероци од податоци донесува заклучоци и предвидувања за целата популација врз основа на веројатноста и

моделирањето на податоците. Основни елементи на инференцијалната статистика се: хипотетички тестови (тестирање статистички хипотези за да се утврди дали постои разлика или однос меѓу податоците), интервали на доверба (мерење на степенот на сигурност дадена вредност да припаѓа на одреден интервал), регресија и корелација (анализа на односи и влијанија меѓу променливите) и веројатносни модели (користење веројатност за предвидување идни настани или трендови во податоците).

1.1. ПОПУЛАЦИЈА И ПРИМЕРОК

Множеството објекти кои имаат заедничка карактеристика, ознака Ω , се нарекува **популација**. Заедничката карактеристика што се набљудува се нарекува **обележје**. Обележјето X претставува случајна променлива, односно на секој $\omega \in \Omega$ придружуваме број $X(\omega)$.

Делот од популацијата на кој се извршуваат потребните испитувања се нарекува **примерок**. Значи примерокот е подмножество U од популацијата Ω .

Примерокот треба да биде репрезентативен, односно да претставува мини модел на популацијата.

Пример. Се разгледува популацијата: Резидентно население на Македонија на 5-ти септември 2021-ва година. Нејзиниот број е 1836713. За да го определиме параметарот просечната тежина на населението го рагледуваме обележјето: тежина на секој граѓанин. Испитувањето не мора да се изврши на целата популација, доволно е да се земат податоците од едно нејзино репрезентативно подмножество-примерок.

Примерокот треба да биде репрезентативен (секој елемент од популацијата да има еднакви шанси да биде избран во примерокот), доволно броен (најчесто се зема бројот да биде корен од бројот на популацијата, значи во нашиот случај 1350) и објективен (изборот да биде случаен). Тогаш просечната тежина на примерокот се зема за просечна тежина на популацијата.

Обележјето може да биде квантитавно и квалитативно. Квантитативно е кога тоа се карактеризира со некоја мерлива големина, додека квалитативно кога обележјето се однесува на некој атрибут на популацијата, пример боја на коса, пол, раса, пушач или не и слично.

1.2-5. ДЕСКРИПТИВНА СТАТИСТИКА

1.2. ПРЕТСТАВУВАЊЕ СТАТИСТИЧКИ ПОДАТОЦИ

Ако во една популација го разгледуваме обележјето X , тогаш од популацијата на случаен начин се избираат n елементи, и за секој од нив се регистрираат вредностите на обележјето $x_1, x_2, \dots, x_n$, соодветно. На тој начин се добива случаен примерок $x_1, x_2, \dots, x_n$ со големина n . Примерокот може да се разгледа како реализација на n димензионална случајна променлива $X_1, X_2, \dots, X_n$. Ако случајните променливи $X_1, X_2, \dots, X_n$ се независни, тогаш велиме дека примерокот е **прост**.

Податоците добиени со набљудување или експерименти, кои претставуваат вредности на примерокот, потребно е да се средат на соодветен начин да може да се испитува и определи распределбата на примерокот. Првенствено податоците се внесуваат во табела, и се пресметуваат параметрите потребни за истражувањето. Ако бројот на податоци е мал сите податоци се внесуваат во табела, додека ако е голем податоците се делат во групи (серији).

Нека набљудуваното обележје X прима k различни вредности, $x_1, x_2, \dots, x_k$, со апсолутни честоти (фреквенции) $f_1, f_2, \dots, f_k$, $f_1 + f_2 + \dots + f_k = n$, соодветно. Значи дека x_1 се јавува f_1 пати, x_2 - f_2 пати, ..., x_k - f_k пати.

Наместо апсолутни, често се користат релативни честоти (фреквенции):

$$f_{n1} = \frac{f_1}{n}, f_{n2} = \frac{f_2}{n}, \dots, f_{nk} = \frac{f_k}{n}; f_{n1} + f_{n2} + \dots + f_{nk} = 1.$$

Вредностите на примерокот земени со соодветните фреквенции се распоредуваат по растечки редослед и формираат статистичка табела.

X	x_1	x_2	...	x_k	Σ
f_i	f_1	f_2	...	f_k	n
f_{ni}	f_{n1}	f_{n2}	...	f_{nk}	1

Низата од подредени парови $(x_{n_i}, f_{n_i})_{i=1}^k$ се нарекува распределба на фреквенцијата на примерокот.

Откако ќе се одреди распределбата на фреквенцијата $(x_{n_i}, f_{n_i})_{i=1}^k$ на примерокот, се пресметува т.н. **емпириска функција на распределба** на примерокот $F_n(x): \mathbb{R} \rightarrow [0,1]$, дефинирана на следниот начин:

$$F_n(x) = \frac{n(x)}{n}, \text{ за секој } x \in \mathbb{R},$$

каде што со $n(x)$ е означен бројот на вредностите од примерокот помали или еднакви на x , познати како кумулативни честоти. Имено:

За $x < x_{n_1}$ важи $n(x) = 0$, затоа $F_n(x) = 0$;

За $x_{n_1} \leq x < x_{n_2}$ важи $n(x) = f_{n_1}$, затоа $F_n(x) = \frac{f_{n_1}}{n}$;

За $x_{n_2} \leq x < x_{n_3}$ важи $n(x) = f_{n_1} + f_{n_2}$, затоа $F_n(x) = \frac{f_{n_1} + f_{n_2}}{n}$;

.....

За $x_{n_r} \leq x < x_{n_{r+1}}$ важи $n(x) = \sum_{i=1}^r f_{n_i}$, затоа $F_n(x) = \sum_{i=1}^r \frac{f_{n_i}}{n}$;

.....

За $x > x_{n_k}$ важи $n(x) = \sum_{i=1}^k f_{n_i} = n$, затоа $F_n(x) = 1$.

Може да се забележи дека емпириската функција на распределба се добива со постепено собирање на релативните честоти (фреквенции) на вредностите од примерокот, па затоа се нарекува и **распределба на кумулативните релативни** или само кумулативни **честоти (фреквенции)** на примерокот.

Задача 1.2.1. Се набљудува должината на еден производ на една производствена лента. Случајно се избрани 20 производи чии должини измерени во сантиметри изнесува:

10, 8, 11, 6, 7, 9, 8, 8, 9, 10, 7, 7, 10, 8, 7, 10, 11, 10, 9, 7.

Во табелата 1.2.1.1, ќе бидат прикажани низата од вредности што ги прима примерокот, нивните фреквенции и релативни фреквенции, а, емпириската функција на распределба на примерок е дадена во табелата 1.2.1.2.

x_i	6	7	8	9	10	11
n_i	1	5	4	3	5	2
$\frac{n_i}{n}$	$\frac{1}{20}$	$\frac{5}{20}$	$\frac{4}{20}$	$\frac{3}{20}$	$\frac{5}{20}$	$\frac{2}{20}$

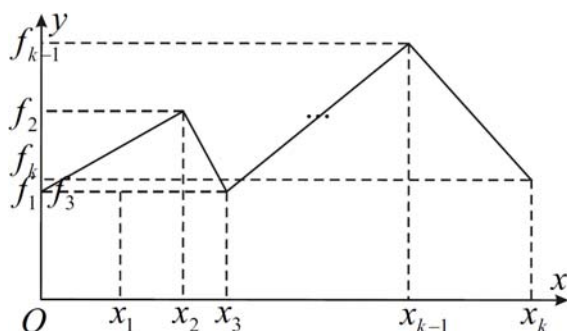
Табела 1.2.1.1.

x	$F_{20}(x)$
$x < 6$	0
$6 \leq x < 7$	$\frac{1}{20}$
$7 \leq x < 8$	$\frac{6}{20}$
$8 \leq x < 9$	$\frac{10}{20}$
$9 \leq x < 10$	$\frac{13}{20}$
$10 \leq x < 11$	$\frac{18}{20}$
$x \geq 11$	1

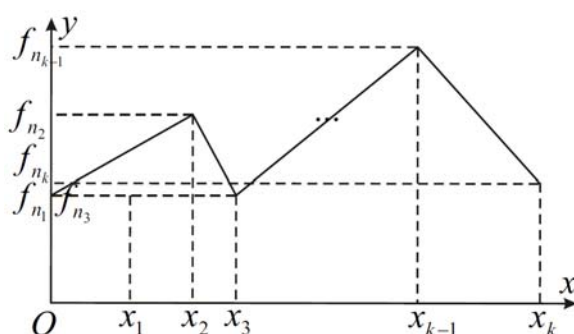
Табела 1.2.1.2. Емпириската функција на распределба на примерокот

Графичкото прикажување на податоците може да биде различно, во зависност од видот на податоците и нивната примена. Најчесто се користат дијаграми кои можат да бидат: точкестистигмограми, линиски-полигони, површински-хистограми и просторни-стереограми.

Податоците од табеларниот приказ може да се претстават така што на координатен систем ќе се нанесат точките (x_i, f_i) и ќе се поврзат со отсечки.



Цртеж 1.2.1.3. Полигон на честоти



Цртеж 1.2.1.4. Полигон на релативни честоти

Се добива искршена линија-полигон и се нарекува **полигон на распределба на апсолутните честоти**. Ако место (x_i, f_i) , $i = 1, 2, \dots, k$; ги нанесеме точките (x_i, f_{ni}) , $i = 1, 2, \dots, k$; го добиваме полигонот на **распределба на релативните честоти**.

Ако бројот на вредности на примерокот е голем тие се задаваат во интервали (групи), како што е случај кога случајната променлива е од апсолутно-непрекинат тип.

Интервалот $[a, b]$ во кој се сите вредности на примерокот и најчесто е минимален со тоа својство, се дели на n подинтервали:

$$[a, a_1), [a_1, a_2), \dots, [a_{n-2}, a_{n-1}), [a_{n-1}, b].$$

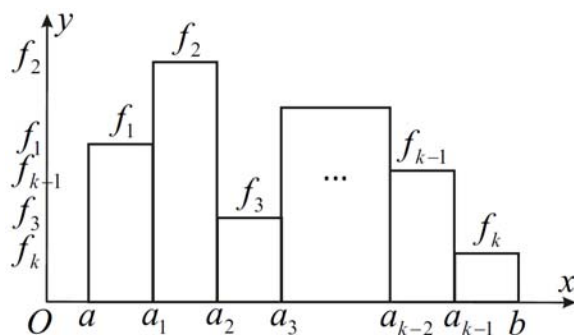
Најчесто интервалите се избираат да бидат со еднаква должина. Нека постојат k различни подинтервали

$$[a, a_1), [a_1, a_2), \dots, [a_{k-2}, a_{k-1}), [a_{k-1}, b];$$

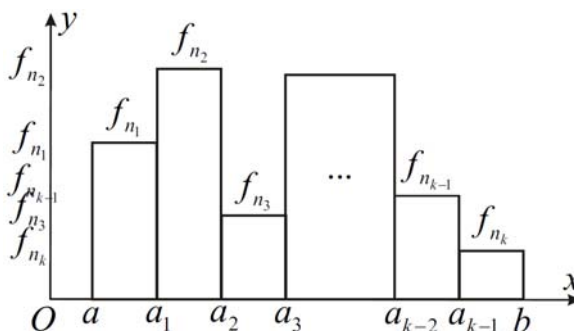
каде што $a = a_0 < a_1 < a_2 < \dots < a_{k-1} < b = a_n$; со честоти $f_1, f_2, \dots, f_k$, $f_1 + f_2 + \dots + f_k = n$. Податоците се прикажуваат во табела.

X	$[a, a_1)$	$[a_1, a_2)$	...	$[a_{k-1}, b]$	Σ
f_i	f_1	f_2	...	f_k	n
f_{ni}	f_{n1}	f_{n2}	...	f_{nk}	1

Податоците може да се претстават графички, така што се цртаат правоаголници со страни: должината на интервалот и соодветната вредност на честотата или релативната честота. Ваквото претставување се нарекува **хистограм** на распределбите на честоти и релативни честоти



Цртеж 1.2.1.5. Хистограм на честоти



Цртеж 1.2.1.6. Хистограм на релативни честоти

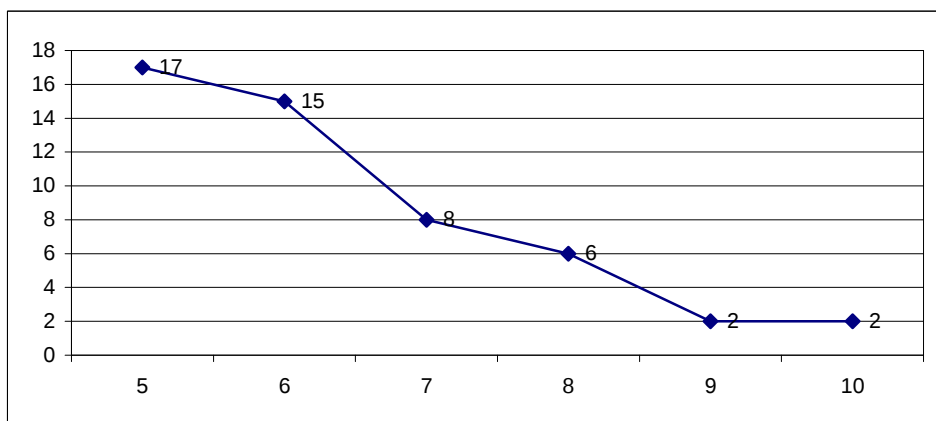
Задача 1.2.2. Во учебната 2022/23 година од запишани 118 студенти, испитот Математика 1 го полагаа 50 студенти. Од нив 2 положили со оценка 10, 2 со 9, 6 со 8, 8 со 7 и 15 со 6. Останатите не го положили испитот, т. е. добиле оценка 5. Да се запишат

податоците во табела, а потоа за се нацртаат полигоните и хистограмите со честоти и релативни честоти.

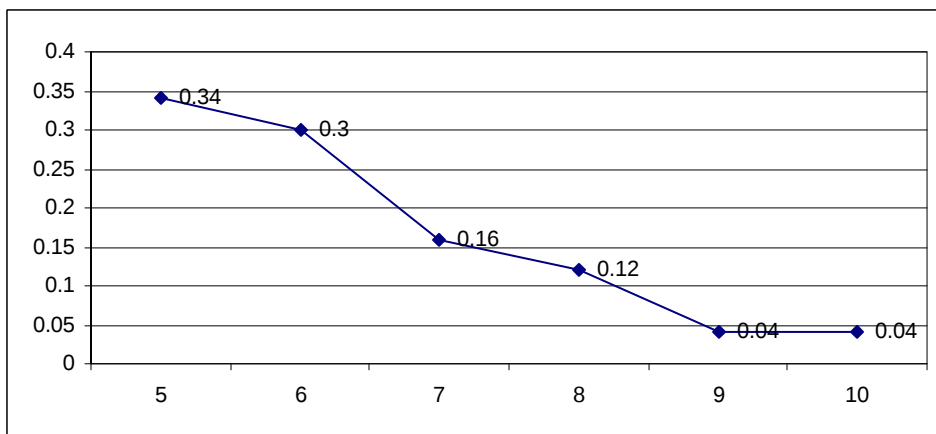
Решение. Прво ги запишуваме податоците во табела.

X	5	6	7	8	9	10	Σ
f_i	17	15	8	6	2	2	50
f_{ni}	$\frac{17}{50}$	$\frac{15}{50} = \frac{3}{10}$	$\frac{8}{50} = \frac{4}{25}$	$\frac{6}{50} = \frac{3}{25}$	$\frac{2}{50} = \frac{1}{25}$	$\frac{2}{50} = \frac{1}{25}$	1

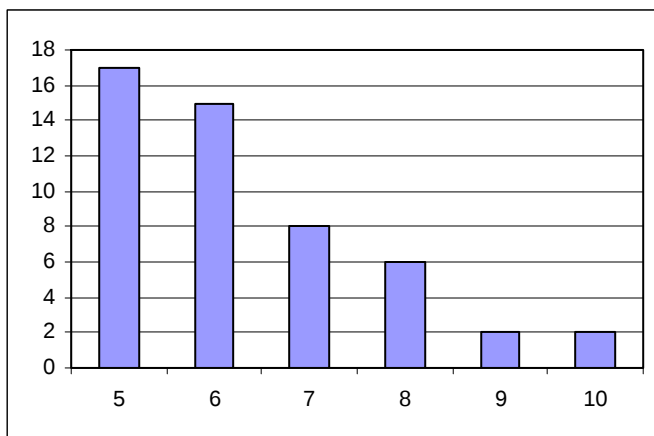
Табела 1.2.2.1.



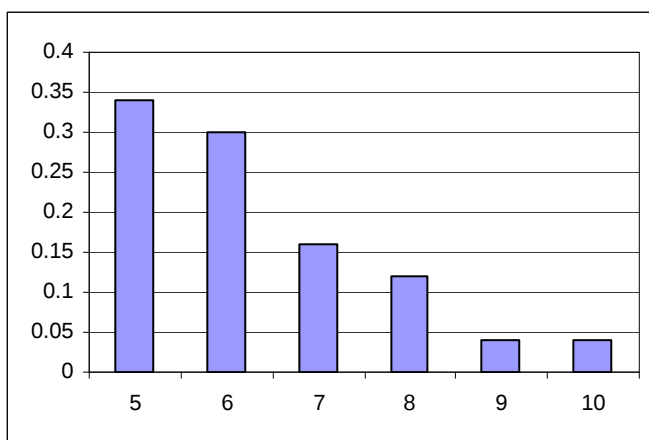
Цртеж 1.2.2.2. Полигон на честоти



Цртеж 1.2.2.3. Полигон на релативни честоти



Цртеж 1.1.2.4. Хистограм на честоти



Цртеж 1.1.2.5. Хистограм на релативни честоти

Да забележиме дека во сите претходни примери обележјата се квантитативни. Доколку обележјето е квалитативно за графичкиот приказ на примерокот е во облик на пита.

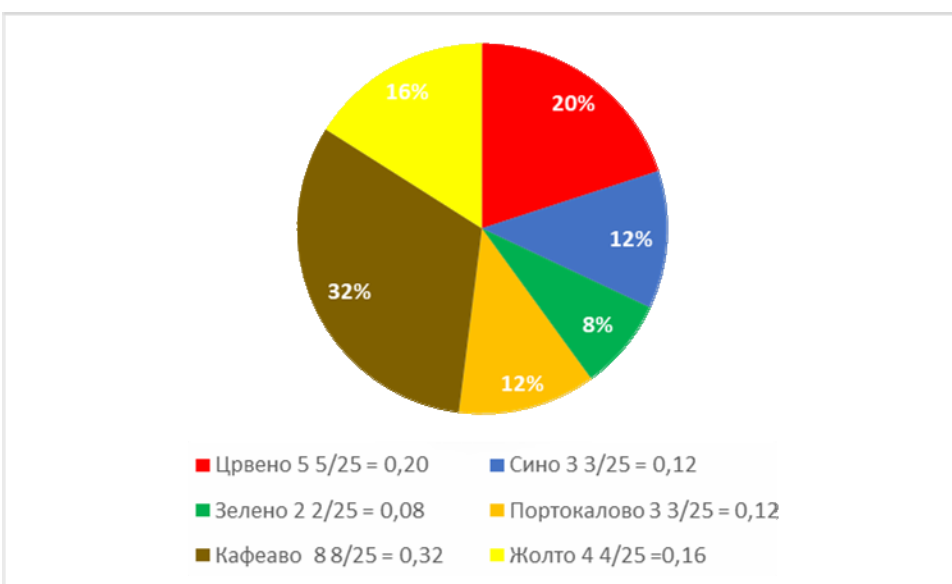
Пример 1.2.3. Да разгледаме едно пакетче со 25 M&M бонбончиња. Статистичката табела за него е следна:

Боја	Фреквенција	Релативна фреквенција	Процент
Црвено	5	$5/25 = 0,20$	20%
Сино	3	$3/25 = 0,12$	12%
Зелено	2	$2/25 = 0,08$	8%

1. Статистика

Портокалово	3	$3/25 = 0,12$	12%
Кафеаво	8	$8/25 = 0,32$	32%
Жолто	4	$4/25 = 0,16$	16%

Табела 1.2.3.1.



Табела 1.2.3.2.

1.3. Бројни карактеристики на примерокот

Покрај графичкиот и табеларниот приказ на податоците на една популација или примерок постојат и т.н. бројни карактеристики на популацијата, односно примерокот. Бројните карактеристики на една популација се викаат **параметри**, додека на примерокот се нарекуваат **статистики**.

Важни карактеристики на примерокот се неговите **средни вредности**. Тие се користат за да го опишат примерокот и да овозможат споредување со други примероци. Се делат на пресметковни и позициони.

Пресметковни средни вредности се аритметичка, геометричка и хармониска средина и др.

Позициони средни вредности се мода, медијана и др.

Од сите средни вредности аритметичката средина најчесто се користи.

Дефиниција 1.3.1. Аритметичката средина на примерокот $x_1, x_2, \dots, x_n$ е:

$$\bar{x} = \frac{x_1 + x_2 + \dots + x_n}{n} = \frac{1}{n} \sum_{i=1}^n x_i.$$

Аритметичката средина на примерокот во кој $x_1, x_2, \dots, x_k$ вредности се јавуваат со честоти $f_1, f_2, \dots, f_k$, $f_1 + f_2 + \dots + f_k = n$ е:

$$\bar{x} = \frac{f_1 x_1 + f_2 x_2 + \dots + f_k x_k}{f_1 + f_2 + \dots + f_k} = \frac{f_1 x_1 + f_2 x_2 + \dots + f_k x_k}{n} = \frac{1}{n} \sum_{i=1}^k f_i x_i.$$

Дефиниција 1.3.2. Групирана аритметичката средина на примерокот во кој вредностите се поделени на k интервали $[a_1, a_1), [a_1, a_2), \dots, [a_{n-2}, a_{n-1}), [a_{n-1}, b]$ е:

$$\bar{x} = \frac{1}{n} \sum_{i=1}^k f_i x_{si}, \quad x_{si} = \frac{x_i + x_{i+1}}{2}, \quad f_1 + f_2 + \dots + f_k = n.$$

Пример 1.3.3. Во табела ги внесуваме вредностите на случајната променлива X : Оценка на испитот Математика 1 (задача 1.2.2). Потоа го внесуваме бројот на кандидати f_i што добиле иста оценка x_i и на крајот го определуваме производот $f_i x_i$.

X	5	6	7	8	9	10	Σ
f_i	17	15	8	6	2	2	50
$f_i x_i$	85	90	56	48	18	20	317

Тогаш, аритметичката средина на оценките од студентите што полагале Математика 1 е:

$$\bar{x} = \frac{317}{50} = 6,34.$$

Дефиниција 1.3.4. Геометриска средина на примерок $x_1, x_2, \dots, x_k$, со честоти $f_1, f_2, \dots, f_k$, $f_1 + f_2 + \dots + f_k = n$ е:

$$G = \sqrt[n]{x_1^{n_1} x_2^{n_2} \dots x_k^{n_k}}.$$

Дефиниција 1.3.5. Хармониска средина на примерок $x_1, x_2, \dots, x_k$, со честоти $f_1, f_2, \dots, f_k$, $f_1 + f_2 + \dots + f_k = n$ е:

$$H = \frac{n}{\sum_{i=1}^k \frac{f_i}{x_i}}.$$

Хармониската средина се користи во задачи во кои се оценува просечната вредност на примерокот, кој е обратно пропорционален на примерокот што се мери. Такви се на пример: продуктивност-време, брзина-време и др.

Пример 1.3.6. За да се испита продуктивноста на една проектантска фирма се набљудува работата на 17 работници на изработка на елаборати. Добиени се следниве податоци за времето на изработка на еден елаборат (во часови):

t_i	6	8	10	12	14	16
f_i	1	2	7	4	2	1

Да се најде просечното време потребно за изработка на еден елаборат.

Решение. Нека t е просечното време на изработка на еден елаборат (во часови), тогаш за еден час се изработува $\frac{1}{t}$ елаборати.

Седумнаесетте работници кои работат со просечна брзина, за еден час ќе изработат $\frac{17}{t}$ елаборати. Работниците кои изработуваат

елаборат за 6 часа, за 1 час ќе изработат $\frac{1}{6}$ од елаборатот; за 8

часа $\frac{1}{8}$ и. т. н. Затоа:

$$\begin{aligned} \frac{1}{6} + 2\frac{1}{8} + 7\frac{1}{10} + 4\frac{1}{12} + 2\frac{1}{14} + \frac{1}{16} &= \frac{17}{t} \Leftrightarrow \frac{1}{6} + \frac{1}{4} + \frac{7}{10} + \frac{1}{3} + \frac{1}{7} + \frac{1}{16} = \frac{17}{t} \Leftrightarrow \\ \frac{230 + 420 + 1176 + 560 + 240 + 105}{1680} &= \frac{17}{t} \Leftrightarrow \end{aligned}$$

$$\frac{2731}{1680} = \frac{17}{t} \Leftrightarrow t = \frac{17 \cdot 1680}{2731} \approx 10,46.$$

Значи, просечното време на изработка е 10,46 часа.

Коментар. Просечното време на изработка е всушност хармониската средина:

$$H = \frac{17}{\frac{1}{6} + 2\frac{1}{8} + 7\frac{1}{10} + 4\frac{1}{12} + 2\frac{1}{14} + \frac{1}{16}} = \frac{17}{\frac{2731}{1680}} \approx 10,46.$$

За споредба, аритметичката средина на примерокот е:

$$\bar{x} = \frac{6+16+70+48+28+16}{17} = \frac{184}{17} \approx 10,82.$$

Мода. Од сите позициони средни вредности, модата најчесто се користи.

Дефиниција 1.3.7. Мода на примерокот е вредноста на примерокот која има најголема честота.

Дефиниција 1.3.8. Ако податоците од примерокот се групирани во интервали, тогаш интервалот со најголема честота се нарекува **модален интервал**. Тогаш **модата** е:

$$M_0 = a_1 + \frac{f_2 - f_1}{f_2 - f_1 + (f_2 - f_3)} d$$

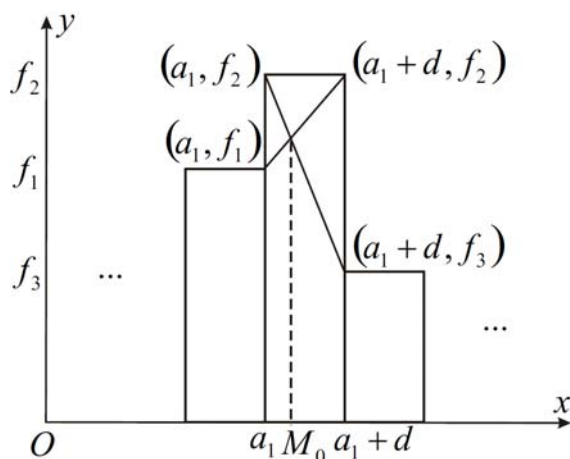
каде што a_1 е левата граница на модалниот интервал; f_1 , f_2 и f_3 се честотите на предмодалниот, модалниот и постмодалниот интервал; и d е должината на интервалот.

Равенката на правата низ точките (a_1, f_1) и $(a_1 + d, f_2)$ е:

$$y - y_0 = \frac{y_2 - y_1}{x_2 - x_1} (x - x_0) \Leftrightarrow y - f_1 = \frac{f_2 - f_1}{d} (x - a_1).$$

Равенката на правата низ точките (a_1, f_2) и $(a_1 + d, f_3)$ е:

$$y - y_0 = \frac{y_2 - y_1}{x_2 - x_1} (x - x_0) \Leftrightarrow y - f_2 = \frac{f_3 - f_2}{d} (x - a_1).$$



Цртеж 1.3.8.1. Геометриска интерпретација на модата

Модата претставува апциса на пресечната точка на правите т.е.

$$f_1 + \frac{f_2 - f_1}{d}(x - a_1) = f_2 + \frac{f_3 - f_2}{d}(x - a_1) \Leftrightarrow$$

$$\frac{f_2 - f_1 - (f_3 - f_2)}{d}(x - a_1) = f_2 - f_1 \Leftrightarrow x = a_1 + \frac{f_2 - f_1}{f_2 - f_1 + (f_2 - f_3)} d.$$

Пример 1.3.9. Примерок е должината на скокот од место на 32 студенти на Градежниот факултет на часот по физичко образование, даден во интервали од 10 единици (центиметри). Податоците се внесени во следнава табела:

cm	180-190	190-200	200-210	210-220	220-230	230-240
n_i	1	5	8	9	7	2

Да се најде модалниот интервал и модата на примерокот.

Решение. Модалниот интервал е 210 - 220, бидејќи во тој интервал е должината на скоковите на најмногу студенти (девет). Понатаму, $a_1 = 210$, $f_1 = 8$, $f_2 = 9$, $f_3 = 7$ и $d = 10$, од каде што модата е:

$$M_0 = a_1 + \frac{f_2 - f_1}{f_2 - f_1 + (f_2 - f_3)} d =$$

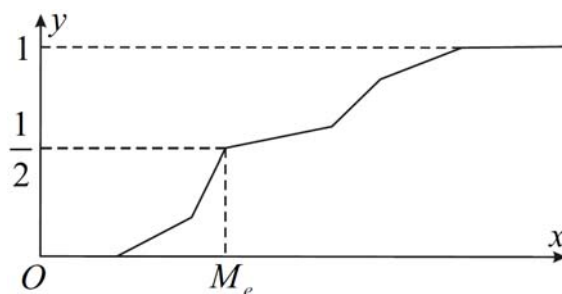
$$210 + \frac{9 - 8}{9 - 8 + 9 - 7} 10 = 210 + \frac{10}{3} = 213, (3).$$

Медијана. Медијана на примерокот е вредноста на елементот на примерокот кој го дели примерокот на два еднакви делови такви што половина елементи имаат помали, а половина поголеми вредности од медијаната.

Дефиниција 1.3.10. Медијана е средната вредност на вредностите на примерокот $x_1, x_2, \dots, x_n$, подредени по големина $x_1 \leq x_2 \leq \dots \leq x_n$.

Ако бројот на вредности на примерокот е непарен, тогаш медијаната е средна вредност на примерокот. Ако бројот на вредности е парен, тогаш медијаната е аритметичка средина на двата средни члена.

Медијаната може графички да се определи со помош на полигонот на кумулативни честоти. Имено, медијаната е еднаква на апцисата на точката од полигонот која има ордината 0,5 (цртеж 1.3.10.1).



Цртеж 1.3.10.1. Медијана на интервално зададен примерок

Пример 1.3.11. а) Медијаната на примерокот: 7, 21, 34, 55, 72, 88, 93; е: 55.

б) Медијаната на примерокот: 17, 23, 25, 30, 37, 40; е:

$$M_e = \frac{25 + 30}{2} = 27,5.$$

Ако вредностите на примерокот се дадени со интервали, тогаш прво се определува медијалниот интервал во кој се наоѓа средниот член, од каде што медијаната е:

$$M_e = a_1 + \frac{\frac{n}{2} - \sum_{i=1}^{m-1} n_i}{n_m},$$

каде што: a_1 е левата граница на медијалниот интервал, n е обемот на примерокот, n_m е честотата на медијалниот интервал и n_i , $i=1,2,\dots,m-1$ се честотите на интервалите пред медијалниот интервал.

Модата и медијаната имаат широка примена во статистиката. На пример ако ги испитуваме условите на домување на граѓаните, подобар показател е големината на станбената површина која ја користат најголем број од граѓаните (мода), отколку просечната површина по жител (аритметичка средина), бидејќи може мал процент богати граѓани што поседуваат многу големи станови да ја зголемат многу просечната станбена квадратура по жител. Слично се однесува за платите, модалната и медијалната плата се подобар показател за финансиската состојба на мнозинството од населението во една држава или регион.

Средниот век на траење на печатачите во една работна организација може да се определи, така што ќе се најде медијаната кога бројот на расходувани принтери ќе надмине половина од вкупниот број на печатачи.

1.4. Параметри кои го мерат расејувањето на случајната променлива околу центарот на растурање.

Дефиниција 1.4.1. Ранг (опсег) на примерокот $x_1, x_2, \dots, x_n$ е разликата меѓу најголемата и најмалата вредност на примерокот т.е.

$$I = x_{\max} - x_{\min}.$$

Дефиниција 1.4.2. Дисперзија (варијанса), ознака s^2 , на примерокот $x_1, x_2, \dots, x_n$; е бројот:

$$s^2 = \frac{1}{n} \sum_{i=1}^n (x_i - \bar{x})^2.$$

Својство 1.4.3. Дисперзијата на примерокот $x_1, x_2, \dots, x_n$; е бројот:

$$s^2 = \frac{1}{n} \sum_{i=1}^n x_i^2 - \bar{x}^2.$$

Доказ. Имаме:

$$\begin{aligned} s^2 &= \frac{1}{n} \sum_{i=1}^n (x_i - \bar{x})^2 = \frac{1}{n} \sum_{i=1}^n (x_i^2 - 2x_i\bar{x} + \bar{x}^2) = \\ &= \frac{1}{n} \left(\sum_{i=1}^n x_i^2 - 2\bar{x} \sum_{i=1}^n x_i + \bar{x}^2 \sum_{i=1}^n 1 \right) = \frac{1}{n} \sum_{i=1}^n x_i^2 - 2\bar{x} \frac{\sum_{i=1}^n x_i}{n} + \bar{x}^2 \frac{1}{n} n = \\ &= \frac{1}{n} \sum_{i=1}^n x_i^2 - 2\bar{x}^2 + \bar{x}^2 = \frac{1}{n} \sum_{i=1}^n x_i^2 - \bar{x}^2. \blacksquare \end{aligned}$$

Ако примерокот има $x_1, x_2, \dots, x_k$ вредности, со фреквенции $f_1, f_2, \dots, f_k$, $f_1 + f_2 + \dots + f_k = n$, тогаш:

$$s^2 = \frac{\sum_{i=1}^n f_i (x_i - \bar{x})^2}{\sum_{i=1}^n f_i} = \frac{1}{n} \sum_{i=1}^n f_i x_i^2 - \bar{x}^2.$$

Дефиниција 1.4.4. Поправена дисперзија (варијанса), ознака $\bar{s}^2$, на примерокот $x_1, x_2, \dots, x_n$; е бројот:

$$\bar{s}^2 = \frac{1}{n-1} \sum_{i=1}^n (x_i - \bar{x})^2.$$

Забелешка 1.4.5. Поправената дисперзија на примерокот е непристрасна од дисперзијата на примерокот и во пракса кога сака да се измери дисперзијата на примерокот, таа се пресметува со формулата за поправена дисперзија без да се нагласува дека тоа е поправена дисперзија на примерок.

Забелешка 1.4.6. Доколку податоците во примерокот се групирани во n интервали вредностите и кога $x_1, x_2, \dots, x_n$ се средните вредности на интервалите, тогаш за дисперзијата на примерокот велиме дека е групирана дисперзија на примерокот.

Дефиниција 1.4.7. Стандардна девијација на примерокот $x_1, x_2, \dots, x_n$; е бројот:

$$s = \sqrt{\frac{1}{n} \sum_{i=1}^n (x_i - \bar{x})^2}.$$

Ако примерокот има $x_1, x_2, \dots, x_k$ вредности, со фреквенции $f_1, f_2, \dots, f_k$, $f_1 + f_2 + \dots + f_k = n$, тогаш:

$$s = \sqrt{\frac{1}{n} \sum_{i=1}^n f_i x_i^2 - \bar{x}^2}.$$

Дефиниција 1.4.8. Коефициент на корелација на примерокот $x_1, x_2, \dots, x_n$; е бројот:

$$K = \frac{s}{\bar{x}} = \frac{100s}{\bar{x}}\%.$$

Коефициентот на корелација се користи кога треба да се споредуваат варијации на различни примероци изразени во различни единици мерки или иста единица но со различни аритметички средини.

Пример 1.4.9. Во еден каменолом во просек секој ден се произведува $\bar{x}_1 = 4t$ песок со стандардна девијација од $s_1 = 0,8t$. Во друг каменолом секој ден се произведуваат $\bar{x}_2 = 12t$ песок со стандардна девијација од $s_2 = 1,2t$. Во кој каменолом се произведува повеќе песок?

Решение. За да ја направиме споредбата ќе ги пресметаме коефициентите на корелација на двата каменоломи:

$$K_1 = \frac{100 \cdot 0,8}{4}\% = 20\% \text{ и } K_2 = \frac{100 \cdot 1,2}{12}\% = 10\%.$$

Заклучокот е дека во првиот каменолом се произведува поголема количина на песок.

1.5. Нумерички карактеристики на аритметичките средини на примерокот

Нека обележјето X на популацијата има математичко очекување $E(X) = \mu = \mu_x$ и дисперзија $\sigma^2 = \sigma_x^2$. Елементите $X_1, X_2, \dots, X_n$ од кој било примерок на оваа популација имаат исто математичко очекување $E(X_i) = \mu$ и дисперзија $D(x_i) = \sigma^2 = \sigma_x^2$.

Следува дека математичкото очекување, дисперзијата и коефициентот на корелација на аритметичката средина се:

$$\begin{aligned} E(\bar{X}) &= E\left(\frac{X_1 + X_2 + \dots + X_n}{n}\right) = \\ &= \frac{1}{n}(E(X_1) + E(X_2) + \dots + E(X_n)) = \frac{1}{n}n\mu = \mu; \\ s^2 &= D(\bar{X}) = D\left(\frac{X_1 + X_2 + \dots + X_n}{n}\right) = \\ &= \frac{1}{n^2}(D(X_1) + D(X_2) + \dots + D(X_n)) = \frac{1}{n^2}n\sigma^2 = \frac{\sigma^2}{n}; \end{aligned}$$

и

$$s = \frac{\sigma}{\sqrt{n}}.$$

Значи дека ако случајната променлива X која претставува некоја карактеристика на популацијата има нормална распределба $N(\mu, \sigma^2)$, тогаш нејзината аритметичка средина ќе има исто така

нормална распределба но од обликот $N\left(\mu, \frac{\sigma^2}{n}\right)$. Ако случајната

променлива X нема нормална распределба но $n \geq 30$, тогаш распределбата на аритметичките средини ќе тежи кон нормалната распределба.

1.6. ЗАДАЧИ СО ПОПУЛАЦИЈА И ПРИМЕРОК

Задача 1.6.1. Брзината на автомобилите на дел од патот изнесува:

51 51 39 25 51 53 52 44 51 40
 33 58 60 46 45 56 38 56 60 51
 65 37 53 63 58 57 44 45 39 52
 40 45 48 51 46 48 55 69 54 53

Да се конструираат хистограми на честотите, релативни честоти и кумулативните честоти и да се пресмета аритметичката средина и дисперзијата на примерокот.

Решение. Бројот на елементи во примерокот е $n = 40$. Елементите од примерокот ги групираме во интервали. Избираме бројот на интервали да биде околу $\sqrt{40} \approx 6$. Распонот на примерокот е $\max h = 69 - 25 = 44$. Значи дека елементите од примерокот кои се природни броеви, може да имаат некоја од $44 + 1 = 45$ различни броеви меѓу најголемата и најмалата вредност. Избираме секој интервал да содржи 8 последователни броеви, така што ќе додадеме 3 броја, најдобро 2 со 1, под и над максималната и минималната измерена вредност, на пример 23, 24 и 70 за да завршиме со цела десетка. Значи, интервалите се:

$$[23, 30], [31, 38], \dots, [63, 70].$$

За вредности x_i , $i = 1, 2, \dots, 6$; на секој интервал ја избираме средината на интервалот. Го определуваме бројот на елементи од примерокот во секој интервал, најпрегледно со различно прецртување на изброените елементи од секој интервал и формираме табела во која ги определуваме честотите f_i , кумулативните честоти f_{n_i} и производите $f_i x_i$ и $f_i x_i^2$, $i = 1, 2, \dots, 6$; потребни за пресметување на аритметичката средина и дисперзијата на примерокот.

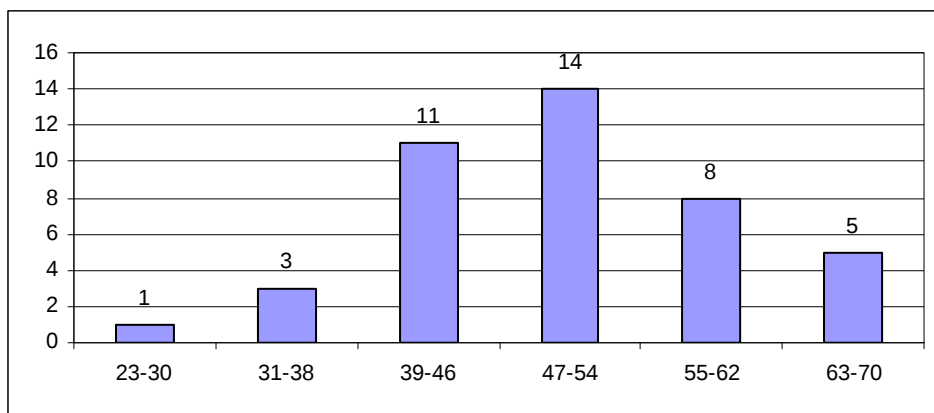
интервал	x_i	f_i	f_{n_i}	$f_i x_i$	$f_i x_i^2$
23-30	26,5	1	0	26,5	702,25
31-38	34,5	3	1	103,5	3570,75
39-46	42,5	11	4	467,5	19868,75

1. Веројатност

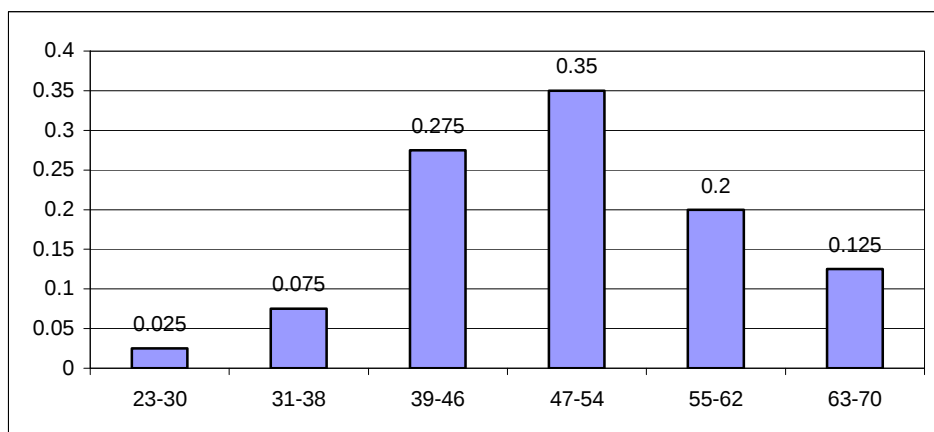
47-54	50,5	14	15	707	35703,5
55-62	58,5	8	29	468	27378
63-70	66,5	5	37	199,5	13266,75
Сума		40	40	1972	100491

Табела 1.6.1.1.

Според податоците од табелата ги цртаме бараните хистограми.



Цртеж 1.6.1.2. Хистограм на честоти



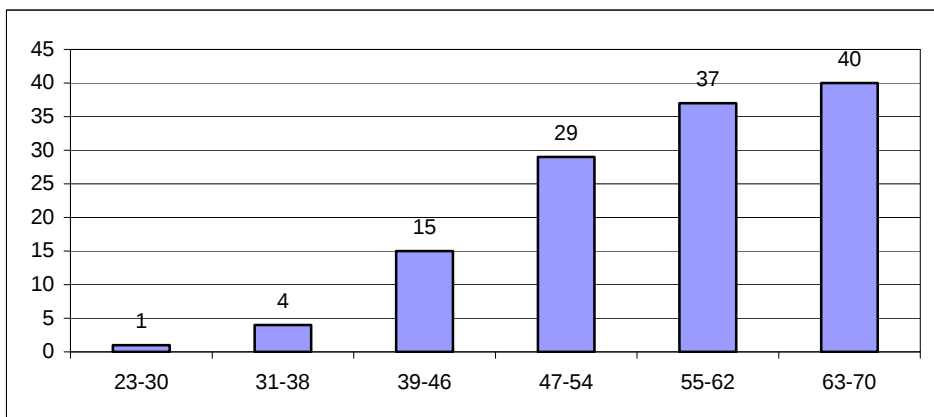
Цртеж 1.6.1.3. Хистограм на релативни честоти

Аритметичката средина на примерокот е:

$$\bar{x} = \frac{1}{n} \sum_{i=1}^6 f_i x_i = \frac{1972}{40} = 49,3;$$

додека дисперзијата:

$$s^2 = \frac{1}{n} \sum_{i=1}^6 f_i x_i^2 - \bar{x}^2 = \frac{100490}{40} - 49,3^2 = 2512,55 - 2430,5 = 81,75.$$



Цртеж 1.6.1.4. Хистограм на кумулативни честоти

Задача 1.6.2. Должината на работата на 30 електрични лампи мерен во часови по 10^{-1} е:

51 56 69 31 56 49 51 53 74 51
 63 48 53 51 64 50 59 84 55 82.
 55 72 70 54 51 77 90 62 73 58

Да се конструираат хистограми на честотите, релативните честоти и кумулативните честоти и да се пресмета аритметичката средина и дисперзијата на примерокот.

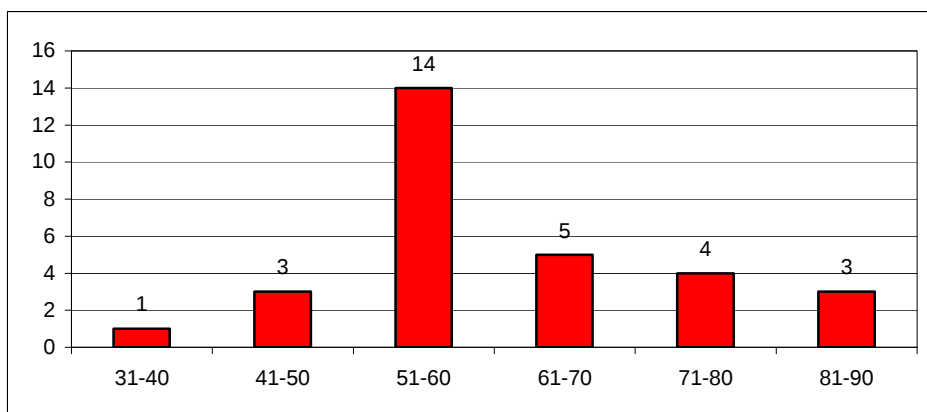
Решение. Бројот на елементи во примерокот е $n = 30$. Распонот е $\max h = 90 - 31 = 59$. Бидејќи $\sqrt{30} \approx 5,5$, и постојат 60 вредности меѓу најголемата и најмалата вредност, податоците ги делиме во 6 интервали и ја формираме следнава табела:

интервал	x_i	f_i	f_{n_i}	$f_i x_i$	$f_i x_i^2$
31-40	34,5	1	0	34,5	34,5
41-50	44,5	3	1	133,5	400,5
51-60	54,5	14	4	763	10632
61-70	64,5	5	18	322,5	1612,5
71-80	74,5	4	23	238	1132
81-90	84,5	3	27	253,5	760,5

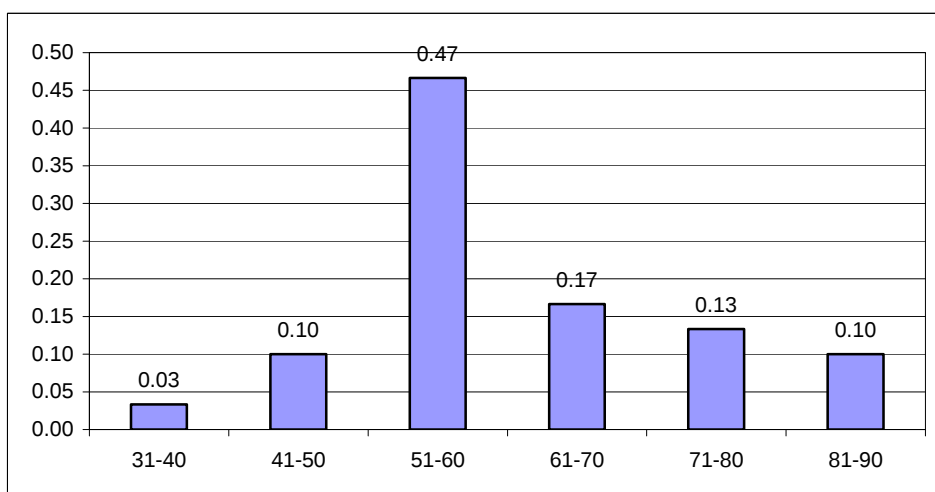
1. Веројатност

Сума		30	30	1001	11764
------	--	----	----	------	-------

Табела 1.6.2.1.



Цртеж 1.6.2.2. Хистограми на честоти



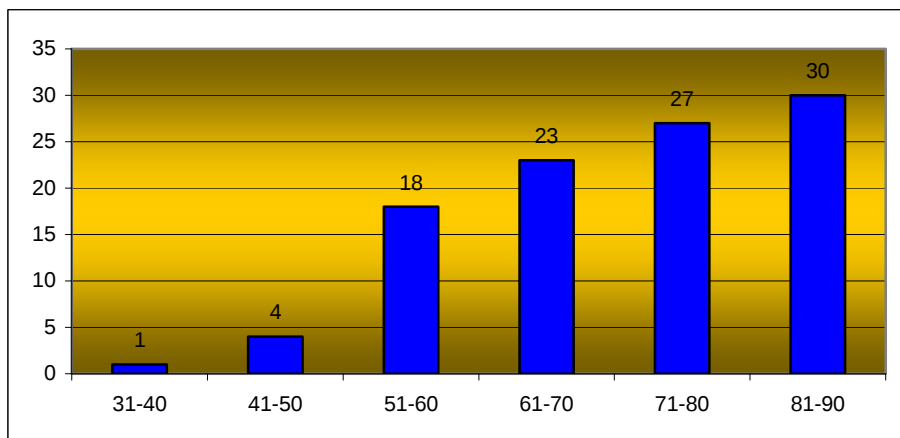
Цртеж 1.6.2.3. Хистограм на релативни честоти

Аритметичката средина на е:

$$\bar{x} = \frac{1}{n} \sum_{i=1}^6 f_i x_i = \frac{1001}{30} = 33,3(6),$$

додека дисперзијата:

$$s^2 = \frac{1}{n} \sum_{i=1}^6 f_i x_i^2 - \bar{x}^2 = \frac{11764}{30} - (33,3(6))^2 = 392,1(3) - 1113, (3) = -278,8.$$



Цртеж 1.6.2.4. Хистограм на кумулативни честоти

Задача 1.6.3. Бројот на дефектни производи при 36 контроли од по 10 производи изнесува:

0 0 1 0 2 0 1 2 1 0 0 0
 0 0 3 1 0 0 0 0 0 2 0 0.
 1 1 0 0 0 1 1 0 1 0 1 1

Да се најдат распределбите на честотите, релативните честоти, кумулативните честоти и емпириските функции на распределба, и да се конструираат соодветните полигони. Да се пресметаат аритметичката средина и дисперзијата на примерокот.

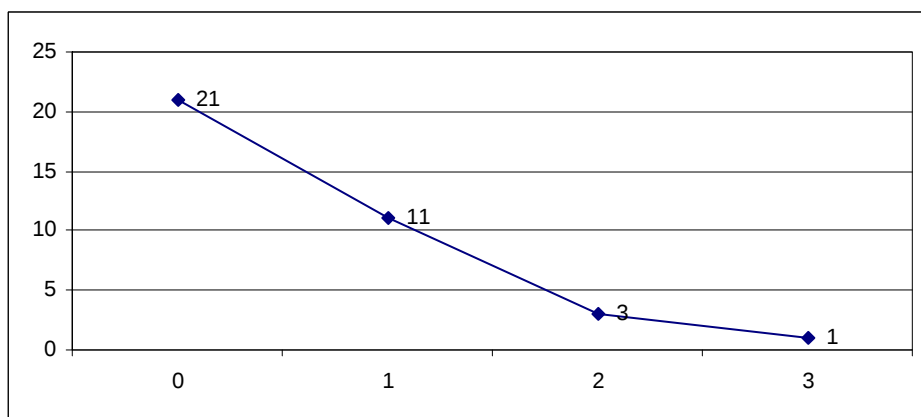
Решение. Имаме $n = 36$ и $\max h = 3$. Во овој случај, бидејќи елементите од примерокот имаат 4 вредности, немаме друг избор освен да ги поделиме во 4 групи во растечки редослед. Ја формираме табелата.

x_i	f_i	$\frac{f_i}{n}$	f_{n_i}	$\frac{f_{n_i}}{n}$	$f_i x_i$	$f_i x_i^2$
0	21	$\frac{21}{36} = \frac{7}{12} \approx 0,58$	0	0	0	0

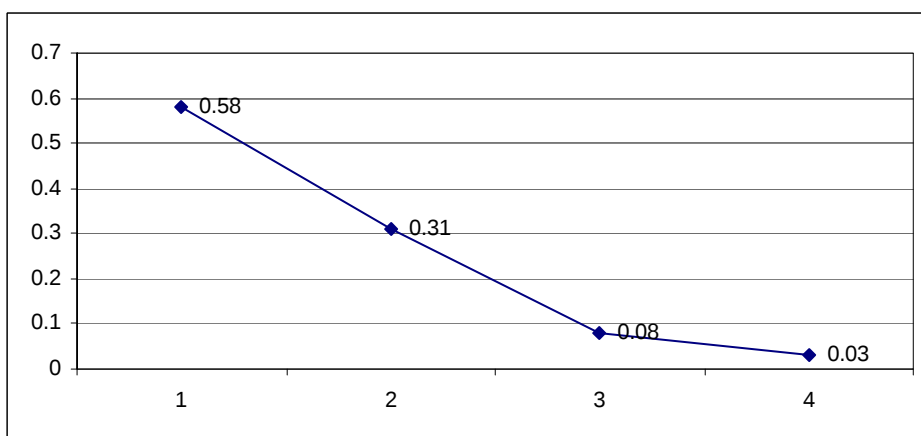
1. Веројатност

1	11	$\frac{11}{36} \approx 0,31$	21	$\frac{21}{36} = \frac{7}{12} \approx 0,58$	11	11
2	3	$\frac{3}{36} = \frac{1}{12} \approx 0,08$	32	$\frac{32}{36} = \frac{8}{9} \approx 0,89$	6	12
3	1	$\frac{1}{36} \approx 0,03$	35	$\frac{35}{36} \approx 0,97$	3	9
Сума	36	1	36	1	20	32

Табела 1.6.3.1.



Цртеж 1.6.3.2. Полигон на честоти



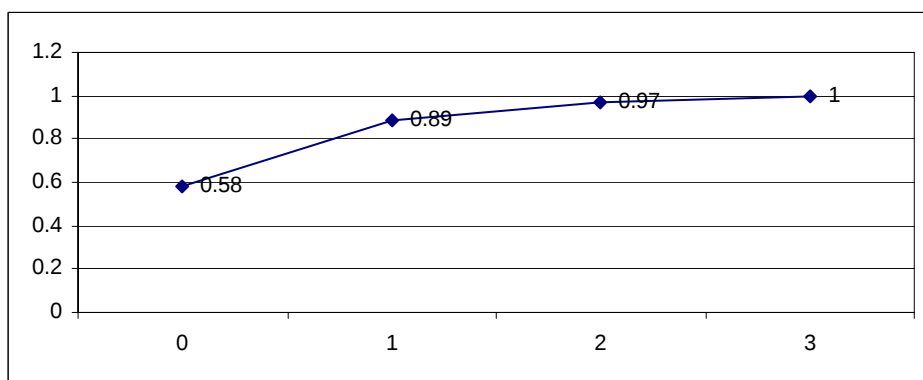
Цртеж 1.6.3.3. Полигон на релативни честоти

Аритметичката средина на е:

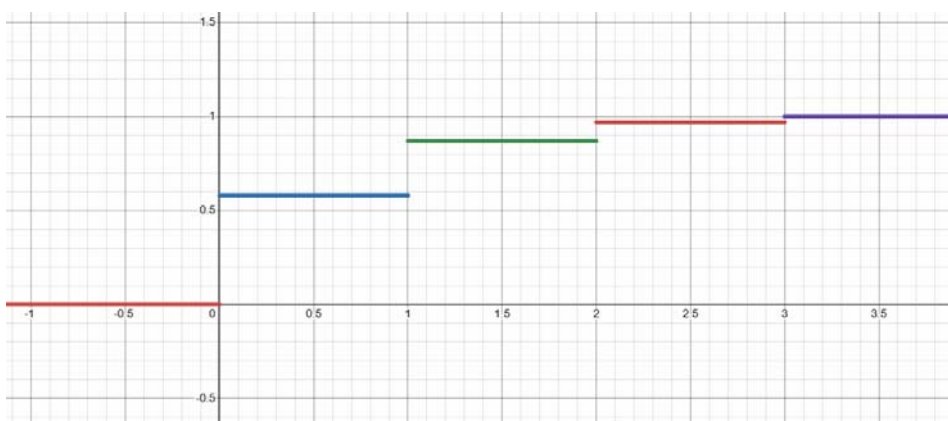
$$\bar{x} = \frac{1}{36} \sum_{i=1}^4 f_i x_i = \frac{20}{36} = \frac{5}{9},$$

додека дисперзијата:

$$s^2 = \frac{1}{36} \sum_{i=1}^4 f_i x_i^2 - \bar{x}^2 = \frac{32}{36} - \frac{25}{81} = \frac{188}{324} = \frac{47}{81}.$$



Цртеж 1.6.3.4. Полигон на кумулативни честоти



Цртеж 1.6.3.5. Емпириска функција на распределба

Задача 1.6.4. Мерен е напонот на електричната мрежа во волти, при што се добиени следниве податоци:

217 218 219 217 218 220 220 219 220 221
 218 219 220 218 217 220 219 221 221 220.
 220 216 220 219 220 218 222 219 219 220

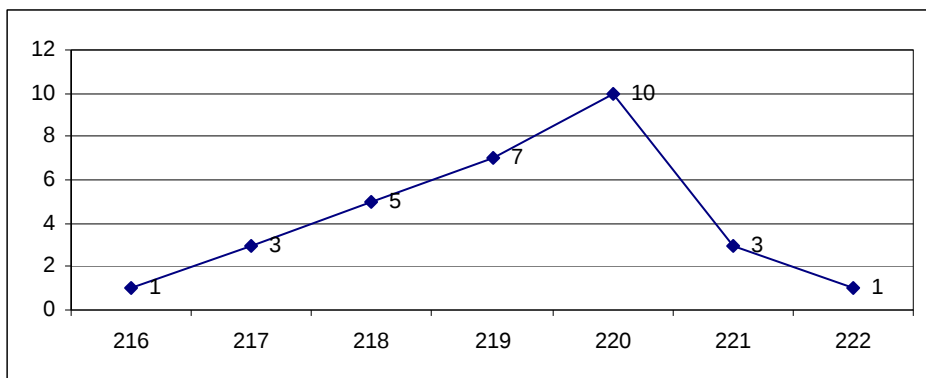
Да се најдат распределбите на честотите, релативните честоти, кумулативните честоти и емпириските функции на распределба, и да се конструираат соодветните полигони. Да се пресметаат аритметичката средина и дисперзијата на примерокот.

Решение. Распонот е $\max h = 222 - 216 = 6$, т. е. постојат 7 вредности меѓу крајните и $\sqrt{30} \approx 5,5$. Ако ги поделеме податоците во 5 или 6 интервали, секој интервал би содржел две вредности и би постоел интервал кој не содржи измерени вредности. Бидејќи бројот 7 не отстапува многу, нема многу да се зголемат пресметките, примерокот ќе го поделеме во 7 групи, така што секоја од нив ќе содржи по една вредност.

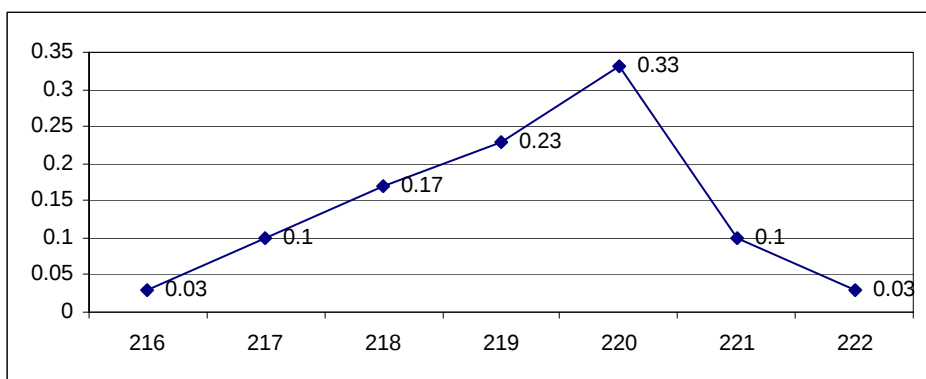
x_i	f_i	$\frac{f_i}{n}$	f_{n_i}	$\frac{f_{n_i}}{n}$	$f_i x_i$	$f_i x_i^2$
216	1	$\frac{1}{30} = 0,0(3)$	0	0	216	216
217	3	$\frac{3}{30} = \frac{1}{10} = 0,1$	1	$\frac{1}{30} = 0,0(3)$	651	1953
218	5	$\frac{5}{30} = \frac{1}{6} = 0,1(6)$	4	$\frac{4}{30} = \frac{2}{15} = 0,1(3)$	1090	5450
219	7	$\frac{7}{30} = 0,2(3)$	9	$\frac{9}{30} = \frac{3}{10} = 0,3$	1533	10731
220	10	$\frac{10}{30} = \frac{1}{3} = 0,3(3)$	16	$\frac{16}{30} = \frac{8}{15} = 0,5(3)$	2200	22000
221	3	$\frac{3}{30} = \frac{1}{10} = 0,1$	26	$\frac{26}{30} = \frac{13}{15} = 0,8(6)$	663	1989
222	1	$\frac{1}{30} = 0,0(3)$	29	$\frac{29}{30} = 0,9(6)$	222	222
Сума	30	1	30	1	6575	42561

Табела 1.6.4.1.

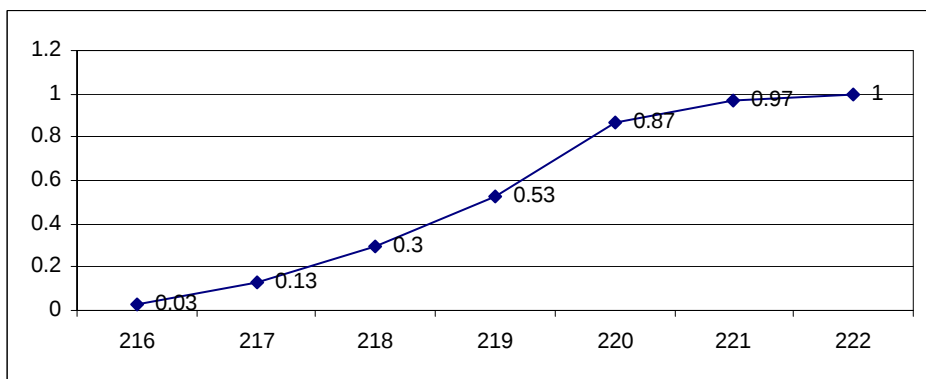
1. Веројатност



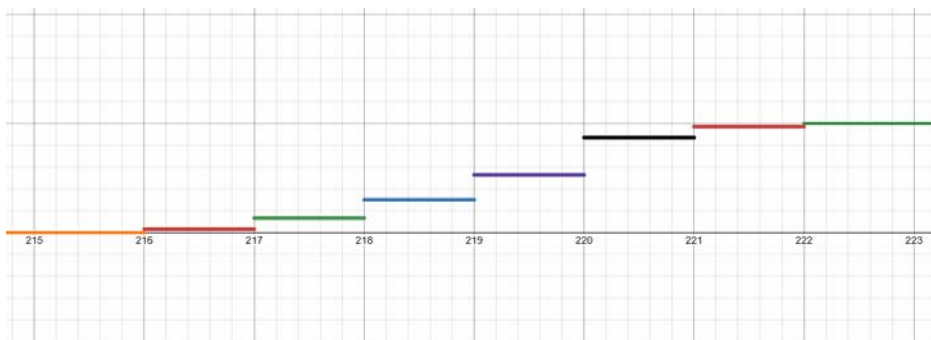
Цртеж 1.6.4.2. Полигон на честоти



Цртеж 1.6.4.3. Полигони на релативни честоти



Цртеж 1.6.4.4. Полигон на кумулативни честоти



Цртеж 1.6.4.5. Полигон на емпириска функција на распределба

Аритметичката средина на е:

$$\bar{x} = \frac{1}{30} \sum_{i=1}^7 f_i x_i = \frac{6575}{30} = 219,1(6),$$

додека дисперзијата:

$$s^2 = \frac{1}{30} \sum_{i=1}^7 f_i x_i^2 - \bar{x}^2 = \frac{42561}{30} - (219,1(6))^2 = 1418,7 - 48034,02(7) = -46615,32(7).$$

2. СТАТИСТИЧКО ОЦЕНУВАЊЕ. ИНТЕРВАЛИ НА ДОВЕРБА

Ако не се располага со податоци врз основа на кои би можеле да се претпостават вредностите на одредени параметри на популацијата (најчесто тоа се математичко очекување, варијанса, односно стандардна девијација), овие вредности се оценуваат со постапката за **статистичко оценување**. Бидејќи нумеричките вредности на параметрите ги оценуваме врз основа на информациите од примерокот, не можеме да бидеме целосно сигурни во исправноста на изведениот заклучок. Затоа, заклучокот со оценувањето го прифаќаме со сигурност помала од 100%.

Методите на статистичкото оценување можат да се поделат по повеќе критериуми. Најопштата поделба е на **параметарски** и **непараметарски** методи. Основниот критериум за оваа поделба е степенот на строгост на условите кои треба да бидат исполнети за да се примени одреден метод. Од наш интерес ќе бидат параметарските методи на статистичко оценување, кои се засновани на построги претпоставки отколку непараметарските. Друга поделба е на **точкести оценки** и **интервали на доверба**. Во математичката статистика од посебен интерес се интервалите на доверба.

2.1. ОЦЕНКИ И НИВНИ КАРАКТЕРИСТИКИ

Во постапката за статистичко оценување, заклучокот за непозната вредност на параметарското множество се изведува врз основа на соодветната статистика на следниот начин непознатата вредност на константата се оценува врз основа на соодветната случајна променлива X од која се разгледува примерокот $X_1, X_2, \dots, X_n$, потоа врз основа на неговата реализирана вредност $x_1, x_2, \dots, x_n$ се пресметува нумеричката (бројната) вредноста на статистиката која претставува оценка на параметарот. Во општ случај, непознатиот параметар на популацијата се означува со θ . Статистиката која се избира врз основа на примерокот и со која се оценува тој параметар се означува со $\tilde{\theta} = f(X_1, X_2, \dots, X_n)$, додека реализираната вредност на оценката во избраниот примерок се нарекува **вредност на оценката** на параметарот и се означува со $v = f(x_1, x_2, \dots, x_n)$. Притоа за друга реализација $x'_1, x'_2, \dots, x'_n$ на

примерокот се добива друга бројна вредност на оценката $v' = f(x'_1, x'_2, \dots, x'_n)$.

Оценетата вредност на параметарот е еден број, т.е. претставува точка на реалната права. Затоа, v се нарекува **точкеста вредност на оценката** на параметарот, а соодветната оценка $\tilde{\theta}$ се нарекува **точкеста оценка**. Точкестата оценка е функција $\tilde{\theta} = f(X_1, X_2, \dots, X_n)$, претставена со алгебарски израз (формула), а нумеричката вредност на таа функција за дадените вредности (реализации) во примерокот, претставува точкеста вредност на оценката на параметарот $v = f(x_1, x_2, \dots, x_n)$.

Најчести карактеристики на статистичката оценка, за кои се дискутира во поголемиот дел на статистички литератури се: непристрасност, ефикасност, конзистентност и доволност на оценката.

Оценката $\tilde{\theta}$ е **непристрасна** ако нејзината очекувана вредност е еднаква на параметарот θ , т. е. $E(\tilde{\theta}) = \theta$. Оваа карактеристика е неопходна, но не е доволна. На пример, оценката $\tilde{\theta}$ може да биде непристрасна и истовремено да има голема дисперзија. Тогаш, оценетите вредности v во различни примероци значајно отстапуваат од очекуваната вредност на параметарот θ , па во заклучокот може да се појави голема случајна грешка. Затоа, покрај непристрасност, добрата оценка треба да се карактеризира со што помал варијабилитет. За една оценка велиме дека е поефикасна од друга, ако има помала варијанса, односно стандардна грешка, за истиот примерок.

Оценката е **конзистентна**, ако со зголемување на големината на примерокот, таа тежи кон параметарот θ . Односно, како расте големината на примерокот θ , така оценетите вредности v за соодветните примероци сè повеќе се концентрираат околу параметарот θ .

За оценката на еден параметар велиме дека е **доволна**, ако ги користи сите информации кои примерокот ги содржи за дадениот параметар.

Наведените карактеристики за параметрите претставуваат критериуми за подредување на повеќе оценки за ист параметар. Врз основа на овие критериуми се избира најдобра оценка со која се оценува непознатата вредност на параметарот θ .

Следно, ќе разгледаме два метода за точкеста оценка на непознати параметри, методот на максимална веродостојност и методот на моменти.

2.2. МЕТОД НА МАКСИМАЛНА ВЕРОДОСТОЈНОСТ

Најопштиот метод за добивање точкести оценки на параметри е **методот на максимална веродостојност**. Притоа оценките добиени со овој метод не мораат да бидат непристрасни, но кога примерокот е значајно голем, овие оценки се најефикасни. Примената на методот на максимална веродостојност се реализира со сложени пресметувања и таа во општ случај ќе биде презентирана во следната дискусија.

Имено, нека распределбата на обележјето X е дадена со густината на распределба $p(x; \theta_1, \theta_2, \dots, \theta_k)$ или во дискретен случај со законот на распределба $P(x; \theta_1, \theta_2, \dots, \theta_k)$, $x = x_1, x_2, \dots$, каде што $\theta_1, \theta_2, \dots, \theta_k$ се непознати параметри. Дефинираме т.н **функција на веродостојност**:

$$L(x_1, x_2, \dots, x_n; \theta_1, \theta_2, \dots, \theta_k) = \prod_{i=1}^n p(x_i; \theta_1, \theta_2, \dots, \theta_k)$$

односно

$$L(x_1, x_2, \dots, x_n; \theta_1, \theta_2, \dots, \theta_k) = \prod_{i=1}^n P(x_i; \theta_1, \theta_2, \dots, \theta_k)$$

каде што $x_1, x_2, \dots, x_n$ е реализација на примерокот со големина n . Да забележиме дека оценките на непознатите параметри:

$$\begin{aligned} \tilde{\theta}_1 &= \tilde{\theta}_1(X_1, X_2, \dots, X_n) \\ \tilde{\theta}_2 &= \tilde{\theta}_2(X_1, X_2, \dots, X_n) \\ &\dots\dots\dots \\ \tilde{\theta}_k &= \tilde{\theta}_k(X_1, X_2, \dots, X_n) \end{aligned} \quad (1)$$

се статистики со кои се добива максималната вредност на функцијата на веродостојност, односно се функции со кои појавата на примерокот ќе биде најверојатна. Значи дека функцијата на веродостојност има максимум за вредностите на параметрите во решенијата на следниов систем равенки:

$$\frac{\partial L}{\partial \theta_i} = 0, \quad i = 1, 2, \dots, k$$

Бидејќи функцијата на веродостојност има екстрими во истите точки како и логаритамската функција на веродостојност:

$$\ln L(x_1, x_2, \dots, x_n; \theta_1, \theta_2, \dots, \theta_k) = \sum_{i=1}^n \ln p(x_i; \theta_1, \theta_2, \dots, \theta_k),$$

ОДНОСНО

$$\ln L(x_1, x_2, \dots, x_n; \theta_1, \theta_2, \dots, \theta_k) = \sum_{i=1}^n \ln P(x_i; \theta_1, \theta_2, \dots, \theta_k)$$

А постапката за решавање на системот равенки:

$$\frac{\partial \ln L}{\partial \theta_i} = 0, \quad i = 1, 2, \dots, k;$$

со која се добива максималната вредност на функцијата на веродостојност, односно статистиките (1) се поедноставни. Затоа, бараните оценки се решенијата на овој систем и врз основа на истите со реализација на примерокот ги добиваме точкестите оценки на параметрите од облик:

$$\begin{aligned} v_1 &= \tilde{\theta}_1(x_1, x_2, \dots, x_n) \\ v_2 &= \tilde{\theta}_2(x_1, x_2, \dots, x_n) \\ &\dots\dots\dots \\ v_k &= \tilde{\theta}_k(x_1, x_2, \dots, x_n) \end{aligned}$$

Задача 2.2.1. Со методот на максимална веродостојност, врз основа на примерок со големина $n = 50$, да се одреди точкестата оценка на параметарот λ во Пуасоновата распределба:

$$P(x, \lambda) = P(X = x) = \frac{\lambda^x}{x!} e^{-\lambda}, \quad x = 0, 1, 2, \dots;$$

ако во примерокот се регистрирани следните вредности:

x_i	0	1	2	3	4
n_i	16	18	8	5	3

Решение. Функцијата на максимална веродостојност на Пуасоновата распределба е од обликот:

$$L(x_1, x_2, \dots, x_n, \lambda) = \prod_{i=1}^n P(X_i, \lambda) = \prod_{i=1}^n \frac{\lambda^{x_i}}{x_i!} e^{-\lambda}.$$

Логаритамската функција на веродостојност е:

$$\begin{aligned} \ln L(x_1, x_2, \dots, x_n, \lambda) &= \sum_{i=1}^n \ln P(X_i, \lambda) = \sum_{i=1}^n \ln \frac{\lambda^{x_i}}{x_i!} e^{-\lambda} = \\ &= \sum_{i=1}^n (x_i \ln \lambda - \ln(x_i!) - \lambda). \end{aligned}$$

Оттука, системот равенки е од видот:

$$\begin{aligned} \frac{\partial \ln L(x_1, x_2, \dots, x_n, \lambda)}{\partial \lambda} &= \frac{\partial}{\partial \lambda} \left(\sum_{i=1}^n (x_i \ln \lambda - \ln(x_i!) - \lambda) \right) = \\ &= \sum_{i=1}^n \left(\frac{\partial}{\partial \lambda} (x_i \ln \lambda) - \frac{\partial}{\partial \lambda} (\ln(x_i!)) - \frac{\partial}{\partial \lambda} (\lambda) \right) = \sum_{i=1}^n \left(\frac{x_i}{\lambda} - 1 \right) = \frac{1}{\lambda} \sum_{i=1}^n x_i - n = 0. \end{aligned}$$

Значи, оценката на параметарот е од обликот:

$$\tilde{\lambda} = \frac{1}{n} \sum_{i=1}^n x_i,$$

додека точкестата оценка има вредност:

$$\lambda = \frac{0 \cdot 16 + 1 \cdot 18 + 2 \cdot 8 + 3 \cdot 5 + 4 \cdot 5}{50} = 3,45.$$

Задача 2.2.2. Бројот на сообраќајни несреќи во Скопје, во 10 случајно избрани неврнежливи денови во 2023 година е:

$$4, 0, 6, 5, 2, 1, 2, 0, 4, 3.$$

Проценете го процентот на неврнежливи денови во кои имало 2 или помалку несреќи таа година.

Решение. Прво ќе одредиме која случајна променлива X е соодветна на распределбата на обележјето дневен број на сообраќајни несреќи во Скопје, во неврнежливите денови. Имено, бидејќи станува збор за голем број возачи, секој со мала веројатност да биде вклучен во сообраќајна несреќа во даден ден, дневниот број на сообраќајни несреќи е Пуасоновата случајна променлива. Притоа, за Пуасоновата распределба е познато дека

точкестиот оценувач за нејзиниот параметар според задачата 2.2.1 е одреден со статистиката $\tilde{\lambda} = \frac{1}{n} \sum_{i=1}^n x_i$, додека точкестата оценка во согласност со дадениот примерок има вредност:

$$\lambda = \frac{4+0+6+5+2+1+2+0+4+3}{10} = 2,7$$

Проценката на процентот на неврнежливи денови во кои имало 2 или помалку несреќи таа година претставува веројатноста:

$$P(X \leq 2) = \sum_{x=0}^2 \frac{2,7^x}{x!} e^{-2,7} = 0,4936$$

Односно, во помалку од половината неврнежливи денови имало 2 или помалку несреќи.

Задача 2.2.3. Со методот на максимална веродостојност да се одреди статистиката на точкестата оценка на параметрите на нормалната случајна променлива $X : N(\mu, \sigma^2)$ со распределба:

$$p(x) = \frac{1}{\sqrt{2\pi}\sigma} e^{\frac{-(x-\mu)}{2\sigma^2}}, \quad x \in \mathbb{R};$$

Потоа да се провери нивната непристрасност, и да се направи корекција за да статистиката на параметрите на нормалната распределба се непристрасни.

Решение. Функцијата на максимална веродостојност на нормалната распределба е од обликот:

$$L(x_1, x_2, \dots, x_n, \mu, \sigma^2) = \prod_{i=1}^n p(x_i, \mu, \sigma^2) = \prod_{i=1}^n \frac{1}{\sqrt{2\pi}\sigma} e^{\frac{-(x_i-\mu)}{2\sigma^2}}.$$

Логаритамската функција на веродостојноста е:

$$\begin{aligned} \ln L(x_1, x_2, \dots, x_n, \mu, \sigma^2) &= \sum_{i=1}^n \ln \frac{1}{\sqrt{2\pi}\sigma} e^{\frac{-(x_i-\mu)}{2\sigma^2}} = \\ &= \sum_{i=1}^n \left(\ln 1 + \ln \left(e^{\frac{-(x_i-\mu)}{2\sigma^2}} \right) - \ln(\sqrt{2\pi}\sigma) \right) = \sum_{i=1}^n \left(\frac{-(x_i-\mu)}{2\sigma^2} - \ln(\sqrt{2\pi}) - \ln \sigma \right) = \end{aligned}$$

$$-\frac{n}{2}\ln(2\pi) - \frac{n}{2}\ln\sigma - \sum_{i=1}^n \frac{(x_i - \mu)^2}{2\sigma^2}.$$

Оттука, системот равенки е од видот:

$$\begin{aligned} \frac{\partial}{\partial \mu} \log f(x_1, \dots, x_n | \mu, \sigma) &= \left(-\frac{n}{2} \log(2\pi) \right)' - (n \log \sigma)' - \left(\frac{\sum_{i=1}^n (x_i - \mu)^2}{2\sigma^2} \right)' = \\ &= 0 - 0 - \frac{\sum_{i=1}^n ((x_i - \mu)^2)'}{2\sigma^2} = - \frac{\sum_{i=1}^n 2(x_i - \mu)((x_i - \mu))'}{2\sigma^2} = \\ &= - \frac{\sum_{i=1}^n 2(x_i - \mu)(-1)}{2\sigma^2} = \frac{\sum_{i=1}^n (x_i - \mu)}{\sigma^2} = 0; \end{aligned}$$

и

$$\begin{aligned} \frac{\partial}{\partial \sigma} \log f(x_1, \dots, x_n | \mu, \sigma) &= \left(-\frac{n}{2} \log(2\pi) \right)' - (n \log \sigma)' - \left(\frac{\sum_{i=1}^n (x_i - \mu)^2}{2\sigma^2} \right)' = \\ &= 0 - \frac{n}{\sigma} - \sum_{i=1}^n \left(\frac{(x_i - \mu)^2}{2} \right) (\sigma^{-2})' = -\frac{n}{\sigma} - \sum_{i=1}^n \left(\frac{(x_i - \mu)^2}{2} \right) (-2\sigma^{-3}) = \\ &= -\frac{n}{\sigma} + \sum_{i=1}^n \frac{(x_i - \mu)^2}{\sigma^3} = 0 \end{aligned}$$

Тогаш,

$$\begin{aligned} \frac{\sum_{i=1}^n (x_i - \mu)}{\sigma^2} = 0 &\Rightarrow \sum_{i=1}^n (x_i - \mu) = 0 \Rightarrow \sum_{i=1}^n x_i = \sum_{i=1}^n \mu \Rightarrow \sum_{i=1}^n x_i = n\mu \Rightarrow \tilde{\mu} = \frac{\sum_{i=1}^n x_i}{n} \\ -\frac{n}{\sigma} + \sum_{i=1}^n \frac{(x_i - \mu)^2}{\sigma^3} = 0 &\Rightarrow -\frac{n}{\sigma} + \frac{\sum_{i=1}^n (x_i - \mu)^2}{\sigma^3} = 0 \Rightarrow \frac{n}{\sigma} = \frac{\sum_{i=1}^n (x_i - \mu)^2}{\sigma^3} \\ &\Rightarrow n\sigma^3 = \sigma \sum_{i=1}^n (x_i - \mu)^2 \Rightarrow n\sigma^3 - \sigma \sum_{i=1}^n (x_i - \mu)^2 = 0 \end{aligned}$$

$$\Rightarrow \sigma \left(n\sigma^2 - \sum_{i=1}^n (x_i - \mu)^2 \right) = 0 \Rightarrow n\sigma^2 - \sum_{i=1}^n (x_i - \mu)^2 = 0 \Rightarrow \tilde{\sigma}^2 = \frac{\sum_{i=1}^n (x_i - \mu)^2}{n}$$

Значи, статистиките за оценките на параметрите се од облик:

$$\tilde{\mu} = \frac{\sum_{i=1}^n X_i}{n} = \bar{X}_n \text{ и } \tilde{\sigma}^2 = \frac{\sum_{i=1}^n (X_i - \tilde{\mu})^2}{n} = \frac{\sum_{i=1}^n (X_i - \bar{X}_n)^2}{n}.$$

Притоа, променливите X_i за секое $i=1, \dots, n$ имаат иста нормална распределба со параметри $E(X_i) = \mu, Var(X_i) = \sigma^2, i=1, \dots, n$.

Да забележиме дека важи равенството

$$\sum_{i=1}^n (X_i - \mu)^2 = \sum_{i=1}^n (X_i - \bar{X}_n)^2 + n(\bar{X}_n - \mu)^2$$

Сега проверуваме дали овие оценки се непристрасни. Имено,

$$E(\tilde{\mu}) = E\left(\frac{\sum_{i=1}^n X_i}{n}\right) = \frac{1}{n} \sum_{i=1}^n E(X_i) = \frac{1}{n} \sum_{i=1}^n \mu = \frac{n\mu}{n} = \mu,$$

Значи, статистиката е непристрасна. Но,

$$\begin{aligned} E(\tilde{\sigma}^2) &= E\left(\frac{\sum_{i=1}^n (X_i - \bar{X}_n)^2}{n}\right) = E\left(\frac{\sum_{i=1}^n (X_i - \mu)^2 - n(\bar{X}_n - \mu)^2}{n}\right) = \\ &= \frac{\sum_{i=1}^n E((X_i - \mu)^2) - E(n(\bar{X}_n - \mu)^2)}{n} = \frac{\sum_{i=1}^n V(X_i) - nE(\bar{X}_n - \mu)^2}{n} = \\ &= \frac{n\sigma^2}{n} - V(\bar{X}_n) = \sigma^2 - \frac{\sigma^2}{n} = \frac{n-1}{n} \sigma^2. \end{aligned}$$

Оттука, оваа статистика не е непристрасна, а и со мала корекција може да ја претвориме во непристрасна статистика за оценка за параметарот σ^2 од следниот облик:

$$\tilde{\sigma}^2 = \frac{n}{n-1} \frac{\sum_{i=1}^n (X_i - \tilde{\mu})^2}{n} = \frac{\sum_{i=1}^n (X_i - \bar{X}_n)^2}{n-1}.$$

Задача 2.2.4. Да претпоставиме дека се извршени n независни испитувања, од кои секоја е успешна со веројатност p . Која е статистиката за точкеста оценка на параметарот p ?

Решение. Да забележиме дека експериментот се однесува на n независни Бернулиеви случајни променливи со еднаков параметар, односно примерок $X_1, X_2, \dots, X_n$ каде што $X_i, i = 1, \dots, n$; има распределба:

$$P(X_i = x) = p^x (1-p)^{1-x}, x = 0, 1.$$

Функцијата на максимална веродостојност на Бернулиевата распределба е од обликот:

$$L(x_1, x_2, \dots, x_n, p) = \prod_{i=1}^n p^{x_i} (1-p)^{1-x_i}.$$

Логаритамската функција на веродостојност е:

$$\begin{aligned} \ln L(x_1, x_2, \dots, x_n, p) &= \ln \prod_{i=1}^n p^{x_i} (1-p)^{1-x_i} = \sum_{i=1}^n \ln p^{x_i} (1-p)^{1-x_i} = \\ &= \sum_{i=1}^n (x_i \ln p + (1-x_i) \ln(1-p)) = \ln p \sum_{i=1}^n x_i + \left(n - \sum_{i=1}^n x_i \right) \ln(1-p). \end{aligned}$$

Оттука, системот равенки е од видот:

$$\begin{aligned} \frac{d}{dp} \ln L(x_1, \dots, x_n, p) &= \frac{d}{dp} \left(\ln p \sum_{i=1}^n x_i + \left(n - \sum_{i=1}^n x_i \right) \ln(1-p) \right) = \\ &= \frac{d}{dp} \left(\ln p \sum_{i=1}^n x_i \right) + \frac{d}{dp} \left(\left(n - \sum_{i=1}^n x_i \right) \ln(1-p) \right) = \\ &= \frac{1}{p} \sum_{i=1}^n x_i + \frac{1}{1-p} \frac{d}{dp} (1-p) \left(n - \sum_{i=1}^n x_i \right) = \frac{1}{p} \sum_{i=1}^n x_i - \frac{1}{1-p} \left(n - \sum_{i=1}^n x_i \right) = 0. \end{aligned}$$

Оттука со решавање на равенката погоре:

$$\frac{\sum_{i=1}^n X_i}{p} = \frac{\left(n - \sum_{i=1}^n X_i\right)}{1-p} \Rightarrow (1-p) \sum_{i=1}^n X_i = \left(n - \sum_{i=1}^n X_i\right) p$$

$$\sum_{i=1}^n X_i - p \sum_{i=1}^n X_i = np - p \sum_{i=1}^n X_i \Rightarrow np = \sum_{i=1}^n X_i,$$

статистиката на точкестотиот оценувач за параметарот p изнесува:

$$\tilde{p} = \frac{\sum_{i=1}^n X_i}{n}.$$

Задача 2.2.5: Земен е примерок од 1000 RAM-чипови независно произведени, со веројатност p за прифатлив чип. Со тестирање на примерокот, 921-RAM чипови се прифатливи. Да се одреди веројатноста p за прифатлив чип.

Решение: Прво да забележиме дека обележјето дека чипот е прифатлив е Бернулиева распределба со параметар p . Од

задачата 2.2.4, статистиката за параметарот p е: $\tilde{p} = \frac{\sum_{i=1}^n X_i}{n}$.

Оттука, точкестата оценка на параметарот p е $p = \frac{921}{1000} = 0,921$.

Значи, веројатноста за RAM прифатлив чип во овој процес на производство изнесува 92,1 %.

2.3. МЕТОД НА МОМЕНТИ

Друг метод што се употребува за наоѓање на точкести оценки е методот на моменти. Пресметувањата на оценките со овој метод се поедноставни од методот на максимална веродостојност, но оценките добиени со овој метод со помалку ефикасни.

Имено, ако X е случајна променлива, тогаш за очекуваната вредност $E(X^k)$ велиме дека е k -тиот момент на случајната променлива. Ако постои k -тиот момент на случајната променлива, тогаш k -тиот централен момент на примерокот $X_1, X_2, \dots, X_n$ е

статистиката: $E(X - E(X))^k = \frac{1}{n} \sum_{i=1}^n (X_i - E(X_i))^k$ и за реализација

на примерокот $x_1, x_2, \dots, x_n$ е бројот: $\bar{x}^k = \frac{1}{n} \sum_{i=1}^n (x_i - E(X_i))^k$.

Нека распределбата на обележјето X е дадена со густината на распределба $p(x; \theta_1, \theta_2, \dots, \theta_k)$ или во дискретен случај со законот на распределба $P(x; \theta_1, \theta_2, \dots, \theta_k)$, $x = x_1, x_2, \dots$; каде што $\theta_1, \theta_2, \dots, \theta_k$ се непознати параметри. Да се најдат оценките на параметрите $\theta_1, \theta_2, \dots, \theta_k$; врз основа на примерок со големина n , $X_1, X_2, \dots, X_n$, значи дека треба да се најдат статистики:

$$\tilde{\theta}_1 = \tilde{\theta}_1(X_1, X_2, \dots, X_n)$$

$$\tilde{\theta}_2 = \tilde{\theta}_2(X_1, X_2, \dots, X_n)$$

.....

$$\tilde{\theta}_k = \tilde{\theta}_k(X_1, X_2, \dots, X_n)$$

врз основа на кои за реализација на примерокот $x_1, x_2, \dots, x_n$ се добиваат нумеричките вредности:

$$v_1 = \tilde{\theta}_1(x_1, x_2, \dots, x_n)$$

$$v_2 = \tilde{\theta}_2(x_1, x_2, \dots, x_n)$$

.....

$$v_k = \tilde{\theta}_k(x_1, x_2, \dots, x_n)$$

кои што претставуваат точкестите оценки на непознатите параметри $\theta_1, \theta_2, \dots, \theta_k$.

Со методот на моменти овие статистики се добиваат како решенија на системот равенки:

$$f_1(X; \theta_1, \theta_2, \dots, \theta_k) = g_1(X_1, X_2, \dots, X_n)$$

$$f_2(X; \theta_1, \theta_2, \dots, \theta_k) = g_2(X_1, X_2, \dots, X_n),$$

.....

$$f_k(X; \theta_1, \theta_2, \dots, \theta_k) = g_k(X_1, X_2, \dots, X_n)$$

каде што $f_i, i=1,2,\dots,k$ претставуваат теориските централни моменти на обележјето X , на пример математичко очекување, дисперзија, додека $g_i, i=1,2,\dots,k$ се соодветните централни моменти на примерокот.

Задача 2.3.1. Да се определат точкестите оценки на обележјето X кое има двојна Пуасонова распределба, во однос на примерокот со вредности дадени во табелата:

x_i	0	1	2	3	4	5	6	7	8	9
n_i	15	13	20	16	12	9	6	3	5	1

Законот на двојната Пуасонова распределба е од обликот:

$$P(X=x) = \frac{1}{2} \frac{\alpha^x}{x!} e^{-\alpha} + \frac{1}{2} \frac{\beta^x}{x!} e^{-\beta}, x=0,1,2,\dots$$

Решение. Бидејќи се прави оценка на два параметри треба да се реши системот од две равенки од обликот:

$$\begin{cases} \mu = \bar{X} \\ \sigma^2 = S^2 \end{cases}$$

каде што μ е математичкото очекување на случајната променлива X , а σ^2 е дисперзијата на случајната променлива X која има двојна Пуасонова распределба, додека $\bar{X}$ е аритметичката средина на примерокот, а S^2 неговата дисперзија.

Математичкото очекување на случајната променлива која има двојна Пуасонова распределба е:

$$\begin{aligned} \mu &= \sum_{x=0}^{+\infty} xP(X=x) = \sum_{x=0}^{+\infty} x \left(\frac{1}{2} \frac{\alpha^x}{x!} e^{-\alpha} + \frac{1}{2} \frac{\beta^x}{x!} e^{-\beta} \right) = \\ &= \sum_{x=0}^{+\infty} x \frac{1}{2} \frac{\alpha^x}{x!} e^{-\alpha} + \sum_{x=0}^{+\infty} x \frac{1}{2} \frac{\beta^x}{x!} e^{-\beta} = \frac{1}{2} e^{-\alpha} \sum_{x=1}^{+\infty} x \frac{\alpha^x}{x!} + \frac{1}{2} e^{-\beta} \sum_{x=1}^{+\infty} x \frac{\beta^x}{x!} = \\ &= \frac{1}{2} e^{-\alpha} \sum_{x=1}^{+\infty} x \frac{\alpha^x}{x(x-1)!} + \frac{1}{2} e^{-\beta} \sum_{x=1}^{+\infty} x \frac{\beta^x}{x(x-1)!} = \end{aligned}$$

$$\begin{aligned} \frac{1}{2}e^{-\alpha}\sum_{x=1}^{+\infty}\frac{\alpha^x}{(x-1)!} + \frac{1}{2}e^{-\beta}\sum_{x=1}^{+\infty}\frac{\beta^x}{(x-1)!} &= \frac{1}{2}e^{-\alpha}\alpha\sum_{x=1}^{+\infty}\frac{\alpha^{x-1}}{(x-1)!} + \frac{1}{2}e^{-\beta}\beta\sum_{x=1}^{+\infty}\frac{\beta^{x-1}}{(x-1)!} = \\ \frac{1}{2}e^{-\alpha}\alpha\sum_{x=0}^{+\infty}\frac{\alpha^x}{x!} + \frac{1}{2}e^{-\beta}\beta\sum_{x=0}^{+\infty}\frac{\beta^x}{x!} &= \frac{1}{2}e^{-\alpha}\alpha e^{\alpha} + \frac{1}{2}e^{-\beta}\beta e^{\beta} = \frac{\alpha + \beta}{2}. \end{aligned}$$

Бидејќи,

$$\begin{aligned} E(X^2) &= \sum_{x=0}^{+\infty} x^2 \cdot P(X=x) = \sum_{x=0}^{+\infty} x^2 \left(\frac{1}{2} \frac{\alpha^x}{x!} e^{-\alpha} + \frac{1}{2} \frac{\beta^x}{x!} e^{-\beta} \right) = \\ \sum_{x=0}^{+\infty} x^2 \frac{1}{2} \frac{\alpha^x}{x!} e^{-\alpha} + \sum_{x=0}^{+\infty} x^2 \frac{1}{2} \frac{\beta^x}{x!} e^{-\beta} &= \frac{1}{2} e^{-\alpha} \sum_{x=1}^{+\infty} x^2 \frac{\alpha^x}{x!} + \frac{1}{2} e^{-\beta} \sum_{x=1}^{+\infty} x^2 \frac{\beta^x}{x!} = \\ \frac{1}{2} e^{-\alpha} \sum_{x=1}^{+\infty} x^2 \cdot \frac{\alpha^x}{x(x-1)!} + \frac{1}{2} e^{-\beta} \sum_{x=1}^{+\infty} x^2 \cdot \frac{\beta^x}{x(x-1)!} &= \\ \frac{1}{2} e^{-\alpha} \sum_{x=1}^{+\infty} x \frac{\alpha^x}{(x-1)!} + \frac{1}{2} e^{-\beta} \sum_{x=1}^{+\infty} x \frac{\beta^x}{(x-1)!} &= \\ \frac{1}{2} e^{-\alpha} \alpha \sum_{x=1}^{+\infty} x \cdot \frac{\alpha^{x-1}}{(x-1)!} + \frac{1}{2} e^{-\beta} \beta \sum_{x=1}^{+\infty} x \cdot \frac{\beta^{x-1}}{(x-1)!} &= \\ \frac{1}{2} e^{-\alpha} \alpha \sum_{x=0}^{+\infty} (x+1) \frac{\alpha^x}{x!} + \frac{1}{2} e^{-\beta} \beta \sum_{x=0}^{+\infty} (x+1) \frac{\beta^x}{x!} &= \\ \frac{1}{2} e^{-\alpha} \alpha \sum_{x=0}^{+\infty} x \frac{\alpha^x}{x!} + \frac{1}{2} e^{-\alpha} \alpha \sum_{x=0}^{+\infty} \frac{\alpha^x}{x!} + \frac{1}{2} e^{-\beta} \beta \sum_{x=0}^{+\infty} x \frac{\beta^x}{x!} + \frac{1}{2} e^{-\beta} \beta \sum_{x=0}^{+\infty} (x+1) \frac{\beta^x}{x!} &= \\ \frac{1}{2} e^{-\alpha} \alpha^2 \sum_{x=0}^{+\infty} \frac{\alpha^x}{x!} + \frac{1}{2} e^{-\alpha} \alpha e^{\alpha} + \frac{1}{2} e^{-\beta} \beta^2 \sum_{x=0}^{+\infty} \frac{\beta^x}{x!} + \frac{1}{2} e^{-\beta} \beta e^{\beta} &= \frac{\alpha^2 + \alpha + \beta^2 + \beta}{2}. \end{aligned}$$

Дисперзијата на случајната променлива која има двојна Пуасонова распределба е:

$$\begin{aligned} V(X) &= E(X^2) - (E(X))^2 = \frac{\alpha + \alpha^2 + \beta + \beta^2}{2} - \left(\frac{\alpha + \beta}{2} \right)^2 = \\ \frac{2\alpha + 2\alpha^2 + 2\beta + 2\beta^2 - \alpha^2 - 2\alpha\beta - \beta^2}{4} &= \frac{2\alpha + 2\beta + \alpha^2 - 2\alpha\beta + \beta^2}{4} \\ \frac{2\alpha + 2\beta}{4} + \frac{\alpha^2 - 2\alpha\beta + \beta^2}{4} &= \frac{\alpha + \beta}{2} + \left(\frac{\alpha - \beta}{2} \right)^2, \end{aligned}$$

Односно:

$$\mu = \frac{\alpha + \beta}{2}, \sigma^2 = \frac{\alpha + \beta}{2} - \left(\frac{\alpha - \beta}{2} \right)^2,$$

од каде што пресметуваме дека:

$$\alpha = \mu + \sqrt{\sigma^2 - \mu}, \beta = \mu - \sqrt{\sigma^2 - \mu}.$$

Ако ги изедначиме μ и σ^2 со аритметичката средина и дисперзијата на примерокот, тогаш оценките на параметрите α и β ќе бидат од обликот:

$$\tilde{\alpha} = \bar{X} + \sqrt{S^2 - \bar{X}}, \tilde{\beta} = \bar{X} - \sqrt{S^2 - \bar{X}}.$$

Врз основа на податоците од примерокот ги добиваме следните точкести оценки: $\alpha = 4,48$, $\beta = 1,52$.

2.4. ОЦЕНУВАЊЕ НА МАТЕМАТИЧКОТО ОЧЕКУВАЊЕ НА ПОПУЛАЦИЈА

Да претпоставиме дека го оценуваме просечниот личен доход во основното образование во Македонија, и дека во примерок од 250 вработени лица во образованието, просечниот личен доход изнесува 40000 денари. Ако ја прифатиме оваа вредност како оценка на вредноста на параметарот просечен личен доход, скоро е сигурно дека ќе згрешиме, бидејќи аритметичката средина на примерокот е блиска, но ретко еднаква на средната вредност, односно математичкото очекување на популацијата.

За да бидеме посигурни во точноста на заклучокот подобро е наместо со една вредност да ги оцениме параметрите на популацијата, како математичко очекување и дисперзија со интервал од вредности. Таквата оценка се нарекува **интервална оценка**, а интервалот се нарекува **интервал на доверба**. Овој интервал би требало да биде доволно голем, така што имајќи ги предвид случајните колебања на параметрите на примероците околу параметрите на популацијата, тој треба да ја опфаќа и нивната реална вредност. Притоа, ако интервалот е многу голем, тој не е информативен. Од друга страна, ако формираме мал интервал, тој може да не го содржи параметарот на популацијата, па постои ризик изведениот заклучок да е погрешен. Сепак, врз основа на резултатите од теоријата на веројатност, големината на овој ризик може да се контролира, доколку избереме случаен примерок со

помош на една реална вредност која се задава однапред, т.н **ниво на доверба** или **веројатност на доверба** β .

За да можеме да го оцениме непознатиот параметар на популацијата со примена на некоја параметарски метод на оценување, неопходно е да бидат исполнети некои строги претпоставки за карактеристиките на основното множество. На пример, при оценувањето на математичкото очекување на популацијата од посебно значење се две информации: каков е обликот на распределбата на веројатности на основното множество и дали е позната дисперзијата. Во зависност од овие информации се применуваат различни методи за оценување на параметарот математичко очекување.

2.4.1. ИНТЕРВАЛ НА ДОВЕРБА ЗА МАТЕМАТИЧКО ОЧЕКУВАЊЕ μ

КОГА Е ПОЗНАТА ДИСПЕРЗИЈАТА НА ПОПУЛАЦИЈАТА σ^2

Популацијата има нормална распределба на веројатност

Нека претпоставиме дека треба да го оцениме математичкото очекување на популацијата која има нормална распределба на веројатност $X : N(\mu, \sigma^2)$, при што е позната нејзината дисперзија σ^2 .

Бидејќи популацијата има нормална распределба, тогаш аритметичката средина $\bar{X}$ на примерокот со големина n земен од таа популација, има нормална распределба со параметри μ и $\frac{\sigma^2}{n}$,

односно $\bar{X} : N\left(\mu, \frac{\sigma^2}{n}\right)$. Оттука, статистиката $Z = \frac{\bar{X} - \mu}{\sigma} \sqrt{n}$ има

$N(0,1)$ распределба.

Ако нивото на доверба е β , тогаш со помош на таблицата за нормална распределба наоѓаме реален број z_β таков што:

$$P\left(-z_\beta \leq \frac{\bar{X} - \mu}{\sigma} \sqrt{n} \leq z_\beta\right) = \beta,$$

односно:

$$P\left(-z_{\beta} \leq \frac{\bar{X} - m}{\sigma} \sqrt{n} \leq z_{\beta}\right) = \Phi_0(z_{\beta}) - \Phi_0(-z_{\beta}) = \\ \Phi_0(z_{\beta}) - (1 - \Phi_0(z_{\beta})) = 2\Phi_0(z_{\beta}) - 1 = \beta$$

оттука:

$$\Phi_0(z_{\beta}) = \frac{1 + \beta}{2}.$$

Со решавање на неравенството во заградата по $\bar{X}$, го добиваме следното равенство:

$$P\left(\mu - z_{\beta} \frac{\sigma}{\sqrt{n}} \leq \bar{X} \leq \mu + z_{\beta} \frac{\sigma}{\sqrt{n}}\right) = \beta.$$

Ако, на пример, нивото на доверба изнесува $\beta = 0,95$, $\Phi_0(z_{\beta}) = \frac{1 + 0,95}{2} = 0,975$ тогаш $z_{\beta} = 1,96$. Затоа, равенството $P\left(m - 1,96 \frac{\sigma}{\sqrt{n}} \leq \bar{X} \leq m + 1,96 \frac{\sigma}{\sqrt{n}}\right) = 0,95$ го толкуваме на следниов начин: „Во 95 % од случајните примероци со големина n , нивната аритметичка средина отстапува од математичкото очекување на популацијата m за најмногу од $1,96$ стандардни грешки $\frac{\sigma}{\sqrt{n}}$ “.

Нека го избереме кој било од овие 95 % примероци, ја пресметаме неговата аритметичка средина и нека го одредиме околу неа следниов интервал $\left[\bar{x} - 1,96 \frac{\sigma}{\sqrt{n}}, \bar{x} + 1,96 \frac{\sigma}{\sqrt{n}}\right]$. Бидејќи $\bar{x}$ отстапува од μ за помалку од $1,96 \frac{\sigma}{\sqrt{n}}$, во пресметаниот интервал ќе се најде вредноста на параметарот μ . Значи, сите 95 % вака формирани интервали ќе ја содржат вредноста μ . Аритметичките средини $\bar{x}$ во другите 5 % примероци ќе се разликуваат од μ за повеќе од $1,96 \frac{\sigma}{\sqrt{n}}$, затоа интервалите околу овие вредности не би го содржеле параметарот μ . Овој заклучок попрецизно ќе го формулираме со следново равенство:

$$P\left(\bar{X} - 1,96 \frac{\sigma}{\sqrt{n}} \leq m \leq \bar{X} + 1,96 \frac{\sigma}{\sqrt{n}}\right) = 0,95.$$

Веројатноста дека интервалот околу аритметичката средина на примерокот ќе ја содржи вредноста на параметарот m изнесува 0,95.

Овој резултат се користи при оценувањето на непознатото математичко очекување на популацијата. Оценувањето се спроведува со помош на еден случаен примерок и неговата аритметичка средина $\bar{x}$, околу која се формира интервалот

$$\left[\bar{x} - z_{\beta} \frac{\sigma}{\sqrt{n}}, \bar{x} + z_{\beta} \frac{\sigma}{\sqrt{n}}\right],$$

кој се нарекува **интервал на доверба за математичкото очекување на популацијата**.

Задача 2.4.1. Од популацијата со распределба $N(\mu, 100)$ земен е примерок со големина $n = 25$ и пресметана е средината на примерокот $\bar{x} = 120$. Ако нивото на доверба изнесува $\beta = 0,95$, тогаш интервалот на доверба за математичкото очекување на популацијата изнесува:

$$\left[120 - 1,96 \frac{10}{\sqrt{25}}, 120 + 1,96 \frac{10}{\sqrt{25}}\right], \text{ односно } [116,08; 123,92].$$

Заклучуваме дека, со веројатност од 0,95 математичкото очекување на популацијата е еднакво на една вредност од интервалот $[116,08; 123,92]$. Притоа, постои ризик од 0,05 дека вредноста на математичкото очекување на популацијата е помала од 116,08 или поголема од 123,92.

Ризикот дека при заклучувањето сме направиле грешка се нарекува **грешка на ризик** и се обележува со α и со нивото на доверба се дополнуваат до 1, значи: $\alpha = 1 - \beta$. Често пати грешката на ризик и нивото на доверба се изразуваат во проценти, имено наместо да кажеме вредноста на нивото на доверба изнесува 0,95, велиме дека нивото на доверба е 95 %.

Избор на нивото на доверба

Математичкото очекување може да се оценува со различни нивоа на доверба, а не само со 0,95. Бидејќи крајните вредности на

интервалот на доверба зависат од нивото на доверба, тогаш со промена на неговата вредност се менува и ширината на интервалот. Имено, како расте нивото на доверба, така расте и ширината на интервалот на доверба, па така ако сакаме да изведеме заклучок во чија точност сме 100 % сигурни, тогаш формиранiot интервал мора да биде широк, но тоа значи дека заклучокот не е прецизен. Со други зборови, *со растење на нивото на доверба, се намалува прецизносата на оценката*. Понекогаш интервалот е толку широк, така што резултатот ја губи целосната информативност и практичната корист. Затоа, во пракса најчесто се користи ниво на доверба од 95% бидејќи тогаш истовремено се постигнува висока доверба и релативно прецизна оценка.

Зависност на нивото на доверба и големината на примерокот

Нека претпоставиме дека во задачата 2.4.1, наместо примерок со големина $n = 25$ се разгледува примерок со големина $n = 10000$, и притоа аритметичката средина на овој примерок е еднаква на аритметичката средина на претходниот, т. е. $\bar{x} = 120$. Тогаш, за истото ниво на доверба $\beta = 0,95$, интервалот на доверба изнесува:

$$\left[120 - 1,96 \frac{10}{\sqrt{10000}}; 120 + 1,96 \frac{10}{\sqrt{10000}} \right] = [119,804; 120,196].$$

Ако го споредиме со претходно пресметаниот интервал, забележуваме дека:

за $n = 25$ ширината на интервалот изнесува $123,92 - 116,08 = 7,84$,
за $n = 100$ изнесува $120,196 - 119,804 = 0,392$.

Значи, при одредено ниво на доверба, интервалот на доверба можеме да го намалиме ако го зголемиме примерокот. Со други зборови, *за дадено ниво на доверба, прецизносата на заклучокот се зголемува со зголемување на примерокот*.

Често пати, и нивото на доверба и прецизносата се однапред дадени. Па така потребно е да ја одредиме онаа големина на примерокот со која се задоволени и двете вредности. Бидејќи со растење на големината на примерокот се зголемуваат и трошоците при истражувањето по математички пат може да се пресмета *најмалата големина на примерокот* со која се обезбедуваат однапред дадената прецизност и доверба на оценката.

Популацијата нема нормална распределба на веројатност

Ако популацијата со позната дисперзија σ^2 нема нормална распределба на веројатност, или обликот на распределбата не е познат, при оценувањето на математичкото очекување m на популацијата се користи *централната гранична теорема*. Од статистичкото множество се избира примерок со поголема големина n и тогаш статистиката $\bar{X}$ има нормална распределба со параметри μ и $\frac{\sigma^2}{n}$. Потоа, оценувањето е идентично како при оценувањето на математичкото очекување на популација со нормална распределба.

Забелешка. Оваа оценка е соодветна само кога за големината на примерокот важи $n \geq 30$. Во случај кога популацијата нема нормална распределба и примерокот има големина помала од 30 не постои статистика за пресметување интервал на доверба.

2.4.2. ИНТЕРВАЛ НА ДОВЕРБА ЗА МАТЕМАТИЧКОТО ОЧЕКУВАЊЕ μ КОГА Е НЕПОЗНАТА ДИСПЕРЗИЈАТА НА ПОПУЛАЦИЈАТА

Претпоставката дека при оценувањето на параметарот μ ја знаеме вредноста на дисперзијата σ^2 , по правило не е исполнета. Во пракса, всушност ретко се случува да не е познато математичкото очекување на популацијата, а во исто време да е позната нејзината дисперзија.

Ако не е позната дисперзијата σ^2 , тогаш не можеме да ја пресметаме стандардната грешка $\frac{\sigma}{\sqrt{n}}$, а без неа не може да се одреди интервалот на доверба на математичкото очекување на популацијата. Затоа, прво е потребно да се оцени непозната вредност σ . Кога во формулата за стандардната грешка $\frac{\sigma}{\sqrt{n}}$, ќе се замени стандардната девијација на популацијата σ со нејзината точкеста оценка, односно со непристрасната статистика:

$$S = \sqrt{\frac{1}{n-1} \left(\sum_{i=1}^n X_i^2 - n\bar{X}^2 \right)},$$

тогаш се добива *оценката на стандардната грешка*:

$$\frac{S}{\sqrt{n}}.$$

Реализираната вредност на оваа оценка со вредностите од примерокот што се разгледува, односно точката вредност на оценката на стандардната грешка изнесува $\frac{s}{\sqrt{n}}$, каде што s претставува средно квадратно отстапување на примерокот, односно:

$$s = \sqrt{\frac{1}{n-1} \left(\sum_{i=1}^n x_i^2 - n\bar{x}^2 \right)}.$$

Во случај кога е позната дисперзијата на популацијата, статистиката $Z = \frac{\bar{X} - \mu}{\sigma} \sqrt{n}$ има $N(0,1)$ распределба, но кога во неа ќе се замени стандардната девијација на популацијата σ со стандардна девијација на примерокот, тогаш се добива следната статистика $t = \frac{\bar{X} - \mu}{s} \sqrt{n}$ и има т.н. Студентова распределба или t -распределба.

Значи кога се одредува интервалот на доверба на математичкото очекување на популацијата ако не е позната нејзината дисперзија при примерок со големина n и ниво на доверба β , се разгледува статистиката $t = \frac{\bar{X} - \mu}{s} \sqrt{n}$ која има t -распределба со $n-1$ степени на слобода. За даденото ниво на грешка β , ризикот изнесува $\alpha = 1 - \beta$ и за најденото α определуваме реален број $t_\alpha(n-1)$, таков што:

$$P\left(\left|\frac{\bar{X} - \mu}{s} \sqrt{n}\right| > t_\alpha(n-1)\right) = \alpha,$$

односно:

$$P\left(\left|\frac{\bar{X}-m}{s}\sqrt{n}\right|\leq t_{1-\beta}(n-1)\right)=1-\alpha=\beta.$$

Оттука, важи:

$$P\left(-t_{1-\beta}(n-1)\leq\frac{\bar{X}-m}{s}\sqrt{n}\leq t_{1-\beta}(n-1)\right)=\beta.$$

Решавајќи ја неравенката во претходното равенство по μ , при што се зема вредноста на точкестата оценка $\bar{x}$ на математичкото очекување $\bar{X}$ пресметана со реализирани вредности на разгледуваниот примерок се добива интервалот на доверба на математичкото очекување на популацијата во следниот облик:

$$\left[\bar{x}-\frac{s}{\sqrt{n}}t_{1-\beta}(n-1), \bar{x}+\frac{s}{\sqrt{n}}t_{1-\beta}(n-1)\right],$$

каде што бројот $t_{1-\beta}(n-1)$ се чита од табелата на Студентова распределба на двостран тест.

Забелешка. Интервалот на доверба за математичкото очекување μ кога е непозната дисперзијата на популацијата, а популацијата има нормална распределба се пресметува со претходниот израз кога за големината на примерокот важи дека $n < 30$. Во случај кога големината на примерокот е поголема или еднаква на 30, бидејќи со растењето на степените на слобода на t -распределбата, таа се приближува на нормалната распределба, тогаш на статистиката $\frac{\bar{X}-m}{s}\sqrt{n}$ соодветствува $N(0,1)$ распределба.

Затоа интервалот на доверба за математичкото очекување m кога е непозната дисперзијата на популацијата, а за големината на примерокот важи дека $n \geq 30$ попрецизно се пресметува со изразот:

$$\left[\bar{x}-z_{\beta}\frac{s}{\sqrt{n}}, \bar{x}+z_{\beta}\frac{s}{\sqrt{n}}\right],$$

каде што реалниот број z_{β} е број прочитан од таблицата за нормална распределба за кој важи:

$$\Phi_0(z_{\beta})=\frac{1+\beta}{2}=\frac{1+1-\alpha}{2}=1-\frac{\alpha}{2}.$$

Задача 2.4.2. Една законодавна комисија сака да провери дали има значајна разлика во даночните приходи помеѓу предложениот нов даночен закон и постојниот даночен закон. Комисијата по случаен избор избира 100 репрезентативни даночни пријави. За тие пријави, таа го пресметува данокот $\bar{y}$ користејќи го предложениот нов даночен закон и го споредува со данокот $\bar{x}$ платен според постоечкиот закон. Комисијата потоа ја пресметува просечната разлика помеѓу овие два закона, таа изнесува -219, и има стандардно отстапување 725. Што може да заклучи законодавната комисија врз основа на овој примерок, со прифатлив ризик од 0,01?

Решение. Разгледуваме зависни примероци со големини $n_1 = n_2 = 100 > 30$, аритметички средина $\bar{z}_n = -219$ и стандардно отстапување $s_n = 725$ на променливата $Z = X - Y$. Она што сакаме да го провериме е дали постои значајна разлика во даночните приходи помеѓу предложениот нов даночен закон и постојниот даночен закон (алтернативна хипотеза), затоа употребуваме z -тест за следната хипотеза:

$$H_0 : m_1 = m_2 \text{ наспроти алтернативната хипотеза } H_1 : m_1 < m_2 ,$$

а статистиката за тестот е $Z = \frac{\bar{Z}_n}{S_z} \sqrt{n}$.

Задача 2.4.3. Разгледуваме примерок со големина $n = 100$. Најдете 90 % интервал на доверба на средната вредност на популацијата од која е земен примерокот, ако аритметичката средина на примерокот е 20, а варијансата на популацијата е 16.

Решение. Бидејќи:

$$\bar{x} = 20, \sigma^2 = 16, \sigma = \sqrt{16} = 4 \text{ и}$$

$$\Phi(1,64) < \Phi(z_\beta) = 1 - \frac{\alpha}{2} = 1 - \frac{0,1}{2} = 1 - 0,05 = 0,95 < \Phi(1,65) ,$$

$$z_\beta = \frac{1,64 + 1,65}{2} = 1,645 .$$

Тогаш, 90% интервал на доверба на средната вредност на популацијата е интервалот

$$\left(\bar{x} - \frac{\sigma}{\sqrt{n}} z_{\beta}, \bar{x} + \frac{\sigma}{\sqrt{n}} z_{\beta} \right) = \left(20 - \frac{4}{\sqrt{100}} \cdot 1,645; 20 + \frac{4}{\sqrt{100}} \cdot 1,645 \right) \\ (20 - 0,658; 20 + 0,658) = (19,342; 20,658) .$$

Задача 2.4.4. Разгледуваме примерок од податоци со големина N_1 и откриваме дека 90 % интервал на доверба на математичкото очекување на распределбата на популацијата од која е земен има ширина, w . Која треба да биде големината на примерокот N_2 за да се зголеми нивото на доверба на 99,9%, а сепак да се одржи истата ширина на интервалот w ?

Решение. Бидејќи $z_{0,9} = \frac{1,64 + 1,65}{2} = 1,645$, 90% интервалот на доверба е:

$$\left(\bar{x}_{N_1} - \frac{\sigma}{\sqrt{N_1}} z_{0,9}, \bar{x}_{N_1} + \frac{\sigma}{\sqrt{N_1}} z_{0,9} \right) = \left(\bar{x}_{N_1} - 1,645 \frac{\sigma}{\sqrt{N_1}}, \bar{x}_{N_1} + 1,645 \frac{\sigma}{\sqrt{N_1}} \right)$$

Оттука, неговата ширина изнесува:

$$w = \bar{x}_{N_1} + 1,645 \frac{\sigma}{\sqrt{N_1}} - \left(\bar{x}_{N_1} - 1,645 \frac{\sigma}{\sqrt{N_1}} \right) = 2 \cdot 1,645 \frac{\sigma}{\sqrt{N_1}} = 3,29 \frac{\sigma}{\sqrt{N_1}}$$

Бидејќи $\Phi(z_{0,99}) = 1 - \frac{0,001}{2} = 0,9995 = \Phi(3,27)$, тогаш 99 % интервал на доверба е:

$$\left(\bar{x} - \frac{\sigma}{\sqrt{N_2}} z_{0,99}, \mu_X + \frac{\sigma}{\sqrt{N_2}} z_{0,99} \right) = \left(\mu - 3,27 \frac{\sigma}{\sqrt{N_2}}, \mu + 3,27 \frac{\sigma}{\sqrt{N_2}} \right)$$

$$\text{односно неговата ширина е } w = 2 \cdot 3,27 \frac{\sigma}{\sqrt{N_2}} = 6,54 \frac{\sigma}{\sqrt{N_2}} .$$

За да овие интервали на доверба имаат иста ширина, неопходно е:

$$2 \cdot 1,645 \frac{\sigma}{\sqrt{N_1}} = 2 \cdot 3,27 \frac{\sigma}{\sqrt{N_2}} / : 2\sigma \Rightarrow \\ 1,645 \frac{1}{\sqrt{N_1}} = 3,27 \frac{1}{\sqrt{N_2}} \Rightarrow \sqrt{\frac{N_2}{N_1}} = \frac{3,27}{1,645} \Rightarrow \frac{N_2}{N_1} = \left(\frac{3,27}{1,645} \right)^2 = 3,95$$

Ова значи дека големината на примерокот N_2 треба да биде околу 4 пати поголема од N_1 .

Задача 2.4.5. Една компанија произведува матични плочи за компјутери. Притоа, од нивни интерес е средното време до откажување на матичните плочи, кое тие редовно го мерат. Определете го бројот на матичните плочи чие време на откажување мора да се измери пред нивното средно време до откажување да биде во рамките на 20% од точното со 90 % веројатност. Да претпоставиме дека распределбата на времето X до откажување има експоненцијална распределба (т.е. $p(x) = \left(\frac{1}{\mu}\right)e^{-\frac{x}{\mu}}$, ако $x > 0$).

Решение. Бидејќи времето до откажување на матичните плочи има експоненцијална распределба тогаш:

$$E(X) = \mu, \quad V(X) = \sigma^2 = \mu^2.$$

Половина од ширината на 90 % интервал на доверба треба да изнесува $20\% \mu = 0,2\mu$.

Од друга страна, половина од ширината на 90 % интервал на доверба треба да изнесува $\frac{\sigma}{\sqrt{n}} z_{0,9} = 1,645 \frac{\mu}{\sqrt{n}}$. Оттука,

$$0,2\mu = 1,645 \frac{\mu}{\sqrt{n}} \Rightarrow 0,2\mu = 1,645 \frac{\mu}{\sqrt{n}} : \mu \Rightarrow n = 8,225^2 = 67,65.$$

Значи, треба да се измери времето на откажување на 68 матични плочи.

3. ТЕСТИРАЊЕ ХИПОТЕЗИ

Во една област на математичката статистика, таканаречена тестирање на хипотези, информациите од примерокот се употребуваат за да се провери дали некое тврдење или претпоставка кое се однесува на некоја карактеристика на популацијата може да се прифати. Најчесто се разгледува некоја вредност на некој параметар на популацијата, обликот на распределбата на статистичкото множество, разликата меѓу вредностите на ист параметар на повеќе статистички множества, зависноста меѓу две или повеќе обележја на статистичкото множество и слично. На пример, со тестирање може да се провери дали е точно тврдењето дека просечната должина на даден производ отстапува од пропишаните вредности, дали просечниот личен доход на вработените во образование и култура меѓу себе се разликуваат, дали популацијата има рамномерна распределба, од кој тип е заемната поврзаност на рекламите и продажбата и слично.

Точноста на овие претпоставки и тврдења можат да се одредат само со целосно опфаќање на сите елементи од разгледуваните множества, но бидејќи тоа обично е тешко да се изведе, посебно заради обемноста на множествата, постапката за тестирање на т.н статистички хипотези ја изведуваме со испитување врз основа на примерок. **Статистичка хипотеза** е прецизно формулирано тврдење или претпоставка за некоја карактеристика на популацијата. Научниот метод со кој се проверува прифатливоста на ова тврдење или претпоставка се нарекува **тестирање статистички хипотези**.

Бидејќи заклучокот за прифаќањето на претпоставката го донесуваме врз основа на информациите од примерокот, при тестирањето е присутен *ризик* дека ќе згрешиме во заклучокот. Во оваа глава ќе се задржиме само на тестирањето хипотези кои се однесуваат на вредности на параметрите на популацијата, како и на разликата меѓу вредностите на истиот параметар од две популации.

3.1. МЕТОД ЗА ТЕСТИРАЊЕ СТАТИСТИЧКИ ХИПОТЕЗИ

Класичниот метод за тестирање на статистичките хипотези се изведува во пет чекори:

- 1) Формулирање на нултата и на алтернативната хипотеза;

- 2) избор на статистика за тестот и одредување на нејзина распределба;
- 3) избор на т.н. ниво на значајност α и според него областите на отфрлање или прифаќање на нултата хипотеза;
- 4) донесување одлука дали нултата хипотеза да ја отфрлиме или не;
- 5) донесување заклучок.

3.1.1. НУЛТА И АЛТЕРНАТИВНА ХИПОТЕЗА

Нултата и алтернативната хипотеза претставуваат две прецизни, меѓу себе исклучувачки тврдења или претпоставки за вредностите на популацијата.

Нултата хипотеза (H_0) е тврдење за вредноста на параметарот θ кое настојуваче да го оспориме со постапката за тестирање. Се претпоставува дека е вистина, додека не го докажеме спротивното. Таа може да биде проста или сложена.

Нултата хипотеза е **проста** ако со неа се тврди дека параметарот е еднаков на точно една, однапред позната нумеричка вредност. Хипотезата кратко се запишува во следниов облик:

$H_0: \theta = \theta_0$ и се чита: „Нултата хипотеза гласи дека параметарот θ е еднаков на θ_0 “.

Ако нултата хипотеза опфаќа поголем број вредности, на пример, $H_0: \theta \leq \theta_0$, тогаш е **сложена**.

На секоја нулта хипотеза ѝ придружуваме **алтернативна хипотеза** (H_1). Алтернативната хипотеза најчесто ги содржи сите вредности на параметарот θ , кои не се опфатени со нултата хипотеза. Таа е тврдењето што сакаме да го докажеме како вистинито.

Нултата и алтернативната хипотеза се јавуваат во следниве три облици:

$$H_0: \theta = \theta_0 \text{ наспроти } H_1: \theta \neq \theta_0;$$

$$H_0: \theta \leq \theta_0 \text{ наспроти } H_1: \theta > \theta_0;$$

$$H_0: \theta \geq \theta_0 \text{ наспроти } H_1: \theta < \theta_0.$$

Покрај обликот на хипотезите од посебно значење за исходот од тестирањето е **содржината** на нултата спрема алтернативната хипотеза. Имено, претпоставката која ја сметаме дека е точна и која

со постапката за тестирање настојуваме да ја потврдиме ја прикажуваме во облик на алтернативната хипотеза.

Причината за тоа е следна: во постапката за тестирање на хипотези, H_0 ја отфрламе само ако информацијата од примерокот значајно противречи на тврдењето кое е содржано во нултата хипотеза, во спротивно H_0 ја прифаќаме. Со други зборови, за да ја отфрлиме нултата хипотеза, потребни ни се многу посилни докази од примерокот од оние докази врз основа на кој ја прифаќаме H_0 . Ако врз основа на примерокот заклучиме дека треба да ја отфрлиме H_0 , можеме прилично да бидеме сигурни дека H_1 е вистинита. Наспроти ова, ако примерокот ја потврдува нултата хипотеза, тогаш тоа не значи дека таа е вистинита, туку дека немаме доволно докази против неа. Заради тоа, нашето уверување кое сакаме да го провериме го формулираме во облик на алтернативна хипотеза, па ако податоците од примерокот противречат на H_0 , можеме скоро да бидеме сигурни дека сме во право.

Ако претпоставката е потврдена со информацијата од еден случаен примерок, не можеме потполно да бидеме уверени дека таа навистина е вистинита. Резултатот од примерокот не е доволен за да тврдиме дека алтернативната хипотеза е точна, туку врз основа на него заклучуваме дека H_1 *можеме да ја прифатиме како точна*. Од друга страна, ако податоците од примерокот ја подржуваат нултата хипотеза, заклучуваме дека H_0 *не можеме да ја отфрлиме*.

Иако, нашето уверување го формулираме во облик на алтернативна хипотеза, во постапката за тестирање ја проверуваме само нултата хипотеза. *Тргнуваме од претпоставката дека H_0 е вистинита и настојуваме да ја оспориме оваа претпоставка*. Алтернативната хипотеза не ја проверуваме, туку ја прифаќаме како вистинита ако резултатот од анализата противречи на нултата хипотеза.

3.1.2. ИЗБОР НА СТАТИСТИКА НА ТЕСТОТ И ОДРЕДУВАЊЕ НА НЕЈЗИНАТА РАСПРЕДЕЛБА

Следен чекор при постапката за тестирање хипотези, после прецизното формулирање на H_0 и H_1 е избор на статистика за изведување на тестот и одредување на нејзината распределба на веројатност.

Статистика на тестот е критериум врз основа на кој се реализира тестирањето. Најчесто се користи непристрасната оценка на параметарот θ или некоја нејзина трансформација. Ако, на пример, ја тестираме хипотезата за математичкото очекување на популацијата, како критериум за тестирање се применува точкестата оценка на математичкото очекување $\bar{X}$ или една од нејзините стандардизирани вредности Z и t . Бидејќи оценката на параметрите на популацијата е случајна променлива, тогаш и статистиката на тестот е случајна променлива. Од реализираните вредности од статистиката на тестот врз основа на еден случаен примерок зависи дали ќе ја отфрлиме H_0 или не.

3.1.3. НИВО НА ЗНАЧАЈНОСТ И ОБЛАСТИ НА ОТФРЛАЊЕ И НЕОТФРЛАЊЕ НА НУЛТАТА ХИПОТЕЗА

Нултата хипотеза, како тврдење за вредноста на параметарот на популацијата, е или вистинита или неvistинита, а информацијата од случајниот примерок е согласна со H_0 или ѝ противречи. Оттука, возможно е податоците од примерокот да ја подржуваат вистинитата хипотеза или да бидат противречни на вистинитата хипотеза. Бидејќи одлуката дали да ја отфрлиме или не нултата хипотеза се донесува врз основа на еден случаен примерок, заклучокот од тестирањето може да биде погрешен. Во следната табела 3.1 се прикажани исходите од тестирањето во однос на нивната вистинитост.

		H_0	
		Вистинита	Невистинита
Одлука	H_0 не се отфрла	Правилна одлука ($1 - \alpha$)	Грешка од втор вид (β)
	H_0 се отфрла	Грешка од прв вид (α)	Правилна одлука ($1 - \beta$)

Табела 3.1.3. Исходи од тестирање хипотези и нивна веројатност

Според табелата 3.1.3, ако нултата хипотеза е вистинита, а информацијата од примерокот е согласна со неа, нултата хипотеза не ја отфрламе и заклучокот од тестирањето ќе биде исправен.

Правилно ќе постапиме и ако ја отфрлиме нултата хипотеза која е неvistинитата. Меѓутоа, може да се случи нултата хипотеза да е вистинитата, но сепак врз основа на податоците од примерокот да ја отфрлиме. Тогаш заклучокот е погрешен, а грешката што тогаш сме ја направиле се нарекува **грешка од прв вид**. Друга грешка која може да ја направиме при заклучувањето е да не ја отфрлиме неvistинитата нулта хипотеза. Оваа грешка се нарекува **грешка од втор вид**. Јасно е дека при заклучувањето може да направиме само една грешка. Имено, ако одлучиме да не ја отфрлиме нултата хипотеза, можно е да правиме грешка од втор вид, а ако одлучиме да ја отфрлиме нултата хипотеза, тогаш можно е да правиме грешка од прв вид. При тестирањето на хипотезите настојуваме да ја одредиме веројатноста да се појават наведените грешки и истите ако е можно да се намалат.

Веројатноста дека ќе ја отфрлиме вистинитата нулта хипотеза (веројатноста дека ќе направиме грешка од прв вид), се нарекува **ризик за грешка од прв вид** или **ниво на значајност** и се обележува со α :

$$\alpha = P(H_0 \text{ ја отфрламе} \mid H_0 \text{ е вистинитата}).$$

Веројатноста дека нема да ја отфрлиме неvistинитата нулта хипотеза (веројатноста дека ќе направиме грешка од втор вид) се нарекува **ризик за грешка од втор вид** и се обележува со β :

$$\beta = P(H_0 \text{ не ја отфрламе} \mid H_0 \text{ не е вистинитата}).$$

Во класичната постапка за тестирање хипотези, однапред се избира нивото на значајност α , односно сами ја избираме веројатноста со која може да се случи да ја отфрлиме вистинитата нулта хипотеза. Имено со прифаќање на одредено ниво на значајност на тестот, α , а тоа најчесто се нивоата 0,05 и 0,01, ја прифаќаме можноста да згрешиме и да ја отфрлиме вистинитата нулта хипотеза. Додека веројатноста да донесеме правилна одлука, односно да ја прифатиме вистинитата нулта хипотеза се нарекува **коефициент на доверба** и изнесува $1 - \alpha$.

За даден облик на тестот и големина на примерокот, со изборот на нивото на значајност на тестот α целосно е одредена вредноста на ризикот за грешка од втор вид β . Веројатноста за грешка од втор вид може да биде многу голема, а бидејќи последиците од прифаќање на неvistинитата нулта хипотеза често се неповолни колку и последиците од отфрлање на вистинитата нулта хипотеза, пожелно е покрај контролирањето на ризикот α да се контролира и ризикот β .

Наместо ризикот β при тестирање често се споменува **јачина на тестот**, која претставува веројатноста да се отфрли неистинитата нулта хипотеза и е комплементарна вредност на ризикот β , односно таа е еднаква на вредноста $1 - \beta$.

Значи, ризикот β е одреден со нивото на значајност на тестот. За дадена големина на примерок, со растењето на ризикот α се намалува ризикот β и обратно, но притоа овие два ризика не се дополнуваат до 1, односно не се комплементарни. Единствен начин да ја намалиме веројатноста на една грешка и притоа да не ја зголемиме веројатноста на другата грешка е да ја зголемиме големината на примерокот. За дадена вредност на ризикот α , со растење на големината на примерокот се намалува ризикот β , односно се зголемува јачината на тестот. Значи, со примена на големи примероци ја зголемуваме веројатноста дека ќе ја отфрлиме неистинитата нулта хипотеза.

Кога се познати веројатностите на сите тврдења од табелата 3.1, може да ја пресметаме вкупната грешка на тестот со формулата:

$$\begin{aligned} Vg &= P(H_0 \text{ отфрлена} \mid H_0 \text{ е вистинита}) \cdot P(H_0 \text{ е вистинита}) + \\ &+ P(H_0 \text{ не е отфрлена} \mid H_0 \text{ не е вистинита}) \cdot P(H_0 \text{ не е вистинита}) = \\ &\alpha \cdot P(H_0 \text{ е вистинита}) + \beta \cdot P(H_0 \text{ не е вистинита}). \end{aligned}$$

При реализирањето на постапката за тестирање хипотези, се јавуваат различни вредности на статистиката на тестот, при што некои се помалку веројатни, а некои повеќе веројатни кога нултата хипотеза е вистинита. Во третата фаза на тестирање хипотези се разграничуваат овие вредности на т.н област на отфрлање на H_0 и област на неотфрлање (која условно се нарекува, област на прифаќање) на H_0 . **Критична област (област на отфрлање на нултата хипотеза)**, претставува подмножество од малку веројатни вредности на статистиката на тестот кога нултата хипотеза е вистинита. Ако се реализира една од овие вредности, нултата хипотеза се отфрла. Подмножеството од останатите можни вредности на статистиката на тестот претставува **област на прифаќање на нултата хипотеза**; со реализација на една од овие вредности, нултата хипотеза не се отфрла. Критичната област, нејзината големина и положба, се одредени со: обликот на хипотезата, статистиката на тестот и нејзината распределба, како и нивото на значајност.

3.2. ТЕСТИРАЊЕ ХИПОТЕЗИ ВРЗ ОСНОВА НА ЕДЕН ПРИМЕРОК

3.2.1. ТЕСТОВИ ЗА МАТЕМАТИЧКО ОЧЕКУВАЊЕ

Z-тест

Тестирањето на хипотезите за математичко очекување како што погоре споменавме е засновано на две статистики Z и t . Притоа, статистиката Z не е секогаш оправдана, туку се неопходни некои претпоставки кои се однесуваат на карактеристиките на популацијата, а кои беа претходно наспоменати во постапката за интервално оценување на математичкото очекување. Имено, ако популацијата има нормална распределба, или распределбата е произволна, но примерокот е доволно голем, $n \geq 30$, со што според централно граничната теорема може да сметаме дека неговиот просек има нормална распределба, статистиката Z има стандардизирана нормална распределба $N(0,1)$. Во следната табела 3.2 се претставени условите за примена на Z -тестот и постапката за тестирање хипотези за математичкото очекување на популацијата.

Табела 3.2.1.1. Тестирање на хипотези за математичко очекување со примена на Z тестот

<u>Услови за примена на Z -тестот:</u> кога σ^2 е позната
а. Популацијата има нормална распределба и примерокот има произволна големина n .
б. Распределбата на популацијата не е позната, и примерокот има големина $n \geq 30$.
Избор на обликот на нултата и алтернативната хипотеза:
$H_0 : m = m_0, \quad H_1 : m \neq m_0$
$H_0 : m \leq m_0, \quad H_1 : m > m_0$
$H_0 : m \geq m_0, \quad H_1 : m < m_0$
каде што m_0 е дадена нумеричка вредност.
Статистика за тестот:

$$1) Z = \frac{\bar{X} - m_0}{\sigma} \sqrt{n}$$

Примерокот е земен од бесконечна популација,

Примерокот е земен од конечна популација, при што изборот се врши со враќање.

Примерокот е земен од конечна популација, при што изборот се врши без враќање, но големината на примерокот n е помала од 5% од големината на популацијата N

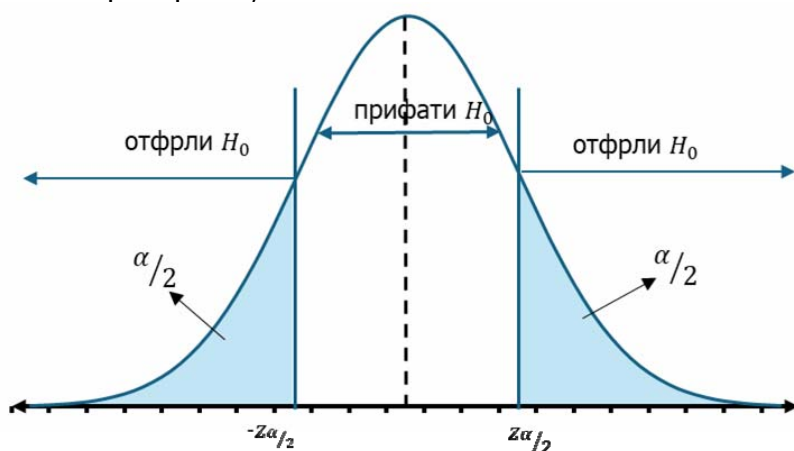
$$2) Z = \frac{\bar{X} - m_0}{\sigma} \sqrt{\frac{n(N-1)}{N-n}}$$

Примерокот е земен од конечна популација, при што изборот се врши без враќање, но големината на примерокот n е поголема од 5 % од големината на популацијата N .

Со ниво на значајност α

A) За хипотезата $H_0 : m = m_0$, наспроти алтернативната хипотеза $H_1 : m \neq m_0$:

i) пресметуваме $\bar{z} = \frac{|\bar{x} - m_0|}{\sigma} \sqrt{n}$ за тестот 1) и $\bar{z} = \frac{|\bar{x} - m_0|}{\sigma} \sqrt{\frac{n(N-1)}{N-n}}$ за тестот 2), каде што $\bar{x}$ е аритметичка средина на примерокот;



ii) одредуваме реален број $z_{\alpha/2}$ таков што $P_{H_0} \left\{ \bar{z} \geq z_{\alpha/2} \right\} = \frac{\alpha}{2}$,

односно $\Phi_0\left(z_{\alpha/2}\right)=1-\frac{\alpha}{2}$ и заклучуваме

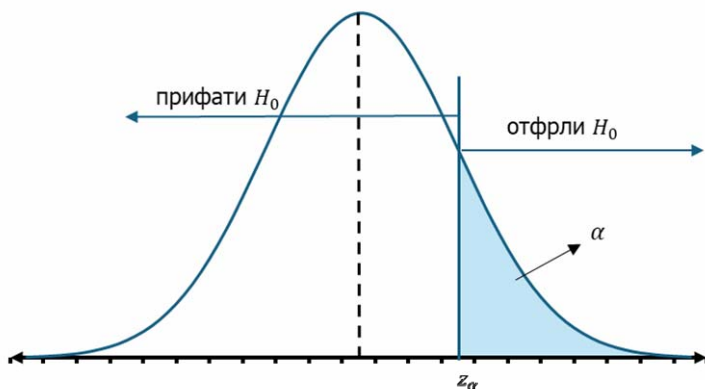
- ако $\bar{z} \leq z_{\alpha/2}$, не ја отфрламе хипотезата H_0 ,
- ако $\bar{z} > z_{\alpha/2}$, ја отфрламе хипотезата H_0 со веројатност на грешка од прв тип помала од α .

В) за хипотезата $H_0 : m \leq m_0$, наспроти алтернативната хипотеза $H_1 : m > m_0$:

- i) пресметуваме $\bar{z} = \frac{\bar{x} - m_0}{\sigma} \sqrt{n}$ за тестот 1) и $\bar{z} = \frac{\bar{x} - m_0}{\sigma} \sqrt{\frac{n(N-1)}{N-n}}$, за тестот 2), каде што $\bar{x}$ е аритметичка средина на примерокот;

ii) одредуваме реален број z_α таков што $P_{H_0}\{\bar{z} \geq z_\alpha\} = \alpha$, односно $\Phi_0(z_\alpha) = 1 - \alpha$ и заклучуваме

- ако $\bar{z} \leq z_\alpha$ не ја отфрламе хипотезата H_0 ,
- ако $\bar{z} > z_\alpha$ ја отфрламе хипотезата H_0 со веројатност на грешка од прв тип помала од α .



С) за хипотезата $H_0 : m \geq m_0$ наспроти алтернативната хипотеза $H_1 : m < m_0$:

- i) пресметуваме $\bar{z} = \frac{\bar{x} - m_0}{\sigma} \sqrt{n}$ за тестот 1) и

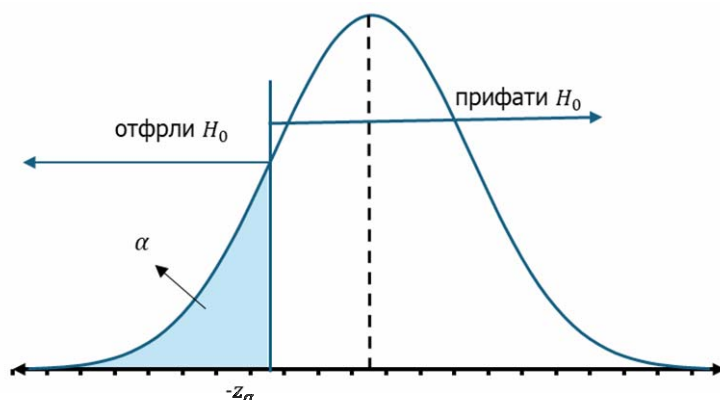
$\bar{z} = \frac{\bar{x} - m_0}{\sigma} \sqrt{\frac{n(N-1)}{N-n}}$ за тестот 2), каде што $\bar{x}$ е аритметичка

средина на примерокот;

ii) одредуваме реален број z_α таков што $P_{H_0} \{ \bar{z} \leq -z_\alpha \} = \alpha$, односно $\Phi_0(z_\alpha) = 1 - \alpha$ и заклучуваме

- ако $\bar{z} \geq -z_\alpha$, не ја отфрламе хипотезата H_0 ,

- ако $\bar{z} < -z_\alpha$, ја отфрламе хипотезата H_0 со веројатност на грешка од прв тип помала од α .



Задача 3.2.1.2. Поради зголемена продуктивност на машините од типот AZX, се предлага нивна надградба. Според пресметките, надградбата е економски оправдана доколку има зголемен број на операции на час и ако просечниот број на операции на час е поголем од 120. На една машина е реализирана надградба, 144 пати е евидентиран бројот на операции на час, при што беше откриено дека просечниот број на операции на час е 125.

До кој заклучок може да се дојде врз основа на спроведената истрага, ако се знае дека бројот на операции на час е нормално распределена случајна променлива со варијанса $\sigma^2 = 400$ и прифатлив ризик $\alpha = 0,05$?

Пресметајте ја вкупната/тоталната грешка ако веројатноста дека просечниот број на операции на час е поголем од 120 е 90 %, а јачината на тестот е 99 %.

Решение. Примерок е со големина $n = 144$ со просечна вредност на примерокот $\bar{x} = 125$, што е земен од бесконечно голема нормално распределена популација со позната стандардна девијација $\sigma = \sqrt{400} = 20$ и сакаме да тестираме дали просечниот

број операции на час е поголем од 120 (алтернативна хипотеза), па затоа ќе користиме z-тест за следнава хипотеза:

$$H_0 : \mu = 120, H_1 : \mu > 120 \text{ и статистика } Z = \frac{\bar{X} - \mu_0}{\sigma} \sqrt{n}.$$

Затоа, ја пресметуваме специфичната вредност

$$\bar{z} = \frac{\bar{x} - \mu_0}{\sigma} \sqrt{n} = \frac{125 - 120}{20} \sqrt{144} = 3$$

и од табелата на стандардна нормална распределба за критичната вредност $z_{0,05}$ имаме:

$$\Phi_0(1,64) = 0,94950 < \Phi_0(z_{0,05}) = 1 - 0,05 = 0,95 < 0,95053 = \Phi_0(1,65)$$

$$\text{или } z_{0,05} = \frac{1,64 + 1,65}{2} = 1,645.$$

Бидејќи $\bar{z} = 3 > 1,645 = z_{0,05}$, нултата хипотеза е отфрлена, со прифатлив ризик од грешка од типот I помала од 0,05. Значи дека имаме доволно причини да мислиме дека надградбата е економски оправдана. Бидејќи:

$$P(H_0 \text{ не е вистинита}) = 0,9, P(H_0 \text{ е вистинита}) = 1 - 0,9 = 0,1 \text{ и} \\ 1 - \beta = 0,99 \Rightarrow \beta = 0,01,$$

вкупната грешка е:

$$Vg = \alpha \cdot P(H_0 \text{ е вистинита}) + \beta \cdot P(H_0 \text{ не е вистинита}) = \\ 0,05 \cdot 0,1 + 0,01 \cdot 0,9 = 0,14 = 14\%.$$

Задача 3.2.1.3. Сите цигари што моментално се на пазарот имаат просечна содржина на никотин од најмалку 1,6 mg по цигара. Фирма која произведува цигари тврди дека открила нов начин за лекување на листовите од тутун што ќе резултира со просечна содржина на никотин во цигара помала од 1,6 mg. За да се тестира ова тврдење, анализиран е примерок од 30 цигари на фирмата. Познато е дека стандардното отстапување на содржината на никотин во цигарата е 8 mg.

а) Какви заклучоци може да се извлечат на 5-процентно ниво на значајност, ако просечната содржина на никотин во 30-те цигари е 1,54?

б) Пресметајте ја вкупната грешка ако веројатноста дека просечната содржина на никотин во цигара е помала од 1,6 mg е 10 %, а моќноста на тестот е 98,5 %.

Решение. а) Примерок е со големина $n=30$ со просечна вредност на примерокот $\bar{x}=1,54$, што е земен од бесконечно голема популација со позната стандардна девијација $\sigma=8$ и сакаме да тестираме дали просечна содржина на никотин во цигара е помала од 1,6 mg (алтернативна хипотеза), па затоа ќе користиме z-тест за следнава хипотеза:

$$H_0: \mu = 1,6 \quad H_1: \mu < 1,6 \text{ и статистика } Z = \frac{\bar{X} - \mu_0}{\sigma} \sqrt{n}.$$

Затоа, ја пресметуваме специфичната вредност:

$$\bar{z} = \frac{1,54 - 1,6}{8} \sqrt{30} = -0,0075 \cdot 5,477 = -0,041;$$

и од табелата на стандардна нормална распределба за критичната вредност $z_{0,05}$ имаме:

$$\Phi_0(1,64) = 0,94950 < \Phi_0(z_{0,05}) = 1 - 0,05 = 0,95 < 0,95053 = \Phi_0(1,65)$$

$$\text{или } z_{0,05} = \frac{1,64 + 1,65}{2} = 1,645.$$

Бидејќи $\bar{z} = -0,041 > -1,645 = z_{0,05}$ нултата хипотеза не може да ја отфрлиме, со прифатлив ризик од грешка од типот I помала од 0,05. Значи дека фирмата за цигари нема доволно причини да мислиме дека новиот начин за лекување на листовите од тутун ќе резултира со просечна содржина на никотин во цигара помала од 1,6 mg. Бидејќи:

$$P(H_0 \text{ не е вистинита}) = 0,1, \quad P(H_0 \text{ е вистинита}) = 1 - 0,1 = 0,9$$

и

$$1 - \beta = 0,985 \Rightarrow \beta = 0,015,$$

вкупната грешка е:

$$Vg = \alpha \cdot P(H_0 \text{ е вистинита}) + \beta \cdot P(H_0 \text{ не е вистинита}) = 0,05 \cdot 0,9 + 0,015 \cdot 0,1 = 0,0465 = 4,65\%.$$

Задача 3.2.1.4. Еден производител тврди дека нивните светилки траат во просек 1000 часа. Примерок од 100 светилки има просечен животен век од 985 часа. Познато е дека стандардното отстапување на животниот век на светилките е 50 часа. На ниво на значајност од 5 %, дали има докази за отфрлање на тврдењето на производителот?

Решение. а) Примерок е со големина $n = 100$ со просечна вредност на примерокот $\bar{x} = 985$, што е земен од бесконечно голема популација со позната стандардна девијација $\sigma = 50$ и сакаме да тестираме дали просечниот животен век на светилките е 1000 часа, па затоа ќе користиме z-тест за следнава хипотеза:

$$H_0 : \mu = 1000, H_1 : \mu \neq 1000 \text{ и статистика } Z = \frac{|\bar{X} - \mu_0|}{\sigma} \sqrt{n}.$$

Затоа, ја пресметуваме специфичната вредност:

$$\bar{z} = \frac{|985 - 1000|}{50} \sqrt{100} = 0,3 \cdot 10 = 3,$$

и од табелата на стандардна нормална распределба за критичната вредност $z_{0,05}$ имаме:

$$\Phi_0(z_{0,025}) = 1 - 0,025 = \Phi_0(1,96) \text{ или } z_{0,025} = 1,96.$$

Бидејќи $\bar{z} = 3 > 1,96 = z_{0,025}$, нултата хипотеза е отфрлена со прифатлив ризик од грешка од типот I помала од 0,05. Значи дека имаме доволно причини да мислиме дека животниот век на светилките од дадениот производител не е 1000 часа.

Студентов t -тест

Наведените претпоставки кои се неопходни за да при тестирањето се користи Z -тестот во пракса најчесто не се исполнети, пред сè бидејќи *дисперзијата на популацијата не е позната*. Тогаш не може да се пресмета стандардната грешка

$$\frac{\sigma}{\sqrt{n}},$$

па затоа не може да се пресмета и вредноста на статистиката на Z -тестот. Затоа, прво се оценува стандардното отстапување σ , со помош на стандардната девијација, односно отстапување на примерокот S , и тогаш добиваме нова статистика за тестот, т.н *статистика за t -тест*, за која примерокот има големина $n < 30$:

$$t_{n-1} = \frac{\bar{X} - m_0}{S} \sqrt{n},$$

која има Студентова распределба со $n-1$ степени на слобода.

Во следната табела 3.3 се претставени условите за примена на t -тестот и постапката за тестирање хипотези за математичкото очекување на популацијата при непозната дисперзија.

Табела 3.2.1.5. Тестирање хипотези за математичко очекување со примена на t -тестот

Услови за примена на t -тестот:

Популацијата има нормална распределба, σ^2 не е позната и примерокот има големина $n < 30$

Избор на обликот на нултата и алтернативната хипотеза:

$$H_0 : m = m_0, \quad H_1 : m \neq m_0;$$

$$H_0 : m \leq m_0, \quad H_1 : m > m_0;$$

$$H_0 : m \geq m_0, \quad H_1 : m < m_0;$$

каде што m_0 е дадена нумеричка вредност.

Статистика за тестот:

$$t_{n-1} = \frac{\bar{X} - m_0}{S} \sqrt{n}, \text{ кога:}$$

- примерокот е земен од бесконечна популација;
- примерокот е земен од конечна популација, при што изборот се врши со враќање и
- примерокот е земен од конечна популација, при што изборот се врши без враќање, но големината на примерокот n е помала од 5 % од големината на популацијата N со ниво на значајност α .

А) За хипотезата $H_0 : m = m_0$ наспроти алтернативната хипотеза $H_1 : m \neq m_0$:

i) пресметуваме $\bar{t}_{n-1} = \frac{|\bar{x} - m_0|}{s} \sqrt{n}$, каде што $\bar{x}$ е

аритметичка средина на примерокот, т.е просечна вредност, а s е стандардно отстапување на примерокот

ii) читаме вредност $t_\alpha(n-1)$ од таблицата за Студентова распределба и заклучуваме

- ако $\bar{t}_{n-1} \leq t_{\alpha}(n-1)$, не ја отфрламе хипотезата H_0 ,

- ако $\bar{t}_{n-1} > t_{\alpha}(n-1)$, ја отфрламе хипотезата H_0 со веројатност на грешка од прв тип помала од α .

Забелешка. Вредноста $t_{\alpha}(n-1)$ ја читаме од табелата на Студентовата распределба за двостран тест.

В) За хипотезата $H_0 : m \leq m_0$ наспроти алтернативната хипотеза $H_1 : m < m_0$:

i) пресметуваме $\bar{t}_{n-1} = \frac{\bar{x} - m_0}{s} \sqrt{n}$, каде што $\bar{x}$ е аритметичка

средина на примерокот, т.е просечна вредност, а s е стандардно отстапување на примерокот;

ii) читаме вредност $t_{\alpha}(n-1)$ од таблицата за Студентова распределба и заклучуваме:

- ако $\bar{t}_{n-1} \leq t_{\alpha}(n-1)$ не ја отфрламе хипотезата H_0 ,

- ако $\bar{t}_{n-1} > t_{\alpha}(n-1)$ ја отфрламе хипотезата H_0 со веројатност на грешка од прв тип помала од α .

Забелешка. Вредноста $t_{\alpha}(n-1)$ ја читаме од табелата на Студентовата распределба за едностран тест.

С) За хипотезата $H_0 : m \geq m_0$, наспроти алтернативната хипотеза $H_1 : m < m_0$:

i) пресметуваме $\bar{t}_{n-1} = \frac{\bar{x} - m_0}{s} \sqrt{n}$, каде што $\bar{x}$ е аритметичка

средина на примерокот, т.е просечна вредност, а s е стандардно отстапување на примерокот:

ii) читаме вредност $t_{\alpha}(n-1)$ од таблицата за Студентова распределба и заклучуваме

- ако $\bar{t}_{n-1} \geq -t_{\alpha}(n-1)$, не ја отфрламе хипотезата H_0 ,

- ако $\bar{t}_{n-1} < -t_{\alpha}(n-1)$, ја отфрламе хипотезата H_0 со веројатност на грешка од прв тип помала од α .

Забелешка: Вредноста $t_{\alpha}(n-1)$ ја читаме од табелата на Студентовата распределба за едностран тест.

Задача 3.2.1.6. Замислете дека собравме случаен примерок од 24 енергетски чоколатца од неколку различни продавници за да ја претставиме популацијата на енергетски чоколатца достапни за општиот потрошувач. Етикетите на чоколатцата тврдат дека секое чоколадце содржи 20 грама протеини. Тестирајте го ова тврдење со веројатност на грешка од прв тип 0,05, ако податоците од примерокот се дадени во следнава табела.

20,70	27,46	22,15	19,85	21,29	24,75	22,14	19,56
20,75	22,91	25,34	20,33	21,54	19,95	19,72	18,28
20,53	22,12	25,06	22,44	19,08	21,10	18,04	17,46

Решение. Примерок е со големина $n = 24 < 30$ со просечна вредност на примерокот $\bar{x} = 21,36$ и стандардна девијација на примерокот $s = 2,44$, што е земен од популација со непозната стандардна девијација и сакаме да тестираме дали секое чоколадце содржи 20 грама протеини, па затоа ќе користиме t -тест за следнава хипотеза:

$$H_0 : \mu = 20, H_1 : \mu \neq 20 \text{ и статистика } t_{n-1} = \frac{\bar{X} - m_0}{S} \sqrt{n}.$$

Затоа, ја пресметуваме специфичната вредност:

$$\bar{t}_{n-1} = \frac{|\bar{x} - m_0|}{s} \sqrt{n} = \frac{|21,36 - 20|}{2,44} \sqrt{24} = 0,557 \cdot 4,899 = 2,729,$$

и од табелата за Студентова распределба, двостран тест за $\alpha = 0,05$ и $n - 1 = 23$ степени на слобода за критичната вредност имаме:

$$t_{0,05}(23) = 2,069.$$

Бидејќи $\bar{t}_{23} = 2,729 > 2,069 = t_{0,05}(23)$, нултата хипотеза е отфрлена, со прифатлив ризик од грешка од типот I помала од 0,05. Значи дека имаме доволно причини да мислиме дека вредноста на протеини во едно чоколадце е различна од 20 грама протеини.

Задача 3.2.1.7. Во едно училиште редовно се практикува дополнителна помош при совладување на училишните обврски, но неа ја користат само мал број на ученици. Наставникот по математика кој има податок дека просекот на сите ученици во училиштето е 2,5; со цел да ги мотивира учениците дека со дополнителната помош нивниот просек ќе се зголеми, разгледува

примерок од 30 ученици кои ја користат дополнителната помош и добива дека нивниот просек е 2,75 со стандардна девијација од 0,5. Што покажува неговото тестирање со веројатност ниво на зачајност од 0,05?

Решение. Примерок е со големина $n=30$ со просечна вредност на примерокот $\bar{x}=2,5$ и стандардна девијација на примерокот $s=0,5$ што е земен од популација со непозната стандардна девијација и сакаме да тестираме дали просекот на учениците кои практикуваат дополнителна помош при совладување на училишните обврски е поголем од 2,5, па затоа ќе користиме t -тест за следнава хипотеза:

$$H_0: \mu = 2,5, \quad H_1: \mu > 2,5 \quad \text{и статистика} \quad t_{n-1} = \frac{\bar{X} - m_0}{S} \sqrt{n}.$$

Затоа, ја пресметуваме специфичната вредност:

$$\bar{t}_{n-1} = \frac{\bar{x} - m_0}{s} \sqrt{n} = \frac{2,75 - 2,5}{0,5} \sqrt{30} = 2,739$$

и од таблицата за Студентова распределба, едностран тест за $\alpha=0,05$ и $n-1=29$ степени на слобода за критичната вредност имаме: $t_{0,05}(29)=1,699$.

Бидејќи $\bar{t}_{29} = 2,739 > 1,699 = t_{0,05}(29)$ нултата хипотеза е отфрлена, со прифатлив ризик од грешка од тип I помала од 0,05. Значи дека имаме доволно причини да сметаме дека просекот на учениците кои ја користат дополнителната помош се зголемува.

Забелешка. Бидејќи, со зголемување на големината на примерокот Студентовата распределба тежи кон Z -распределбата, тогаш за многу големи вредности на n , критичните t -вредности стануваат приближно еднакви на критичните z -вредности. Затоа, во пракса се прифаќа дека, *ако статистичкото множество има нормална распределба, со непозната дисперзија и примерокот е многу голем (по некои автори доволно е $n > 30$, а по други дури $n \geq 150$), тестирањето може да се спроведе со примена на Z -тестот наместо t -тестот.*

Во следната табела 3.4 се претставени условите за примена на Z -тестот и постапката за тестирање на хипотези за математичкото очекување на популацијата при непозната дисперзија во случај на големи примероци.

Табела 3.2.1.8. Тестирање хипотези за математичко очекување при непозната дисперзија со големи примероци

Услови за примена на Z - тестот:

Популацијата има нормална распределба, σ^2 не е позната и примерокот има произволна големина $n > 30$.

Избор на обликот на нултата и алтернативната хипотеза:

$$H_0 : m = m_0, \quad H_1 : m \neq m_0;$$

$$H_0 : m \leq m_0, \quad H_1 : m > m_0;$$

$$H_0 : m \geq m_0, \quad H_1 : m < m_0,$$

каде што m_0 е дадена нумеричка вредност.

Статистика за тестот:

$$Z = \frac{\bar{X} - m_0}{S} \sqrt{n}, \text{ кога:}$$

-примерокот е земен од бесконечна популација,

-примерокот е земен од конечна популација, при што изборот се врши со враќање и

-примерокот е земен од конечна популација, при што изборот се врши без враќање, но големината на примерокот n е помала од 5 % од големината на популацијата N

$$Z = \frac{\bar{X} - m_0}{S} \sqrt{\frac{n(N-1)}{N-n}},$$

кога примерокот е земен од конечна популација, при што изборот се врши без враќање, но големината на примерокот n е поголема од 5 % од големината на популацијата N со ниво на значајност α .

А) За хипотезата $H_0 : m = m_0$, наспроти алтернативната хипотеза $H_1 : m \neq m_0$:

i) пресметуваме $\bar{z} = \frac{|\bar{x} - m_0|}{s} \sqrt{n}$ за тестот 1) и

$\bar{z} = \frac{|\bar{x} - m_0|}{s} \sqrt{\frac{n(N-1)}{N-n}}$ за тестот 2), каде што $\bar{x}$ е аритметичка средина на примерокот, s е стандардното отстапување на примерокот;

ii) одредуваме реален број $z_{1-\alpha}$ таков што

$P_{H_0} \left\{ \bar{z} \geq z_{\alpha/2} \right\} = \frac{\alpha}{2}$, односно $\Phi_0 \left(z_{\alpha/2} \right) = 1 - \frac{\alpha}{2}$ и заклучуваме:

- ако $\bar{z} \leq z_{\alpha/2}$, не ја отфрламе хипотезата H_0 ,

- ако $\bar{z} > z_{\alpha/2}$, ја отфрламе хипотезата H_0 со веројатност

на грешка од прв тип помала од α .

В) За хипотезата $H_0 : m \leq m_0$ наспроти алтернативната хипотеза $H_1 : m > m_0$:

i) пресметуваме $\bar{z} = \frac{\bar{x} - m_0}{s} \sqrt{n}$ за тестот 1) и

$\bar{z} = \frac{\bar{x} - m_0}{s} \sqrt{\frac{n(N-1)}{N-n}}$ за тестот 2), каде што $\bar{x}$ е аритметичка средина на примерокот, s е стандардното отстапување на примерокот;

ii) одредуваме реален број $z_{1-\alpha}$ таков што $P_{H_0} \left\{ \bar{z} \geq z_{\alpha} \right\} = \alpha$, односно $\Phi_0 \left(z_{\alpha} \right) = 1 - \alpha$ и заклучуваме

- ако $\bar{z} \leq z_{\alpha}$ не ја отфрламе хипотезата H_0 ,

- ако $\bar{z} > z_{\alpha}$ ја отфрламе хипотезата H_0 со веројатност на грешка од прв тип помала од α .

С) за хипотезата $H_0 : m \geq m_0$ наспроти алтернативната хипотеза $H_1 : m < m_0$:

i) пресметуваме $\bar{z} = \frac{\bar{x} - m_0}{s} \sqrt{n}$ за тестот 1) и

$\bar{z} = \frac{\bar{x} - m_0}{s} \sqrt{\frac{n(N-1)}{N-n}}$ за тестот 2), каде што каде што $\bar{x}$ е аритметичка средина на примерокот, s е стандардното отстапување на примерокот;

ii) одредуваме реален број $z_{1-\alpha}$ таков што $P_{H_0} \left\{ \bar{z} \leq -z_{\alpha} \right\} = \alpha$, односно $\Phi_0 \left(z_{\alpha} \right) = 1 - \alpha$ и заклучуваме

- ако $\bar{z} \geq -z_{\alpha}$, не ја отфрламе хипотезата H_0 ,

-ако $\bar{z} < -z_{\alpha}$, ја отфрламе хипотезата H_0 со веројатност на грешка од прв тип помала од α .

Задача 3.2.1.9. Ресторанот Ексклузив рекламира достава на готова храна. Во соопштението се наведува дека нарачаната храна ќе биде доставена за помалку од 50 минути од приемот на нарачката. Врз основа на случаен примерок од 100 нарачки, просечното време за испорака на нарачките се пресметува на 49 минути, а стандардното отстапување од 5 минути. Што може да се заклучи од овој случаен примерок за времето на испорака со ниво на значајност од 0,05?

Решение. Разгледуваме примерок со големина $n = 100 > 30$ аритметичка средина $\bar{x} = 49$ и стандардно отстапување $s = 5$, земен од бесконечна популација. Она што сакаме да го провериме е дали времето на достава на храната во просек е помалку од 50 минути (алтернативна хипотеза), затоа употребуваме z -тест за следнава хипотеза:

$$H_0 : m = 50, \text{ наспроти алтернативната хипотеза } H_1 : m < 50,$$

а статистиката за тестот е: $Z = \frac{\bar{X} - m_0}{S} \sqrt{n}$.

Тогаш специфичната вредност изнесува:

$$\bar{z} = \frac{\bar{x} - \mu_0}{s} \sqrt{n} = \frac{49 - 50}{5} \sqrt{100} = -2.$$

а од табелата за нормална распределба за критичната вредност $z_{0,05}$ имаме:

$$\begin{aligned} \Phi_0(1,64) &= 0,94950 < \Phi_0(z_{0,05}) = \\ 1 - 0,05 &= 0,95 < 0,95053 = \Phi_0(1,65) \end{aligned}$$

или

$$z_{0,05} = \frac{1,64 + 1,65}{2} = 1,645.$$

Бидејќи

$$\bar{z} = -2 < -1,645 = z_{0,05},$$

нултата хипотеза ја отфрламе, со прифатлив ризик од грешка од типот I помала од 0,05. Значи дека во согласност со примерокот постојат доволно причини да сметаме дека ресторанот Ексклузив ќе ја достави нарачаната храна за помалку од 50 минути од приемот на нарачката.

3.2.2. ТЕСТОВИ ЗА НЕПОЗНАТА ВЕРОЈАТНОСТ

Нека за секој елемент на популацијата се интересираме дали се реализирал настанот A , со веројатноста дека тој да се реализира $P(A) = p$. Тестирањето на хипотезите за оваа веројатност со помош на мали примероци се реализира со нејзината непристрасна оценка P која има биномна распределба (кај примероците со повторување), односно $P = \frac{S_n}{n}$, каде што со S_n го означуваме бројот на реализациите на настанот A , или хипергеометриска распределба (кај примероците без повторување). Меѓутоа, ако имаме *големи примероци*, претходната оценка како статистика на тестот се стандардизира до следната променлива:

$$Z = \frac{S_n - np_0}{\sqrt{np_0q_0}}.$$

Статистиката Z има приближно стандардизирана нормална распределба $N(0,1)$, каде што p_0 е хипотетичката вредност на веројатноста p на настанот A , а $q_0 = 1 - p_0$.

Во следната табела 3.5 се претставени условите за примена на Z -тестот и постапката за тестирање хипотези за веројатноста на настан во случај на големи примероци.

Табела 3.2.2.1. Тестирање хипотези за веројатноста на настан во случај на големи примероци

Услови за примена на Z -тестот:

Веројатноста за „успех“ е константна при изборот на елемент од примерокот:

$$np_0 \geq 5, nq_0 \geq 5 \text{ и } n \geq 30$$

Избор на обликот на нултата и алтернативната хипотеза:

$$H_0 : p = p_0, \quad H_1 : p \neq p_0;$$

$$H_0 : p \leq p_0, \quad H_1 : p > p_0;$$

$$H_0 : p \geq p_0, \quad H_1 : p < p_0,$$

каде што p_0 е дадена нумеричка вредност.

Статистика за тестот:

$$1) Z = \frac{S_n - np_0}{\sqrt{np_0q_0}}, \text{ кога:}$$

- примерокот е земен од бесконечна популација,
- примерокот е земен од конечна популација, при што изборот се врши со враќање, и
- примерокот е земен од конечна популација, при што изборот се врши без враќање, но големината на примерокот n е помала од 5% од големината на популацијата N .

$$1) Z = \frac{S_n - np_0}{\sqrt{np_0q_0}} \sqrt{\frac{N-1}{N-n}}, \text{ кога:}$$

Примерокот е земен од конечна популација, при што изборот се врши без враќање, но големината на примерокот n е поголема од 5% од големината на популацијата N со ниво на значајност α .

А) За хипотезата $H_0: p = p_0$, наспроти алтернативната хипотеза $H_1: p \neq p_0$:

i) пресметуваме $\bar{z} = \frac{|S_n - np_0|}{\sqrt{np_0q_0}}$ за тестот 1) и

$$\bar{z} = \frac{|S_n - np_0|}{\sqrt{np_0q_0}} \sqrt{\frac{N-1}{N-n}}$$

за тестот 2), каде што S_n е бројот на реализациите на настанот A ;

ii) одредуваме реален број $z_{\alpha/2}$ таков што $P_{H_0} \left\{ \bar{z} \geq z_{\alpha/2} \right\} = \frac{\alpha}{2}$,

односно $\Phi_0 \left(z_{\alpha/2} \right) = 1 - \frac{\alpha}{2}$ и заклучуваме

- ако $\bar{z} \leq z_{\alpha/2}$, не ја отфрламе хипотезата H_0 ,

- ако $\bar{z} > z_{\alpha/2}$, ја отфрламе хипотезата H_0 со веројатност

на грешка од прв тип помала од α .

В) за хипотезата $H_0: p \leq p_0$, наспроти алтернативната хипотеза $H_1: p > p_0$:

i) пресметуваме $\bar{z} = \frac{S_n - np_0}{\sqrt{np_0q_0}}$ за тестот 1) и

$\bar{z} = \frac{S_n - np_0}{\sqrt{np_0q_0}} \sqrt{\frac{N-1}{N-n}}$ за тестот 2), каде што S_n е бројот на

реализациите на настанот A ;

ii) одредуваме реален број z_α таков што $P_{H_0} \{ \bar{z} \geq z_\alpha \} = \alpha$,
односно $\Phi_0(z_\alpha) = 1 - \alpha$ и заклучуваме

- ако $\bar{z} \leq z_\alpha$, не ја отфрламе хипотезата H_0 ,

- ако $\bar{z} > z_\alpha$, ја отфрламе хипотезата H_0 со веројатност на грешка од прв тип помала од α .

С) За хипотезата $H_0: p \geq p_0$ наспроти алтернативната хипотеза $H_1: p < p_0$:

i) пресметуваме $\bar{z} = \frac{S_n - np_0}{\sqrt{np_0q_0}}$ за тестот 1) и

$\bar{z} = \frac{S_n - np_0}{\sqrt{np_0q_0}} \sqrt{\frac{N-1}{N-n}}$ за тестот 2), каде што S_n е бројот на

реализациите на настанот

ii) одредуваме реален број z_α таков што $P_{H_0} \{ \bar{z} \leq -z_\alpha \} = \alpha$,
односно $\Phi_0(z_\alpha) = 1 - \alpha$ и заклучуваме

- ако $\bar{z} \geq -z_\alpha$, не ја отфрламе хипотезата H_0 ,

- ако $\bar{z} < -z_\alpha$, ја отфрламе хипотезата H_0 со веројатност на грешка од прв тип помала од α .

Задача 3.2.2.2. Истражувачка компанија тврди дека 60 % од гласачите поддржуваат нова политика. Во примерок од 1000 гласачи, 570 ја поддржуваат политиката. На ниво на значајност од 5%, дали има докази за поддршка на тврдењето на компанијата?

Решение. Примерок е со големина $n=1000$ и земен е од голема популација на гласачи. Притоа $S_{1000}=570$ гласачи ја поддржуваат новата политика и сакаме да провериме дека веројатноста гласачите да ја поддржуваат новата политика е различен од 60%. Од тие причини го користиме Z -тестот за следнава хипотеза:

$$H_0: p = 0,6, H_1: p \neq 0,6 \text{ и статистика } \bar{z} = \frac{|S_n - np_0|}{\sqrt{np_0q_0}}.$$

Затоа, ја пресметуваме специфичната вредност

$$\bar{z} = \frac{|S_n - np_0|}{\sqrt{np_0q_0}} = \frac{|570 - 1000 \cdot 0,6|}{\sqrt{1000 \cdot 0,6 \cdot 0,4}} = \frac{|570 - 1000 \cdot 0,6|}{\sqrt{1000 \cdot 0,6 \cdot 0,4}} = 1,936$$

и од табелата на стандардна нормална распределба за критичната вредност $z_{0,05}$ имаме:

$$\Phi_0(z_{0,025}) = 1 - 0,025 = \Phi_0(1,96) \text{ или } z_{0,025} = 1,96.$$

Бидејќи $\bar{z} = 1,936 < 1,96 = z_{0,025}$ немаме доволно причини да ја отфрлиме нултата хипотеза, со веројатност од грешка од типот I помала од 0,05. Значи дека немаме доволно причини да тврдиме дека новата политика ја подржуваат различно од 60 % од гласачите.

Задача 3.2.2.3. Во една пратка од 10 000 пара чевли од истиот модел е избран примерок со големина 4 %, во кој се пронајдени 19 неисправни чевли. Дали купувачот ќе ја прифати пратката ако договорот предвидува дека пратката ќе толерира најмногу 5 % од неисправни чевли? Значајното ниво е 0,05.

Решение. Разгледуваме конечна популација со големина $N = 10000$ и примерокот е со големина:

$$n = 4\% \cdot N = 400 < 500 = 5\% \cdot N.$$

Притоа, бројот на успеси, односно бројот на неисправни пара чевли е $S_n = 19$. Сакаме да провериме дали купувачот ќе ја прифати пратката ако договорот предвидува дека пратката ќе толерира најмногу 5 % од неисправни чевли. Од тие причини го користиме Z -тестот за следнава хипотеза:

$$H_0 : p \geq 0,5, H_1 : p < 0,05 \text{ и статистика } \bar{z} = \frac{S_n - np_0}{\sqrt{np_0q_0}}.$$

Затоа, ја пресметуваме специфичната вредност

$$\bar{z} = \frac{S_n - np_0}{\sqrt{np_0q_0}} = \frac{19 - 400 \cdot 0,05}{\sqrt{400 \cdot 0,05 \cdot 0,95}} = \frac{19 - 20}{\sqrt{19}} = -0,23$$

и од табелата на стандардна нормална распределба за критичната вредност $z_{0,05}$ имаме:

$$\Phi_0(1,64) = 0,94950 < \Phi_0(z_{0,055}) = 1 - 0,05 = 0,95 < 0,95053 = \Phi_0(1,65)$$

$$\text{или } z_{0,95} = \frac{1,64 + 1,65}{2} = 1,645.$$

Бидејќи $\bar{z} = -0,23 > -1,645 = -z_{0,05}$, немаме доволно причини да ја отфрлиме нултата хипотеза, со веројатност од грешка од тип I помала од 0,05. Значи дека купувачот нема да ја прифати пратката.

Задача 3.2.2.4. Во една фабрика процентот на неисправни производи во прва смена е 10 %. Дали процентот на неисправни производи во втора смена е значително поголем ако се најдат 59 неисправни производи во примерок од 450 производи со прифатлив ризик од 0,01.

Решение. Примерок е со големина $n = 450$ и земен е од голема популација, притоа $S_{450} = 59$ се дефектни и сакаме да провериме дали процентот за неисправни производи во втората смена е поголем од 10 %. Од тие причини го користиме Z -тестот за следнава хипотеза:

$$H_0 : p = 0,1, H_1 : p > 0,1 \text{ и статистика } \bar{z} = \frac{S_n - np_0}{\sqrt{np_0q_0}}.$$

Затоа, ја пресметуваме специфичната вредност:

$$\bar{z} = \frac{S_n - np_0}{\sqrt{np_0q_0}} = \frac{59 - 450 \cdot 0,1}{\sqrt{450 \cdot 0,1 \cdot 0,90}} = \frac{59 - 45}{\sqrt{40,5}} = 2,2$$

и од табелата на стандардна нормална распределба за критичната вредност $z_{0,01}$ имаме:

$$\Phi_0(2,32) = 0,98983 < \Phi_0(z_{0,01}) = 1 - 0,01 = 0,99 < 0,99010 = \Phi_0(2,33)$$

$$\text{или } z_{0,01} = \frac{2,32 + 2,33}{2} = 2,325.$$

Бидејќи $\bar{z} = 2,2 < 2,325 = z_{0,01}$, немаме доволно причини да ја отфрлиме нултата хипотеза, со веројатност од грешка од типот I помала од 0,01. Значи во согласност со примерокот немаме доволно причини да сметаме дека процентот на несиправни производи во втората смена купувачот ќе биде поголем од 10 %.

Задача 3.2.2.5. Финансиското одделение на големопродажба ги анализира причините за слабиот прилив на средства. Се претпоставува дека една од причините е

неправилноста во плаќањето на достасаните фактури. Одделот претпоставува дека повеќе од 75 % од клиентите не плаќаат во пропишаниот рок од 60 дена. Од 4 270 неплатени фактури земен е примерок од 235 неплатени фактури, а 198 се пронајдени со пречекорување на рокот од 60 дена. Што можете да заклучите врз основа на овој примерок, со прифатлив ризик од 0,02?

Решение. Разгледуваме примерок со големина $n = 235$ што е земен од популација со големина $N = 4270$ така што успех е настанот „клиентите плаќаат со пречекорување на рокот од 60 дена“, па бројот на успеси е $S_n = 198$.

Сакаме да тестираме дали процентот на клиенти кои плаќаат со пречекорување на рокот од 60 дена е повеќе од 75 % и бидејќи големината на примерокот е $\frac{235}{4270} = 5,5\%$, значи дека е поголема од 5 % од големината на популацијата, така што ќе користиме Z -тест статистика за следниве хипотези:

$$H_0 : p = 0,75, H_1 : p > 0,75 \text{ и статистика } Z = \frac{S_n - np_0}{\sqrt{np_0q_0}} \sqrt{\frac{N-1}{N-n}}.$$

Затоа, ја пресметуваме специфичната вредност:

$$\bar{z} = \frac{S_n - np_0}{\sqrt{np_0q_0}} \sqrt{\frac{N-1}{N-n}} = \frac{198 - 235 \cdot 0,75}{\sqrt{235 \cdot 0,75 \cdot 0,25}} \sqrt{\frac{4270-1}{4270-235}} = 3,4;$$

и од табелата на стандардна нормална распределба за критичната вредност $z_{0,02}$ имаме:

$$\Phi_0(2,05) = 0,97982 < \Phi_0(z_{0,02}) = 1 - 0,02 = 0,98 < 0,98030 = \Phi_0(2,06)$$

$$\text{или } z_{0,02} = \frac{2,05 + 2,06}{2} = 2,055$$

Бидејќи $\bar{z} = 3,4 > 2,055 = z_{0,02}$ нултата хипотеза е отфрлена, со прифатлив ризик од грешка од тип I помала од 0,02. Значи дека имаме доволно причини да мислиме дека повеќе од 75% од клиентите не плаќаат во пропишаниот рок од 60 дена.

3.3. ТЕСТИРАЊЕ ХИПОТЕЗИ ВРЗ ОСНОВА НА ДВА ПРИМЕРОКА

Покрај тестирањето на хипотезите за вредноста на еден параметар на популацијата, во областа на тестирање се опфатени и

методи со кои се споредуваат две статистички множества, односно се испитува еднаквост или разлика на вредностите на ист параметар на две статистички множества. На пример, кога се проверува претпоставката дека домашните и увозните производи од ист вид имаат ист квалитет или, пак, учеството на домашните туристи во вкупниот број на туристи е поголем од учеството на домашните туристи во вкупниот број на туристи во одреден регион и слично. Значи, во овој случај предмет на тестирање не е вредноста на еден ист параметар во две статистички множества, туку односот меѓу вредностите на истиот параметар од статистичките множества.

3.3.1. ТЕСТОВИ ЗА ЕДНАКВОСТ НА МАТЕМАТИЧКИТЕ ОЧЕКУВАЊА ПРИ НЕЗАВИСНИ ПРИМЕРОЦИ

Z-тест

Нека $X_1, X_2, \dots, X_{n_1}$ и $Y_1, Y_2, \dots, Y_{n_2}$ се два независни примероци земени од популации со нормални распределби $X: N(m_1, \sigma_1^2)$ и $Y: N(m_2, \sigma_2^2)$, соодветно при што дисперзиите σ_1^2 и σ_2^2 се познати. Аритметичките средини на примероците $\bar{X}$ и $\bar{Y}$ се статистики кои се оценка на математичките очекувања $E(X) = m_1$ и $E(Y) = m_2$ на двете популации X и Y , соодветно. Притоа важи:

$$E(\bar{X}) = m_1, E(\bar{Y}) = m_2, D(\bar{X}) = \frac{\sigma_1^2}{n_1} \text{ и } D(\bar{Y}) = \frac{\sigma_2^2}{n_2}.$$

Тогаш, разликата на аритметичките средини на примероците $\bar{X} - \bar{Y}$ е непристрасна оценка за разликата $m_1 - m_2$ меѓу математичките очекувања на популациите, за која важи:

$$E(\bar{X} - \bar{Y}) = E(\bar{X}) - E(\bar{Y}) = m_1 - m_2 \text{ и } D(\bar{X} - \bar{Y}) = D(\bar{X}) + (-1)^2 D(\bar{Y}) = \frac{\sigma_1^2}{n_1} + \frac{\sigma_2^2}{n_2}.$$

Бидејќи оценката има нормална распределба, тестирањето на хипотезите се спроведува со примена на Z -тестот, при што претходната оценка како статистика на тестот се стандардизира до следната променлива:

$$Z = \frac{(\bar{X} - \bar{Y}) - (m_1 - m_2)}{\sqrt{n_2\sigma_1^2 + n_1\sigma_2^2}} \sqrt{n_1n_2}.$$

Во следнава табела 3.6, се претставени условите за примена на Z -тестот и постапката за тестирање хипотези за разликата меѓу математичките очекувања на две популации, кога примероците се независни.

Табела 3.3.1.1. Тестирање хипотези за разликата меѓу математичките очекувања на две популации, кога примероците се независни со примена на Z -тестот

<i>Услови за примена на Z тестот:</i> Случајните примероци се независни. Дисперзиите на популациите се познати: а. популациите имаат нормална распределба на веројатност, б. распределбата на популациите не е позната, и примероците се големи, односно $n_1 \geq 30$ и $n_2 \geq 30$. Избор на обликот на нултата и алтернативната хипотеза: $H_0 : m_1 = m_2, \quad H_1 : m_1 \neq m_2;$ $H_0 : m_1 \leq m_2, \quad H_1 : m_1 > m_2;$ $H_0 : m_1 \geq m_2, \quad H_1 : m_1 < m_2.$ Статистика за тестот: 1) $Z = \frac{\bar{X} - \bar{Y}}{\sqrt{n_2\sigma_1^2 + n_1\sigma_2^2}} \sqrt{n_1n_2}$, кога: примероците се земени од бесконечни популации, примероците се земени од конечни популации, при што изборот се врши со враќање, и примероците се земени од конечни популации, при што изборот се врши без враќање, но големината на примероците n_1 и n_2 е помала од 5% од големината на популациите N_1 и N_2, соодветно. 2) $Z = \frac{\bar{X} - \bar{Y}}{\sqrt{\frac{\sigma_1^2}{n_1} \cdot \frac{N_1 - n_1}{N_1 - 1} + \frac{\sigma_2^2}{n_2} \cdot \frac{N_2 - n_2}{N_2 - 1}}}$, кога: Примероците се земени од конечни популации, при што
--

изборот се врши без враќање, но големината на примероците n_1 и n_2 е поголема од 5 % од големината на популациите N_1 и N_2 , соодветно, со ниво на значајност α

А) за хипотезата $H_0 : m_1 = m_2$ наспроти алтернативната хипотеза $H_1 : m_1 \neq m_2$:

i) пресметуваме $\bar{z} = \frac{|\bar{x} - \bar{y}|}{\sqrt{n_2\sigma_1^2 + n_1\sigma_2^2}} \sqrt{n_1n_2}$ за тестот 1) и

$\bar{z} = \frac{|\bar{x} - \bar{y}|}{\sqrt{\frac{\sigma_1^2}{n_1} \cdot \frac{N_1 - n_1}{N_1 - 1} + \frac{\sigma_2^2}{n_2} \cdot \frac{N_2 - n_2}{N_2 - 1}}}$ за тестот 2), каде што $\bar{x}$ и $\bar{y}$ се

аритметички средини, т. е., просечни вредности

ii) одредуваме реален број $z_{\alpha/2}$ таков што $P_{H_0} \left\{ \bar{z} \geq z_{\alpha/2} \right\} = \frac{\alpha}{2}$,

односно $\Phi_0 \left(z_{\alpha/2} \right) = 1 - \frac{\alpha}{2}$ и заклучуваме:

- ако $\bar{z} \leq z_{\alpha/2}$ не ја отфрламе хипотезата H_0 ,

- ако $\bar{z} > z_{\alpha/2}$ ја отфрламе хипотезата H_0 со веројатност на

грешка од прв тип помала од α .

В) за хипотезата $H_0 : m_1 \leq m_2$ наспроти алтернативната хипотеза $H_1 : m_1 > m_2$:

i) пресметуваме $\bar{z} = \frac{\bar{x} - \bar{y}}{\sqrt{n_2\sigma_1^2 + n_1\sigma_2^2}} \sqrt{n_1n_2}$ за тестот 1) и

$\bar{z} = \frac{\bar{x} - \bar{y}}{\sqrt{\frac{\sigma_1^2}{n_1} \cdot \frac{N_1 - n_1}{N_1 - 1} + \frac{\sigma_2^2}{n_2} \cdot \frac{N_2 - n_2}{N_2 - 1}}}$ за тестот 2), каде што $\bar{x}$ и $\bar{y}$ се

аритметички средини, т. е., просечни вредности

ii) одредуваме реален број z_α таков што $P_{H_0} \left\{ \bar{z} \geq z_\alpha \right\} = 1 - \alpha$, односно $\Phi_0(z_\alpha) = 1 - \alpha$ и заклучуваме

- ако $\bar{z} \leq z_\alpha$, не ја отфрламе хипотезата H_0 ,

- ако $\bar{z} > z_\alpha$, ја отфрламе хипотезата H_0 со веројатност на

грешка од прв тип помала од α .

С) за хипотезата $H_0: m_1 \geq m_2$ наспроти алтернативната хипотеза $H_1: m_1 < m_2$:

i) пресметуваме $\bar{z} = \frac{\bar{x} - \bar{y}}{\sqrt{n_2\sigma_1^2 + n_1\sigma_2^2}} \sqrt{n_1n_2}$ за тестот 1) и

$\bar{z} = \frac{\bar{x} - \bar{y}}{\sqrt{\frac{\sigma_1^2}{n_1} \cdot \frac{N_1 - n_1}{N_1 - 1} + \frac{\sigma_2^2}{n_2} \cdot \frac{N_2 - n_2}{N_2 - 1}}}$ за тестот 2), каде што $\bar{x}$ и $\bar{y}$ се

аритметички средини, т. е., просечни вредности;

ii) одредуваме реален број z_α таков што:

$$P_{H_0} \{ \bar{z} \leq -z_\alpha \} = 1 - \alpha, \text{ односно } \Phi_0(z_\alpha) = 1 - \alpha$$

и заклучуваме:

- ако $\bar{z} \geq -z_\alpha$, не ја отфрламе хипотезата H_0 ,

- ако $\bar{z} < -z_\alpha$, ја отфрламе хипотезата H_0 со веројатност на грешка од прв тип помала од α .

t-тест

Наведените претпоставки кои се неопходни за да при тестирањето сè користи Z -тестот во пракса најчесто не се исполнети, пред се бидејќи дисперзиите на популациите σ_1 и σ_2 не се познати. Тогаш, не може да се пресмета вредноста на статистиката на Z -тестот. Затоа, прво се оценуваат дисперзиите σ_1 и σ_2 , со помош на дисперзиите на примероците s_1^2 и s_2^2 . Притоа, разликуваме два случаја:

1) кога се претпоставува дека дисперзиите на двете популации се еднакви, односно $\sigma_1^2 = \sigma_2^2$ и

2) кога се претпоставува дека дисперзиите на двете популации се различни, односно $\sigma_1^2 \neq \sigma_2^2$.

Прво, ќе го разгледаме првиот случај, односно кога важи $\sigma_1^2 = \sigma_2^2$ и тогаш добиваме нова статистика за тестот, т.н *статистика за t-тест*:

$$t_{n_1+n_2-2} = \frac{\bar{X} - \bar{Y}}{\sqrt{n_1s_1^2 + n_2s_2^2}} \sqrt{\frac{n_1n_2}{n_1+n_2}} (n_1+n_2-2),$$

која има Студентова распределба со $n_1 + n_2 - 2$ степени на слобода.

Во следната табела 3.7 се претставени условите за примена на t -тестот и постапката за тестирање хипотези за разликата меѓу математичките очекувања на две популации при непознати еднакви дисперзии, кога примероците се независни и мали.

Табела 3.3.1.2. Тестирање хипотези за разликата меѓу математичките очекувања на две популации при непознати еднакви дисперзии, кога примероците се независни со примена на t -тестот

Услови за примена на t -тестот:

Случајните примероци се независни.

Дисперзиите на популациите се непознати:

а. популациите имаат нормална распределба со еднакви дисперзии, односно $\sigma_1^2 = \sigma_2^2$;

б. примероците се со мала големина, т.е $n_1 \leq 30$ и $n_2 \leq 30$

Избор на обликот на нултата и алтернативната хипотеза:

$$H_0 : m_1 = m_2, \quad H_1 : m_1 \neq m_2;$$

$$H_0 : m_1 \leq m_2, \quad H_1 : m_1 > m_2;$$

$$H_0 : m_1 \geq m_2, \quad H_1 : m_1 < m_2.$$

Статистика за тестот:

$$t_{n_1+n_2-2} = \frac{\bar{X} - \bar{Y}}{\sqrt{n_1 s_1^2 + n_2 s_2^2}} \sqrt{\frac{n_1 n_2}{n_1 + n_2} (n_1 + n_2 - 2)}$$

Со ниво на значајност α

А) За хипотезата $H_0 : m_1 = m_2$ наспроти алтернативната хипотеза $H_1 : m_1 \neq m_2$:

i) пресметуваме:

$$\bar{t}_{n_1+n_2-2} = \frac{|\bar{x} - \bar{y}|}{\sqrt{n_1 s_1^2 + n_2 s_2^2}} \sqrt{\frac{n_1 n_2}{n_1 + n_2} (n_1 + n_2 - 2)},$$

каде што $\bar{x}$ и $\bar{y}$ се аритметички средини, а s_1^2 и s_2^2 се дисперзии на примероците;

ii) читаме вредност $t_\alpha (n_1 + n_2 - 2)$ од таблицата за Студентова распределба за двостран тест и заклучуваме:

- ако $\bar{t}_{n_1+n_2-2} \leq t_\alpha (n_1 + n_2 - 2)$, не ја отфрламе хипотезата H_0 ,

- ако $\bar{t}_{n_1+n_2-2} > t_\alpha (n_1 + n_2 - 2)$, ја отфрламе хипотезата H_0 со веројатност на грешка од прв тип помала од α .

В) За хипотезата $H_0 : m_1 \leq m_2$ наспроти алтернативната хипотеза $H_1 : m_1 > m_2$:

$$\text{i) пресметуваме } \bar{t}_{n_1+n_2-2} = \frac{\bar{x} - \bar{y}}{\sqrt{n_1 s_1^2 + n_2 s_2^2}} \sqrt{\frac{n_1 n_2}{n_1 + n_2} (n_1 + n_2 - 2)},$$

каде што $\bar{x}$ и $\bar{y}$ се аритметички средини, а s_1^2 и s_2^2 се дисперзии на примероците:

ii) читаме вредност $t_\alpha (n_1 + n_2 - 2)$ од таблицата за Студентова распределба за едностран тест и заклучуваме

- ако $\bar{t}_{n_1+n_2-2} \leq t_\alpha (n_1 + n_2 - 2)$, не ја отфрламе хипотезата H_0 ,

- ако $\bar{t}_{n_1+n_2-2} > t_\alpha (n_1 + n_2 - 2)$, ја отфрламе хипотезата H_0 со веројатност на грешка од прв тип помала од α .

С) За хипотезата $H_0 : m \geq m_0$, наспроти алтернативната хипотеза $H_1 : m < m_0$:

$$\text{i) пресметуваме } \bar{t}_{n_1+n_2-2} = \frac{\bar{x} - \bar{y}}{\sqrt{n_1 s_1^2 + n_2 s_2^2}} \sqrt{\frac{n_1 n_2}{n_1 + n_2} (n_1 + n_2 - 2)},$$

каде што $\bar{x}$ и $\bar{y}$ се аритметички средини, а s_1^2 и s_2^2 се дисперзии на примероците

ii) читаме вредност $t_\alpha (n_1 + n_2 - 2)$ од таблицата за Студентова распределба за едностран тест и заклучуваме:

- ако $\bar{t}_{n_1+n_2-2} \geq -t_\alpha (n_1 + n_2 - 2)$ не ја отфрламе хипотезата H_0 ,

- ако $\bar{t}_{n_1+n_2-2} < -t_\alpha (n_1 + n_2 - 2)$ ја отфрламе хипотезата H_0 со веројатност на грешка од прв тип помала од α .

Задача 3.3.1.3. Еден начин да се измери кондицијата на една личност е да се измери процентот на маснотии во телото. Просечниот процент на маснотии во телото варира во зависност од возраста, но според некои упатства, нормалниот опсег за мажите е 15 – 20 % телесни масти, а нормалниот опсег за жените е 20 – 25 % телесни масти. Разгледуваме примерок на податоци од група мажи и жени кои вежбале во теретана три пати неделно во текот на една

година. Потоа, нивниот тренер ги измерил телесните масти. Табелата подолу ги прикажува податоците.

Група	Процент на маснотии				
Мажи	13,3	6,0	20,0	8,0	14,0
	19,0	18,0	25,0	16,0	24,0
	15,0	1,0	15,0		
Жени	22,0	16,0	21,7	21,0	30,0
	26,0	12,0	23,2	28,0	23,0

Што можете да заклучите врз основа на овој примерок, доколу популациите имаат нормална распределба со еднакви стандардни отстапувања и со прифатлив ризик од 0,05?

Решение. Разгледуваме независни примероци со големини $n_1 = 13 < 30$ и $n_2 = 10 < 30$, земени од популации со нормална распределба и еднакви стандардни отстапувања. Сакаме да го провериме дали постои разлика помеѓу просечниот процент на маснотии во телото помеѓу мажите и жените.

Заради претпоставката на проблемот употребуваме t -тест за следнава хипотеза:

$$H_0 : m_1 = m_2, \text{ наспроти алтернативната хипотеза } H_1 : m_1 \neq m_2,$$

а статистиката за тестот е

$$\bar{t}_{n_1+n_2-2} = \frac{|\bar{x} - \bar{y}|}{\sqrt{n_1 s_1^2 + n_2 s_2^2}} \sqrt{\frac{n_1 n_2}{n_1 + n_2} (n_1 + n_2 - 2)}.$$

Притоа, аритметичката средина на примерокот од групата мажи е:

$$\bar{x} = \frac{13,3 + 6 + 20 + 8 + 14 + 19 + 18 + 25 + 16 + 24 + 15 + 1 + 15}{13} = 14,95$$

а варијансата на примерокот од групата мажи е:

$$\begin{aligned} s_1^2 &= \frac{1}{n_1 - 1} \left(\sum_{i=1}^k f_i \cdot x_i^2 - n \bar{x}^2 \right) = \\ &= \frac{1}{12} (13,3^2 + 6^2 + 20^2 + 8^2 + 14^2 + 19^2 + 18^2 + 25^2 + 16^2 + 24^2 + 2 \cdot 15^2 + 1^2 \\ &\quad - 13 \cdot 14,95^2) = \frac{1}{12} (3465,89 - 2905,533) = 46,7. \end{aligned}$$

Притоа, аритметичката средина на примерокот од групата жени е:

$$\bar{y} = \frac{22+16+21,7+21+30+26+12+23,2+28+23}{10} = 22,29$$

а варијансата на примерокот од групата мажи е:

$$\begin{aligned} s_2^2 &= \frac{1}{n_2-1} \left(\sum_{i=1}^k f_i \cdot x_i^2 - n\bar{x}^2 \right) = \\ &= \frac{1}{9} (22^2 + 16^2 + 21,7^2 + 21^2 + 30^2 + 26^2 + 12^2 + 23,2^2 + 28^2 + 23^2 - 10 \cdot 22,9^2) \\ &= \frac{1}{9} (5223,13 - 4968,441) = 28,3. \end{aligned}$$

Затоа, ја пресметуваме специфичната вредност:

$$\bar{t}_{21} = \frac{|14,95 - 22,29|}{\sqrt{13 \cdot 46,7 + 10 \cdot 28,3}} \sqrt{\frac{13 \cdot 10}{13+10} (13+10-2)} = 2,69$$

од табелата за Студентова распределба, двостран тест за $\alpha = 0,05$ и $n_1 + n_2 - 2 = 21$ степени на слобода за критичната вредност имаме:

$$t_{0,05}(21) = 2,08$$

Бидејќи:

$$\bar{t}_{21} = 2,69 > 2,08 = t_{0,05}(21)$$

нултата хипотеза е отфрлена, со прифатлив ризик од грешка од типот I помала од 0,05. Значи дека имаме доволно причини да мислиме дека постои разлика помеѓу просечниот процент на маснотии во телото помеѓу мажите и жените.

2) Во претходниот тест претпоставивме дека дисперзиите на популациите се еднакви. Но, понекогаш дисперзиите се различни, односно $\sigma_1^2 \neq \sigma_2^2$. Во ваков случај, ако примероците се мали се

применува модифицираниот t -тест: $t = \frac{\bar{X} - \bar{Y}}{\sqrt{n_2 s_1^2 + n_1 s_2^2}} \sqrt{n_1 n_2}$, која има

Студентова распределба со:

$$k = \left[\frac{(n_2 s_1^2 + n_1 s_2^2)^2 (n_1 - 1)(n_2 - 1)}{(n_2 - 1)(n_2 s_1^2)^2 + (n_1 - 1)(n_1 s_2^2)^2} \right]$$

степени на слобода.

Во следната табела 3.8. се претставени условите за примена на t -тестот и постапката за тестирање на хипотези за разликата меѓу математичките очекувања на две популации при непознати различни дисперзии, кога примероците се независни и мали.

Табела 3.3.1.4. Тестирање хипотези за разликата меѓу математичките очекувања на две популации при непознати различни дисперзии, кога примероците се независни со примена на t -тестот

Услови за примена на t -тестот:

Случајните примероци се независни.

Дисперзиите на популациите се непознати:

а. Популациите имаат нормална распределба со различни дисперзии, односно $\sigma_1^2 \neq \sigma_2^2$;

б. примероците се со мала големина, т. е. $n_1 \leq 30$ и $n_2 \leq 30$

Избор на обликот на нултата и алтернативната хипотеза:

$$H_0 : m_1 = m_2, \quad H_1 : m_1 \neq m_2;$$

$$H_0 : m_1 \leq m_2, \quad H_1 : m_1 > m_2;$$

$$H_0 : m_1 \geq m_2, \quad H_1 : m_1 < m_2.$$

Статистика за тестот:

$$t = \frac{\bar{X} - \bar{Y}}{\sqrt{n_2 s_1^2 + n_1 s_2^2}} \sqrt{n_1 n_2}$$

Со ниво на значајност α

А) За хипотезата $H_0 : m_1 = m_2$, наспроти алтернативната хипотеза $H_1 : m_1 \neq m_2$:

i) пресметуваме $\bar{t} = \frac{|\bar{x} - \bar{y}|}{\sqrt{n_2 s_1^2 + n_1 s_2^2}} \sqrt{n_1 n_2}$, каде што $\bar{x}$ и $\bar{y}$ се

аритметички средини, а s_1^2 и s_2^2 се дисперзии на примероците;

ii) читаме вредност $t_\alpha(k)$ од таблицата за Студентова распределба за двостран тест, каде што:

$$k = \left[\frac{(n_2 s_1^2 + n_1 s_2^2)^2 (n_1 - 1)(n_2 - 1)}{(n_2 - 1)(n_2 s_1^2)^2 + (n_1 - 1)(n_1 s_2^2)^2} \right]$$

и заклучуваме:

- ако $\bar{t} \leq t_{\alpha}(k)$, не ја отфрламе хипотезата H_0 ,

- ако $\bar{t} > t_{\alpha}(k)$, ја отфрламе хипотезата H_0 со веројатност на грешка од прв тип помала од α .

В) За хипотезата $H_0: m_1 \leq m_2$, наспроти алтернативната хипотеза $H_1: m_1 > m_2$:

i) пресметуваме $\bar{t} = \frac{\bar{x} - \bar{y}}{\sqrt{n_2 s_1^2 + n_1 s_2^2}} \sqrt{n_1 n_2}$, каде што $\bar{x}$ и $\bar{y}$ се

аритметички средини, а s_1^2 и s_2^2 се дисперзии на примероците;

ii) читаме вредност $t_{\alpha}(k)$ од таблицата за Студентова распределба за едностран тест, каде што:

$$k = \left[\frac{(n_2 s_1^2 + n_1 s_2^2)^2 (n_1 - 1)(n_2 - 1)}{(n_2 - 1)(n_2 s_1^2)^2 + (n_1 - 1)(n_1 s_2^2)^2} \right]$$

и заклучуваме:

- ако $\bar{t} \leq t_{\alpha}(k)$ не ја отфрламе хипотезата H_0 ,

- ако $\bar{t} > t_{\alpha}(k)$ ја отфрламе хипотезата H_0 со веројатност на грешка од прв тип помала од α .

С) За хипотезата $H_0: m \geq m_0$, наспроти алтернативната хипотеза $H_1: m < m_0$:

i) пресметуваме $\bar{t} = \frac{\bar{x} - \bar{y}}{\sqrt{n_2 s_1^2 + n_1 s_2^2}} \sqrt{n_1 n_2}$, каде што $\bar{x}$ и $\bar{y}$ се

аритметички средини, а s_1^2 и s_2^2 се дисперзии на примероците

ii) читаме вредност $t_{\alpha}(k)$ од таблицата за Студентова распределба за едностран тест, каде што:

$$k = \left[\frac{(n_2 s_1^2 + n_1 s_2^2)^2 (n_1 - 1)(n_2 - 1)}{(n_2 - 1)(n_2 s_1^2)^2 + (n_1 - 1)(n_1 s_2^2)^2} \right]$$

и заклучуваме:

- ако $\bar{t} \geq -t_{\alpha}(k)$, не ја отфрламе хипотезата H_0 ,

- ако $\bar{t} < -t_{\alpha}(k)$, ја отфрламе хипотезата H_0 со веројатност на грешка од прв тип помала од α .

Задача 3.3.1.5. За да се спореди висината на мажите од САД и од Шведска, по случаен избор е избран примерок од 30 мажи од секоја земја и во согласност со измерените висини добиени се просечната висина на примерокот од САД $\bar{x} = 177,038$, со варијанса на примерокот од САД $s_1^2 = 23,114$ и просечната висина на примерокот од Шведска $\bar{y} = 181,6354$, со варијанса на примерокот од Шведска $s_2^2 = 14,7066$. Дали може да се заклучи дека просечната висина на мажите од САД е помала од просечната висина на мажите од Шведска, доколу популациите имаат нормална распределба со различни стандардни отстапувања и со прифатлив ризик од 0,05?

Решение. Разгледуваме независни примероци со големини $n_1 = 30$ и $n_2 = 30$, земени од популации со нормална распределба и различни стандардни отстапувања. Сакаме да провериме дали просечната висина на мажите од САД е помала од просечната висина на мажите од Шведска (алтернативна хипотеза).

Заради претпоставката на проблемот употребуваме t -тест за следнава хипотеза:

$$H_0 : m_1 \geq m_2, \text{ наспроти алтернативната хипотеза } H_1 : m_1 < m_2,$$

а статистиката за тестот е:

$$\bar{t} = \frac{\bar{x} - \bar{y}}{\sqrt{n_2 s_1^2 + n_1 s_2^2}} \sqrt{n_1 n_2}.$$

и во согласност со истатата специфичната вредност на тестот изнесува

$$\bar{t} = \frac{177,038 - 181,6354}{\sqrt{30 \cdot 23,114 + 30 \cdot 14,7066}} \sqrt{30 \cdot 30} = -4,095$$

од табелата за Студентова распределба, едностран тест за $\alpha = 0,05$ и

$$k = \left[\frac{(n_2 s_1^2 + n_1 s_2^2)^2 (n_1 - 1)(n_2 - 1)}{(n_2 - 1)(n_2 s_1^2)^2 + (n_1 - 1)(n_1 s_2^2)^2} \right] =$$

$$\left[\frac{(30 \cdot 23,114 + 30 \cdot 14,7066)^2 (30 - 1)(30 - 1)}{(30 - 1)(30 \cdot 23,114)^2 + (30 - 1)(30 \cdot 14,7066)^2} \right] = 55$$

степени на слобода, за критичната вредност имаме:

$$t_{0,05}(55) = \frac{1,671 + 1,676}{2} = 1,67.$$

Бидејќи $\bar{t} = -4,095 < -1,67 = -t_{0,05}(55)$, нултата хипотеза е отфрлена, со прифатлив ризик од грешка од типот I помала од 0,05. Значи дека имаме доволно причини да мислиме дека просечната висина на мажите од САД е помала од просечната висина на мажите од Шведска, со прифатлив ризик од грешка од типот I помала од 0,05.

Забелешка. Бидејќи со зголемување на големината на примерокот Студентовата распределба тежи кон Z -распределбата, тогаш за многу големи вредности на n , критичните t -вредности стануваат приближно еднакви на критичните z -вредности. Затоа, во пракса се прифаќа дека, *ако статистичките множества имаат нормална распределба, со непозната дисперзија и примероците се многу големи (по некои автори доволно е $n_1 > 30$ и $n_2 > 30$), тестирањето може да се спроведе со примена на Z -тестот наместо t -тестот.*

Во следните табели 3.9 и 3.10 се претставени условите за примена на Z -тестот и постапката за тестирање хипотези за разликата меѓу математичките очекувања на две популации при непознати еднакви и различни дисперзии, соодветно, кога примероците се независни во случај на големи примероци.

Табела 3.3.1.6. Тестирање хипотези за математичко очекување при непознати еднакви дисперзии со големи примероци

<i>Услови за примена на Z-тестот:</i>	
Случајните примероци се независни.	
Дисперзиите на популациите се непознати и еднакви:	
а. популациите имаат нормална распределба	
б. примероците се големи, т.е $n_1 > 30$ и $n_2 > 30$.	
Избор на обликот на нултата и алтернативната хипотеза:	
$H_0 : m_1 = m_2,$	$H_1 : m_1 \neq m_2;$
$H_0 : m_1 \leq m_2,$	$H_1 : m_1 > m_2;$
$H_0 : m_1 \geq m_2,$	$H_1 : m_1 < m_2.$
Статистика за тестот:	

$$Z = \frac{\bar{X} - \bar{Y}}{\sqrt{n_1 s_1^2 + n_2 s_2^2}} \sqrt{\frac{n_1 n_2}{n_1 + n_2} (n_1 + n_2 - 2)}$$

Со ниво на значајност α

А) За хипотезата $H_0 : m = m_0$ наспроти алтернативната хипотеза $H_1 : m \neq m_0$:

i) пресметуваме

$$\bar{z} = \frac{|\bar{x} - \bar{y}|}{\sqrt{n_1 s_1^2 + n_2 s_2^2}} \sqrt{\frac{n_1 n_2}{n_1 + n_2} (n_1 + n_2 - 2)},$$

каде што $\bar{x}$ и $\bar{y}$ се аритметички средини, а s_1^2 и s_2^2 се дисперзии на примероците:

ii) одредуваме реален број $z_{\alpha/2}$ таков што $P_{H_0} \left\{ \bar{z} \geq z_{\alpha/2} \right\} = \frac{\alpha}{2}$,

односно $\Phi_0 \left(z_{\alpha/2} \right) = 1 - \frac{\alpha}{2}$ и заклучуваме:

- ако $\bar{z} \leq z_{\alpha/2}$ не ја отфрламе хипотезата H_0 ,

- ако $\bar{z} > z_{\alpha/2}$ ја отфрламе хипотезата H_0 со веројатност на

грешка од прв тип помала од α .

В) за хипотезата $H_0 : m \leq m_0$ наспроти алтернативната хипотеза $H_1 : m > m_0$:

i) пресметуваме $\bar{z} = \frac{\bar{x} - \bar{y}}{\sqrt{n_1 s_1^2 + n_2 s_2^2}} \sqrt{\frac{n_1 n_2}{n_1 + n_2} (n_1 + n_2 - 2)}$, каде

што $\bar{x}$ и $\bar{y}$ се аритметички средини, а s_1^2 и s_2^2 се дисперзии на примероците;

ii) одредуваме реален број z_α таков што $P_{H_0} \left\{ \bar{z} \geq z_\alpha \right\} = 1 - \alpha$, односно $\Phi_0 \left(z_\alpha \right) = 1 - \alpha$ и заклучуваме:

- ако $\bar{z} \leq z_\alpha$, не ја отфрламе хипотезата H_0 ,

- ако $\bar{z} > z_\alpha$, ја отфрламе хипотезата H_0 со веројатност на грешка од прв тип помала од α .

С) За хипотезата $H_0: m \geq m_0$, наспроти алтернативната хипотеза $H_1: m < m_0$:

i) пресметуваме
$$\bar{z} = \frac{\bar{x} - \bar{y}}{\sqrt{n_1 s_1^2 + n_2 s_2^2}} \sqrt{\frac{n_1 n_2}{n_1 + n_2} (n_1 + n_2 - 2)},$$
 каде

$\bar{x}$ и $\bar{y}$ се аритметички средини, а s_1^2 и s_2^2 се дисперзии на примероците

ii) одредуваме реален број z_α таков што $P_{H_0} \{ \bar{z} \leq -z_\alpha \} = 1 - \alpha$, односно $\Phi_0(z_\alpha) = 1 - \alpha$ и заклучуваме

- ако $\bar{z} \geq -z_\alpha$, не ја отфрламе хипотезата H_0 ,

- ако $\bar{z} < -z_\alpha$, ја отфрламе хипотезата H_0 со веројатност на грешка од прв тип помала од α .

Задача 3.3.1.7. Истражувачите сакаат да споредат два вида локална анестезија за да утврдат дали има разлика во времето до почетокот на губењето болка. Се набљудуваат две групи од по 151 пациент со големина од 151 и се мери времето до почетокот на дејството на двата анестетика соодветно на двете групи. Просечното време до почетокот на дејството на првата анестезија во првиот примерок е 9 минути, додека просечното време до почетокот на дејството на втората анестезија во вториот примерок е 11 минути. Познато е дека стандардното отстапување за секоја од групите е 4,5 и дека популациите имаат еднакви варијанси. Не е познато кој третман има помало време за губење на болката. Што можете да заклучите врз основа на овие примероци, со прифатлив ризик од 0,02?

Решение. Разгледуваме независни примероци со големини $n_1 = n_2 = 151 > 30$, аритметички средина $\bar{x} = 9$ и $\bar{y} = 11$ и стандардни отстапувања $s_1 = s_2 = 4,5$ земени од бесконечни популации со еднакви варијанси. Она што сакаме да го провериме е дали постои разлика во времето до почетокот на дејството на двете анестезии (алтернативна хипотеза), затоа употребуваме z -тест за следнава хипотеза:

$$H_0: m_1 = m_2, \text{ наспроти алтернативната хипотеза } H_1: m_1 \neq m_2,$$

а статистиката за тестот е:

$$Z = \frac{\bar{X} - \bar{Y}}{\sqrt{n_1 s_1^2 + n_2 s_2^2}} \sqrt{\frac{n_1 n_2}{n_1 + n_2} (n_1 + n_2 - 2)}.$$

Тогаш специфичната вредност изнесува:

$$\begin{aligned} \bar{z} &= \frac{|\bar{x} - \bar{y}|}{\sqrt{n_1 s_1^2 + n_2 s_2^2}} \sqrt{\frac{n_1 n_2}{n_1 + n_2} (n_1 + n_2 - 2)} = \\ &= \frac{|9 - 11|}{\sqrt{151 \cdot 4,5 + 151 \cdot 4,5}} \sqrt{\frac{151 \cdot 151}{151 + 151} (151 + 151 - 2)} = \frac{2}{\sqrt{1359}} \cdot \sqrt{22650} = 8,16, \end{aligned}$$

а од табелата за нормална распределба за критичната вредност $z_{0,02}$ имаме:

$$\Phi_0(2,32) = 0,9898 < \Phi_0(z_{0,02}) = 1 - 0,01 = 0,99 < 0,9901 = \Phi_0(2,33)$$

$$\text{или } z_{0,01} = \frac{2,32 + 2,33}{2} = 2,325.$$

Бидејќи $\bar{z} = 8,16 > 2,325 = z_{0,02}$, нултата хипотеза ја отфрламе, со прифатлив ризик од грешка од типот I помала од 0,01. Значи во согласност со примероците постојат доволно причини да сметаме дека постои статистички значајна разлика во должината на времето до почетокот на дејството на двете локални анестезии.

Табела 3.3.1.8. Тестирање хипотези за математичко очекување при непознати различни дисперзии со големи примероци

<i>Услови за примена на Z -тестот:</i>
Случајните примероци се независни. Дисперзиите на популациите се непознати и различни: а. Популациите имаат нормална распределба б. Примероците се со големи, т.е $n_1 > 30$ и $n_2 > 30$.
Избор на обликот на нултата и алтернативната хипотеза:
$H_0 : m_1 = m_2, \quad H_1 : m_1 \neq m_2;$
$H_0 : m_1 \leq m_2, \quad H_1 : m_1 > m_2;$
$H_0 : m_1 \geq m_2, \quad H_1 : m_1 < m_2.$
Статистика за тестот:

$$Z = \frac{\bar{X} - \bar{Y}}{\sqrt{n_2 s_1^2 + n_1 s_2^2}} \sqrt{n_1 n_2}$$

Со ниво на значајност α

А) За хипотезата $H_0: m = m_0$, наспроти алтернативната хипотеза $H_1: m \neq m_0$:

i) пресметуваме $\bar{z} = \frac{|\bar{x} - \bar{y}|}{\sqrt{n_2 s_1^2 + n_1 s_2^2}} \sqrt{n_1 n_2}$, каде што $\bar{x}$ и $\bar{y}$ се

аритметички средини, а s_1^2 и s_2^2 се дисперзии на примероците

ii) одредуваме реален број $z_{\alpha/2}$ таков што $P_{H_0} \left\{ \bar{z} \geq z_{\alpha/2} \right\} = \frac{\alpha}{2}$,

односно $\Phi_0 \left(z_{\alpha/2} \right) = 1 - \frac{\alpha}{2}$ и заклучуваме

- ако $\bar{z} \leq z_{\alpha/2}$, не ја отфрламе хипотезата H_0 ,

- ако $\bar{z} > z_{\alpha/2}$, ја отфрламе хипотезата H_0 со веројатност на грешка од прв тип помала од α .

В) За хипотезата $H_0: m \leq m_0$, наспроти алтернативната хипотеза $H_1: m > m_0$:

i) пресметуваме $\bar{z} = \frac{\bar{x} - \bar{y}}{\sqrt{n_2 s_1^2 + n_1 s_2^2}} \sqrt{n_1 n_2}$, каде што $\bar{x}$ и $\bar{y}$ се

аритметички средини, а s_1^2 и s_2^2 се дисперзии на примероците

ii) одредуваме реален број z_α таков што $P_{H_0} \left\{ \bar{z} \geq z_\alpha \right\} = 1 - \alpha$, односно $\Phi_0 \left(z_\alpha \right) = 1 - \alpha$ и заклучуваме:

- ако $\bar{z} \leq z_\alpha$, не ја отфрламе хипотезата H_0 ,

- ако $\bar{z} > z_\alpha$, ја отфрламе хипотезата H_0 со веројатност на грешка од прв тип помала од α .

С) За хипотезата $H_0: m \geq m_0$, наспроти алтернативната хипотеза $H_1: m < m_0$:

i) пресметуваме $\bar{z} = \frac{\bar{x} - \bar{y}}{\sqrt{n_2 s_1^2 + n_1 s_2^2}} \sqrt{n_1 n_2}$, каде $\bar{x}$ и $\bar{y}$ се

аритметички средини, а s_1^2 и s_2^2 се дисперзии на примероците

ii) одредуваме реален број z_α таков што

$P_{H_0} \{ \bar{z} \leq -z_\alpha \} = 1 - \alpha$, односно $\Phi_0(z_\alpha) = 1 - \alpha$ и заклучуваме

- ако $\bar{z} \geq -z_\alpha$, не ја отфрламе хипотезата H_0 ,

- ако $\bar{z} < -z_\alpha$, ја отфрламе хипотезата H_0 со веројатност на грешка од прв тип помала од α .

Задача 3.3.1.9. Направен е бајпас на срцето со ограничен пристап на 48 пациенти, а конвенционална операција на 55 пациенти. Должината на престојот во болница на пациентите на кој е извршен бајпас со ограничен пристап се очекува да биде помала од должината на престојот во болница на пациентите со конвенционална операција. Ако средната должина на престој на пациентите на кој е извршен бајпас со ограничен пристап е 3,5 дена со стандардно отстапување на примерокот 1,5 дена, додека средната должина на престој на пациентите со конвенционална операција е 8 дена со стандардно отстапување на примерокот 2 дена, што можете да заклучите врз основа на овие примероци, со прифатлив ризик од 0,01?

Решение. Разгледуваме независни примероци со големини $n_1 = 48 > 30$ и $n_2 = 55 > 30$, аритметички средина $\bar{x} = 3,5$ и $\bar{y} = 8$ и стандардни отстапувања $s_1 = 1,5$ и $s_2 = 2$, соодветно земени од бесконечни популации со различни варијанси. Она што сакаме да го провериме е дали должината на престојот во болница на пациентите на кој е извршен бајпас со ограничен пристап е помала од должината на престојот во болница на пациентите со конвенционална операција (алтернативна хипотеза), затоа употребуваме z -тест за следнава хипотеза:

$H_0 : m_1 = m_2$, наспроти алтернативната хипотеза $H_1 : m_1 < m_2$,

а статистиката за тестот е:

$$Z = \frac{\bar{X} - \bar{Y}}{\sqrt{n_2 s_1^2 + n_1 s_2^2}} \sqrt{n_1 n_2}.$$

Тогаш специфичната вредност изнесува:

$$\begin{aligned} \bar{z} &= \frac{\bar{x} - \bar{y}}{\sqrt{n_2 s_1^2 + n_1 s_2^2}} \sqrt{n_1 n_2} = \frac{3,5 - 8}{\sqrt{55 \cdot 1,5^2 + 48 \cdot 2^2}} \sqrt{48 \cdot 55} = \\ &= \frac{-4,5}{\sqrt{123,75 + 192}} 51,38 = \frac{-231,21}{17,77} = -13. \end{aligned}$$

а од табелата за нормална распределба за критичната вредност $z_{0,01}$ имаме:

$$\Phi_0(2,32) = 0,9898 < \Phi_0(z_{0,01}) = 1 - 0,01 = 0,99 < 0,9901 = \Phi_0(2,33)$$

$$\text{или } z_{0,01} = \frac{2,32 + 2,33}{2} = 2,325.$$

Бидејќи $\bar{z} = -13 < -2,325 = -z_{0,01}$ нултата хипотеза ја отфрламе, со прифатлив ризик од грешка од типот I помала од 0,01. Значи дека во согласност со примерокот постојат доволно причини да сметаме дека должината на престојот во болница на пациентите на кој е извршен бајпас со ограничен пристап е помала од должината на престојот во болница на пациентите со конвенционална операција.

3.3.2. ТЕСТОВИ ЗА ЕДНАКВОСТ НА МАТЕМАТИЧКИТЕ ОЧЕКУВАЊА ПРИ ЗАВИСНИ ПРИМЕРОЦИ

Кога сакаме да тестираме хипотеза за разликата на математичките очекувања со помош на заемно зависни примероци, тогаш претходно споменатите тестови не можеме да ги користиме, за што е причина корелацијата на обележјата. Во пракса, ваков случај вообичаено се јавува кога се споредуваат просечните резултати од ефектот на два анестетици на иста група луѓе.

При тестирањето хипотези за разликата на математичките очекувања со помош на заемно зависни примероци се разгледуваат примероци кои имаат еднакви големини, дадени во вид на парови (X_i, Y_i) , $i = 1, 2, \dots, n$. Тогаш, се формира нов примерок

$$Z_i = X_i - Y_i, i = 1, 2, \dots, n.$$

Ако математичките очекувања на популациите се m_1 и m_2 , соодветно, тогаш математичкото очекување на променливата:

$$\bar{Z}_n = \frac{1}{n} \sum_{i=1}^n Z_i$$

изнесува $m_1 - m_2$. Па, затоа статистиката за тестирањето на хипотезите за разликата на математичките очекувања со помош на заемно зависни примероци е стандардизираната форма на променливата $\bar{Z}$.

Притоа, ако станува збор за примероци со големина $n \leq 30$, стандардизираната форма на $\bar{Z}_n$ е од обликот $\frac{\bar{Z}_n}{S_z} \sqrt{n}$, каде што S_z е стандардно отстапување на променливата Z и има Студентова t -распределба со $n-1$ степени на слобода.

Во следната табела 3.11 се претставени условите за примена на t -тестот и постапката за тестирање хипотези за разликата меѓу математичките очекувања на две популации кога примероците се зависни и мали.

Табела 3.3.2.1. Тестирање хипотези за разликата меѓу математичките очекувања на две популации при зависни примероци со примена на t -тестот

Услови за примена на t -тестот:

Случајните примероци се зависни.

Примероците имаат големина $n \leq 30$.

Избор на обликот на нултата и алтернативната хипотеза:

$$H_0 : m_1 = m_2, \quad H_1 : m_1 \neq m_2 ;$$

$$H_0 : m_1 \leq m_2, \quad H_1 : m_1 > m_2 ;$$

$$H_0 : m_1 \geq m_2, \quad H_1 : m_1 < m_2 .$$

Статистика за тестот:

1) $t = \frac{\bar{Z}_n}{\sigma_z} \sqrt{n}$, кога е позната дисперзијата σ_z на променливата $Z = X - Y$;

2) $t = \frac{\bar{Z}_n}{S_z} \sqrt{n}$, кога не е позната дисперзијата на променливата $Z = X - Y$.

Со ниво на значајност α

А) За хипотезата $H_0 : m_1 = m_2$ наспроти алтернативната хипотеза $H_1 : m_1 \neq m_2$:

и) пресметуваме $\bar{t} = \frac{|\bar{z}_n|}{\sigma_z} \sqrt{n}$ за тестот 1) или $\bar{t} = \frac{|\bar{z}_n|}{S_z} \sqrt{n}$ за тестот 2), каде што $\bar{z}_n = \frac{1}{n} \sum_{i=1}^n (x_i - y_i)$ е аритметичката средина на

променливата $Z = X - Y$, а $s_z^2 = \frac{1}{n-1} \left(\sum_{i=1}^n (x_i - y_i)^2 - n\bar{z}_n^2 \right)$ е

стандардно отстапување на променливата Z ;

ii) читаме вредност $t_\alpha(n-1)$ од таблицата за Студентова распределба за двостран тест и заклучуваме:

- ако $\bar{t} \leq t_\alpha(n-1)$, не ја отфрламе хипотезата H_0 ,

- ако $\bar{t} > t_\alpha(n-1)$, ја отфрламе хипотезата H_0 со веројатност на грешка од прв тип помала од α .

В) За хипотезата $H_0: m_1 \leq m_2$ наспроти алтернативната хипотеза $H_1: m_1 > m_2$:

i) пресметуваме $\bar{t} = \frac{\bar{z}_n}{\sigma_z} \sqrt{n}$ за тестот 1) или $\bar{t} = \frac{\bar{z}_n}{s_z} \sqrt{n}$ за

тестот 2), каде што $\bar{z}_n = \frac{1}{n} \sum_{i=1}^n (x_i - y_i)$ е аритметичката средина на

променливата $Z = X - Y$, а $s_z^2 = \frac{1}{n-1} \left(\sum_{i=1}^n (x_i - y_i)^2 - n\bar{z}_n^2 \right)$ е

стандардно отстапување на променливата Z ;

ii) читаме вредност $t_\alpha(n-1)$ од таблицата за Студентова распределба за едностран тест и заклучуваме

- ако $\bar{t} \leq t_\alpha(n-1)$, не ја отфрламе хипотезата H_0 ,

- ако $\bar{t} > t_\alpha(n-1)$, ја отфрламе хипотезата H_0 со веројатност на грешка од прв тип помала од α .

С) За хипотезата $H_0: m \geq m_0$, наспроти алтернативната хипотеза $H_1: m < m_0$:

i) пресметуваме $\bar{t} = \frac{\bar{z}_n}{\sigma_z} \sqrt{n}$ за тестот 1) или $\bar{t} = \frac{\bar{z}_n}{s_z} \sqrt{n}$ за

тестот 2), каде што $\bar{z}_n = \frac{1}{n} \sum_{i=1}^n (x_i - y_i)$ е аритметичката средина на

променливата $Z = X - Y$, а $s_z^2 = \frac{1}{n-1} \left(\sum_{i=1}^n (x_i - y_i)^2 - n\bar{z}_n^2 \right)$ е

стандардно отстапување на променливата Z

ii) читаме вредност $t_\alpha(n-1)$ од таблицата за Студентова

распределба за едностран тест и заклучуваме:

- ако $\bar{t} \geq -t_{\alpha}(n-1)$, не ја отфрламе хипотезата H_0 ,
- ако $\bar{t} < -t_{\alpha}(n-1)$, ја отфрламе хипотезата H_0 со веројатност на грешка од прв тип помала од α .

Задача 3.3.2.2. Спроведена е социолошка студија за бракот и семејството и еден од неколкуте фактори на интерес е нивото на образование на сопружниците. Во табелата подолу е даден бројот на години на образование надвор од средното образование за 15 парови.

Брачен пар	Сопруг	Сопруга	Разлика
1	4	2	2
2	0	4	-4
3	7	6	1
4	4	4	0
5	2	3	-1
6	8	4	4
7	4	4	0
8	3	1	2
9	1	0	1
10	2	2	0
11	4	4	0
12	8	2	6
13	0	4	-4
14	4	0	4
15	1	4	-3

Овие податоци се искористени за тестирање на хипотезата дека постои разлика помеѓу средната вредност на годините на сопружниците поминати во образование надвор од средното образование со ниво на значајност од 0,05?

Решение. Разгледуваме зависни примероци со еднакви големини $n_1 = n_2 = 15 < 30$. Сакаме да провериме дали постои разлика помеѓу средната вредност на годините на сопружниците поминати во образование надвор од средното образование (алтернативна хипотеза).

Заради претпоставката на проблемот употребуваме t -тест за следнава хипотеза:

$H_0 : m_1 = m_2$ наспроти алтернативната хипотеза $H_1 : m_1 \neq m_2$,

а статистиката за тестот е $t = \frac{\bar{Z}_n}{S_z} \sqrt{n}$ и во согласност со истатата,

специфичната вредност на тестот изнесува: $\bar{t} = \frac{|\bar{Z}_n|}{S_z} \sqrt{n}$ каде што:

$$\bar{Z}_n = \frac{2 + (-4) + 1 + 0 + (-1) + 4 + 0 + 2 + 1 + 0 + 0 + 6 + (-4) + 4 + (-3)}{15} = 0,53$$

$$\begin{aligned} s_z^2 &= \frac{1}{n-1} \left(\sum_{i=1}^n (x_i - y_i)^2 - n\bar{Z}_n^2 \right) = \\ &= \frac{1}{14} \left((2 \cdot 2^2 + 2 \cdot (-4)^2 + 2 \cdot 1^2 + (-1)^2 + 2 \cdot 4^2 + 6^2 + (-3)^2) - 15 \cdot 0,53^2 \right) = \\ &= 8,27. \end{aligned}$$

Оттука, специфичната вредност на тестот изнесува:

$$\bar{t} = \frac{|0,53|}{\sqrt{8,27}} \sqrt{15} = 0,72$$

од табелата за Студентова распределба, двостран тест за $\alpha = 0,05$ и $n_1 - 1 = 14$ степени на слобода за критичната вредност имаме:

$$t_{0,05}(14) = 2,08.$$

Бидејќи $\bar{t} = 0,72 < 2,145 = t_{0,05}(14)$ немаме доволно причини да ја отфрлиме нултата хипотеза, со прифатлив ризик од грешка од типот I помала од 0,05. Значи дека немаме доволно причини за да тврдиме дека постои разлика помеѓу средната вредност на годините на сопругниците поминати во образование надвор од средното образование, со прифатлив ризик од грешка од типот I помала од 0,05.

Ако примероците се со големина $n > 30$, тогаш статистиката $\frac{\bar{Z}_n}{S_z} \sqrt{n}$ има $N(0,1)$ распределба.

Во следната табела 3.12 се претставени условите за примена на Z -тестот и постапката за тестирање хипотези за разликата меѓу математичките очекувања на две популации кога примероците се зависни и големи.

Табела 3.3.2.3. Тестирање хипотези за разликата меѓу математичките очекувања на две популации при зависни примероци со примена на Z -тестот

Услови за примена на Z -тестот:

Случајните примероци се зависни.

Примероците имаат големина $n > 30$.

Избор на обликот на нултата и алтернативната хипотеза:

$$H_0 : m_1 = m_2, \quad H_1 : m_1 \neq m_2;$$

$$H_0 : m_1 \leq m_2, \quad H_1 : m_1 > m_2;$$

$$H_0 : m_1 \geq m_2, \quad H_1 : m_1 < m_2.$$

Статистика за тестот:

1) $Z = \frac{\bar{Z}_n}{\sigma_z} \sqrt{n}$, кога е позната дисперзијата на променливата $Z = X - Y$, σ_z ;

2) $Z = \frac{\bar{Z}_n}{s_z} \sqrt{n}$, кога не е позната дисперзијата на променливата $Z = X - Y$;

Со ниво на значајност α

А) За хипотезата $H_0 : m = m_0$, наспроти алтернативната хипотеза $H_1 : m \neq m_0$:

i) пресметуваме $\bar{z} = \frac{|\bar{Z}_n|}{\sigma_z} \sqrt{n}$ за тестот 1) или $\bar{z} = \frac{|\bar{Z}_n|}{s_z} \sqrt{n}$ за тестот 2), каде што $\bar{z}_n = \frac{1}{n} \sum_{i=1}^n (x_i - y_i)$ е аритметичката средина на променливата $Z = X - Y$, а $s_z = \frac{1}{n-1} \left(\sum_{i=1}^n (x_i - y_i)^2 - n \bar{z}_n^2 \right)$ е стандардно отстапување на променливата Z ;

ii) одредуваме реален број $z_{\alpha/2}$ таков што $P_{H_0} \left\{ \bar{z} \geq z_{\alpha/2} \right\} = \frac{\alpha}{2}$, односно $\Phi_0 \left(z_{\alpha/2} \right) = 1 - \frac{\alpha}{2}$ и заклучуваме:
- ако $\bar{z} \leq z_{\alpha/2}$, не ја отфрламе хипотезата H_0 ,

- ако $\bar{z} > z_{\alpha/2}$, ја отфрламе хипотезата H_0 со веројатност на грешка од прв тип помала од α .

В) За хипотезата $H_0: m \leq m_0$, наспроти алтернативната хипотеза $H_1: m > m_0$:

i) пресметуваме $\bar{z} = \frac{\bar{z}_n}{\sigma_z} \sqrt{n}$ за тестот 1) или $\bar{z} = \frac{\bar{z}_n}{s_z} \sqrt{n}$ за тестот 2), каде што $\bar{z}_n = \frac{1}{n} \sum_{i=1}^n (x_i - y_i)$ е аритметичката средина на променливата $Z = X - Y$, а

$$s_z = \frac{1}{n-1} \left(\sum_{i=1}^n (x_i - y_i)^2 - n \bar{z}_n^2 \right)$$

е стандардно отстапување на променливата Z ;

ii) одредуваме реален број z_α таков што $P_{H_0} \{ \bar{z} \geq z_\alpha \} = 1 - \alpha$, односно $\Phi_0(z_\alpha) = 1 - \alpha$ и заклучуваме

- ако $\bar{z} \leq z_\alpha$, не ја отфрламе хипотезата H_0 ,

- ако $\bar{z} > z_\alpha$, ја отфрламе хипотезата H_0 со веројатност на грешка од прв тип помала од α .

С) за хипотезата $H_0: m \geq m_0$ наспроти алтернативната хипотеза $H_1: m < m_0$:

i) пресметуваме $\bar{z} = \frac{\bar{z}_n}{\sigma_z} \sqrt{n}$ за тестот 1) или $\bar{z} = \frac{\bar{z}_n}{s_z} \sqrt{n}$ за тестот 2), каде што $\bar{z}_n = \frac{1}{n} \sum_{i=1}^n (x_i - y_i)$ е аритметичката средина на

променливата $Z = X - Y$, додека $s_z = \frac{1}{n-1} \left(\sum_{i=1}^n (x_i - y_i)^2 - n \bar{z}_n^2 \right)$ е

стандардното отстапување на променливата;

ii) одредуваме реален број z_α таков што $P_{H_0} \{ \bar{z} \leq -z_\alpha \} = 1 - \alpha$, односно $\Phi_0(z_\alpha) = 1 - \alpha$ и заклучуваме

- ако $\bar{z} \geq -z_\alpha$, не ја отфрламе хипотезата H_0 ,

- ако $\bar{z} < -z_\alpha$, ја отфрламе хипотезата H_0 со веројатност на грешка од прв тип помала од α .

Задача 3.3.2.4. Една законодавна комисија сака да провери дали има значајна разлика во даночните приходи помеѓу предложениот нов даночен закон и постојниот даночен закон. Комисијата по случаен избор избира 100 репрезентативни даночни пријави. За тие пријави, таа го пресметува данокот $\bar{y}$ користејќи го предложениот нов даночен закон и го споредува со данокот $\bar{x}$ платен според постоечкиот закон. Комисијата потоа ја пресметува просечната разлика помеѓу овие два закона и таа изнесува -219, стандардно отстапување на истата од 725. Што може да заклучи законодавната комисија врз основа на овој примерок, со прифатлив ризик од 0,01?

Решение. Разгледуваме зависни примероци со големини $n_1 = n_2 = 100 > 30$, аритметички средина $\bar{z}_n = -219$ и стандардно отстапување $s_n = 725$ на променливата $Z = X - Y$. Она што сакаме да го провериме е дали постои значајна разлика во даночните приходи помеѓу предложениот нов даночен закон и постојниот даночен закон (алтернативна хипотеза), затоа употребуваме z -тест за следнава хипотеза:

$$H_0 : m_1 = m_2 \text{ наспроти алтернативната хипотеза } H_1 : m_1 < m_2,$$

а статистиката за тестот е $Z = \frac{\bar{Z}_n}{S_z} \sqrt{n}$.

Тогаш специфичната вредност изнесува:

$$\bar{z} = \frac{|\bar{z}_n|}{s_z} \sqrt{n} = \frac{219}{725} \sqrt{100} = 3,02;$$

а од табелата за нормална распределба за критичната вредност $z_{0,01}$ имаме:

$$\Phi_0(2,32) = 0,9898 < \Phi_0(z_{0,01}) = 1 - 0,01 = 0,99 < 0,9901 = \Phi_0(2,33)$$

$$\text{или } z_{0,01} = \frac{2,32 + 2,33}{2} = 2,325.$$

Бидејќи $\bar{z} = 3,02 > 2,325 = z_{0,01}$, нултата хипотеза ја отфрламе, со прифатлив ризик од грешка од типот I помала од 0,01. Значи дека во согласност со примерокот постојат доволно причини да сметаме дека постои статистички значајна разлика во даночните приходи помеѓу предложениот нов даночен закон и постојниот даночен закон, со прифатлив ризик од грешка од тип I помала од 0,01.

3.4. ЗАДАЧИ

Задача 3.4.1. Извршени се мерења на едно обележје на примерок со обем $n = 20$, при што е добиена оценка на σ^2 , $s^2 = 3$. Ако со стандардите е пропишано $\sigma^2 = 2$, да се провери дали серијата од која потекнува примерокот ги задоволува тие стандарди. Притоа, $\alpha = 0,05$.

Решение. Треба да се провери хипотезата $H_0: \sigma^2 = 2$; во однос на алтернативната хипотеза $H_0: \sigma^2 > 2$. Имаме:

$$\frac{ns^2}{\sigma^2} = \frac{20 \cdot 3}{2} = 30.$$

Во табелата на χ^2 за $n = 19$, $\alpha = 0,05$; добиваме дека:

$$\chi_{19;0,05}^2 = 30,144.$$

Бидејќи $30 < 30,144$, не ја отфрламе хипотезата H_0 .

Задача 3.4.2. Извршени се мерења на едно обележје на примерок со обем $n = 25$, при што е добиена оценка на σ^2 , $s^2 = 2$. Ако со стандардите е пропишано $\sigma^2 = 1$, да се провери дали серијата од која потекнува примерокот ги задоволува тие стандарди. Притоа, $\alpha = 0,05$.

Решение. Треба да се провери хипотезата $H_0: \sigma^2 = 1$; во однос на алтернативната хипотеза $H_0: \sigma^2 > 1$. Имаме:

$$\frac{ns^2}{\sigma^2} = \frac{25 \cdot 2}{1} = 50.$$

Во табелата на χ^2 за $n = 24$, $\alpha = 0,05$; добиваме дека:

$$\chi_{24;0,05}^2 = 36,415.$$

Бидејќи $50 > 36,415$, хипотезата H_0 се отфрла со веројатност на грешка од I вид $\alpha = 0,05$.

Задача 3.4.3. Извршени се 9 мерења на една големина и добиени се следниве вредности:

5,29; 5,30; 5,33; 5,31; 5,30; 5,32; 5,29; 5,32; 5,33.

Ако стандардната грешка на мерењето е $\sigma = 0,02$, да се провери грешката дека точната големина е 5,30 при $\alpha = 0,05$.

Решение. Вршме проверка на хипотезите:

$$H_0 : EX = 5,3, H_0 : EX \neq 5,3 \text{ при } \alpha = 0,05.$$

Ги пресметуваме вредностите:

$$\bar{x}_9 = \frac{2 \cdot 5,29 + 2 \cdot 5,30 + 5,31 + 2 \cdot 5,32 + 2 \cdot 5,33}{9} \approx 5,31$$

и

$$\bar{u}_9 = \frac{\bar{x}_9 - a_0}{\sigma} \sqrt{n} = \frac{5,31 - 5,30}{0,02} \sqrt{9} = 1,5.$$

Од студентовата t -распределба за $n = \infty$ и $\alpha = 0,05$ ја отчитуваме вредноста $t = 1,96$.

Бидејќи, $1,5 < 1,96$, хипотезата H_0 е точна со веројатност на грешка од I видот $\alpha = 0,05$.

Задача 3.4.4. Извршени се 10 мерења на една големина и добиени се следниве вредности:

5,3; 5,28; 5,28; 5,29; 5,3; 5,31; 5,31; 5,29; 5,3; 5,3.

Ако стандардната грешка на мерењето е $\sigma = 0,02$, да се провери грешката дека точната големина е $a_0 = 5,29$ при $\alpha = 0,05$.

Решение. Вршме проверка на хипотезите:

$$H_0 : EX = 5,3, H_0 : EX \neq 5,3 \text{ при } \alpha = 0,05.$$

Ги пресметуваме вредностите:

$$\bar{x}_9 = \frac{2 \cdot 5,28 + 2 \cdot 5,29 + 4 \cdot 5,3 + 2 \cdot 5,31}{9} \approx 5,296$$

и

$$\bar{u}_{10} = \frac{\bar{x}_{10} - a_0}{\sigma} \sqrt{n} = \frac{5,296 - 5,29}{0,02} \sqrt{10} \approx 0,95.$$

Од студентовата t -распределба за $n = \infty$ и $\alpha = 0,05$ ја отчитуваме вредноста $t = 1,96$.

Бидејќи $0,95 < 1,96$, хипотезата H_0 не се отфрла со веројатност на грешка од I вид $\alpha = 0,05$.

Задача 3.4.5. Со стандардите е пропишано одредена карактеристична вредност да има големина 60 единици. Од извршените мерења на 9 производи добиени се оценките:

$$\bar{x}_9 = 59,2 \text{ и } \bar{s}_9 = 0,5.$$

Да се провери, со веројатност на грешка од прв вид $\alpha = 0,05$; дали серијата од која се земени производите ги задоволува стандардите.

Решение. Треба да се тестира хипотезата $H_0: EX = 60$ во однос на алтернативната хипотеза $H_0: EX \neq 60$.

Во табелата на t -распределбата за $n = 9 - 1 = 8$ (се бара бројот на мерења -1) и $\alpha = 0,05$, се отчитува вредноста:

$$t_{8,0,05} = 2,306.$$

Од друга страна врз основа на оценките добиени од примерокот се пресметува вредноста на тест-статистиката:

$$\bar{t}_{n-1} = \frac{\bar{x}_n - a_0}{\bar{s}_n} \sqrt{n} \Leftrightarrow \bar{t}_8 = \frac{\bar{x}_9 - a_0}{\bar{s}_9} \sqrt{9} \Leftrightarrow \bar{t}_8 = \frac{59,2 - 60}{0,5} 3 = -\frac{0,8}{0,5} 3 = -4,8.$$

Бидејќи

$$|\bar{t}_8| = 4,8 > 2,306 = t_{8,0,05},$$

заклучуваме дека хипотезата H_0 се отфрла со веројатност на грешка од I вид $\alpha = 0,05$.

Во овој случај, податоците од примерокот покажуваат дека се работи за значајно отстапување од стандардното.

4. ПРОСТА ЛИНЕАРНА РЕГРЕСИЈА И КОРЕЛАЦИЈА. ХИ-КВАДРАТ ТЕСТ.

4.0. Проста линейна регресија и корелација

Во претходните поглавја се запознавме со методите на инференцијалната статистика во согласност кои врз основа на информациите од примерокот да се донесат заклучоци (со оценување или тестирање на хипотези) за непознатите карактеристики на популацијата. Постојат тестови на хипотези со кои се проверува зависноста меѓу две обележја, кои не беа споменати во интерес на времето предвидено за совладување на основните методи на статистичкото заклучување. Меѓутоа, таквата анализа не овозможува оценување или предвидување на вредностите на една појава врз основа на познавањето на вредностите на некоја друга појава. Статистичките методи кои тоа го овозможуваат ќе ги разгледаме во ова поглавје. Заради тоа, нашиот интерес ќе биде насочен кон **истражување на меѓусебните врски и влијанието на две појави**. Како и досега, заклучоците повторно ќе се засноваат на разгледувањето примерок. Но, сега врз основа на примерокот ќе испитуваме како варијабилноста на една појава е поврзана со варијабилноста на друга појава. Целта на нашето истражување нема да се однесува само на примерокот, туку пред сè на популацијата од која е извлечен примерокот.

4.1. Детерминистичка и случајна зависност

Врските меѓу појавите (променливите) можат да се поделат во две групи: **детерминистички и случајни**. Детерминистичката врска (се нарекува уште и *функционална* или *егзактна*) се јавува кога на една вредност на независната променлива X одговара само една, точно одредена, вредност на зависната променлива Y . Така на пример, волуменот на коцка се пресметува со формулата $P = a^2$. За која било вредност на работ на коцката a , може *егзактно* да се пресмета волуменот на коцката, со едноставна замена на нумеричките вредности на десната страна од равенството. Во општ случај, ваков вид на врска може да се прикаже со изразот $Y = f(X)$, каде што $f(X)$ е некоја функција од X .

Детерминистичката врска ретко може да се сретне во секојдневниот живот.

На општествените појави влијаат голем број на фактори и специфични непредвидливи влијанија од психолошка природа. Заради тоа не сме во можност врз основа на знаењето на поединечни вредности на независната променлива да ги предвидиме вредностите на зависната променлива. Врските кои тука се јавуваат се послаби од детерминистичките и се нарекуваат **случајна зависност**. Кај случајната зависност на една вредност на независната променлива X одговара цела низа на можни вредности на зависната променлива Y . Секоја од тие вредности на зависната променлива Y може да се земе со одредена веројатност и бидејќи нејзините исходи во поединечни ситуации не можеме со сигурност да ги предвидиме, тогаш таа е случајна променлива. Суштината на случајните врски кои ние ќе ги разгледуваме е дека меѓу поединечни вредности на независната променлива X и просечните вредности на зависната променлива Y (попрецизно очекуваните вредности) постои функционална зависност, односно, Просек на $Y = f(X)$.

Врските во кои растењето (опаѓањето) на вредностите на независната променлива X одговара на растењето (опаѓањето) на вредностите на зависната променлива Y се нарекуваат **директни (право-пропорционални)** врски. А, ако на растењето на вредноста на една променлива одговара опаѓање на вредноста на другата променлива тогаш, врската се нарекува **инверзна (обратно-пропорционална)**. Доколку, се констатира дека со промена на вредностите на едната променлива вредностите на другата променлива остануваат константни, заклучуваме дека меѓу нив не постои никаква зависност.

4.2. Цели на регресионата и корелационата анализа

При истражување на квантитативните релации меѓу две или повеќе појави во статистиката, се користат методите на регресиона и корелациона анализа. Овие методи, иако се блиско поврзани и се надополнуваат, имаат значајни разлики.

Во регресионата анализа потребно е однапред да се определи која појава ќе биде независна променлива, а која зависна. Ова се заснова на претходни теоретски или емпириски сознанија или претпоставки за природата на појавите што се анализираат. Во

корелационата анализа, при испитување на две појави, е безначајно која ќе се именува како зависна или независна променлива – резултатот останува ист. Сепак, при испитување на корелацијата меѓу три или повеќе појави, една од нив треба да се дефинира како зависна, додека другите добиваат улога на независни променливи.

Целта на регресионата анализа е да се утврди видот на врската, односно формата на зависноста меѓу две разгледувани појави, што се постигнува преку соодветен регресионен модел. Регресиониот модел е статистички модел кој преку математички формули и соодветни претпоставки ја опишува квантитативната зависност меѓу варијабилноста на анализираните појави. Поради стохастичката природа на овие врски, регресиониот модел ја покажува просечната конзистентност на варијабилноста на појавите. Сепак, регресиониот модел не е крајна цел, туку алатка за проценка и предвидување на однесувањето на зависната променлива при одредена вредност на независната променлива.

Целта на корелационата анализа е да се утврди дали има совпаѓање меѓу варијабилноста на појавите и, ако постои, колку е силна врската.

Важна напомена е дека ниту регресионата ниту корелационата анализа не можат да ја утврдат причинско-последичната врска меѓу појавите – дали едната е причина, а другата последица. Ова се утврдува со други методи на квантитативна анализа.

При испитување на релациите меѓу две променливи се користат методите на проста (линеарна или нелинеарна) регресиона и корелациона анализа. Ние ќе се ограничине на линеарните врски, бидејќи се полесни за примена и почесто користени во пракса.

4.3. Проста линеарна регресија

Регресионата анализа е една од најважните и најчесто употребувани статистички модели и има голема примена во економијата и во други општествени науки. Терминот *регресија* прв го употребил англискиот научник Францис Галтон 1885 при истражувањето на наследните карактеристики на децата во однос на родителите. Денес, зборот регресија се употребува во значително пошироко значење да покаже на статистичкиот метод кој овозможува предвидување и оценување на една појава врз основа на вредностите на друга појава (или група појави).

4. Проста линеарна регресија и корелација. Хи-квадрат тест

За проста линеарна регресија зборуваме кога разгледуваме две појави меѓу кои постои линеарна (праволиниска) врска.

За да ја објасниме природата на регресионите модели, прво ќе разгледаме еден пример:

Имено, познато е дека притисокот на воздухот зависи од надморската висина на местото на кое таа се мери и дека тој се намалува со зголемување на надморската висина (под воздужен притисок ќе го подразбираме просечниот годишен притисок на воздухот на одредено место).

Ако извршиме мерења на n места со различни висини $x_1, x_2, \dots, x_n$ ќе ги добиеме соодветните вредности на воздушниот притисок $Y_1, Y_2, \dots, Y_n$. Меѓутоа, ако овие мерења ги повториме на истите места, тогаш, воопшто земено, нема да се добијат истите вредности за густината на воздухот, туку ќе добиеме нови вредности $Y'_1, Y'_2, \dots, Y'_n$.

Затоа, променливите $Y_1, Y_2, \dots, Y_n$ со кои се означени воздушните притисоци на одредени n места можеме да ги сметаме за случајни променливи, додека $x_1, x_2, \dots, x_n$ не ги сметаме за случајни променливи, туку сметаме дека тие се определени константи.

Бидејќи Y_i зависи од x_i како од параметар, можеме да сметаме дека случајната променлива Y (воздушниот притисок) зависи од параметарот x (висината на мерното место) кој не е случаен, т. е. $Y = Y(x)$ и притоа за секој $i = 1, 2, 3, \dots, n$ важи $Y_i = Y(x_i)$.

Според тоа, простиот случаен примерок треба да се разбере така што за секој $i = 1, 2, 3, \dots, n$ на вредноста x_i , ја придружуваме случајната променлива Y_i . Значи, простиот случаен примерок е од видот $(x_i, Y_i), i = 1, 2, \dots, n$. Понатаму, во примерокот случајните променливи $Y_1, Y_2, \dots, Y_n$ примаат вредности $y_1, y_2, \dots, y_n$ соодветно, па затоа реализацијата на примерокот е $(x_i, y_i), i = 1, 2, \dots, n$.

Природно е да се постави прашањето врз основа на примерокот $(x_i, Y_i), i = 1, 2, \dots, n$ и неговата реализација $(x_i, y_i), i = 1, 2, \dots, n$ да се најде $Y = Y(x)$ и да се даде оценка за точноста на оваа зависност.

4. Проста линеарна регресија и корелација. Хи-квадрат тест

Претходниот пример, нè упатува на заклучокот дека случајната променлива Y е составена од две компоненти и тоа:

1. **функционална компонента**, односно функција $f(x)$ која зависи од аргументот x и

2. **случајна компонента**, односно случајна променлива ε која ги одразува некои други влијанија врз случајната променлива Y .

Во пракса односот меѓу случајната променлива Y , функционалната компонента $f(x)$ и стохастичката компонента ε вообичаено се задава со помош на една од функциите:

$$Y = f(x) + \varepsilon \text{ или } Y = f(x)\varepsilon.$$

Тука, ние ќе ја разгледуваме функцијата $Y = f(x) + \varepsilon$, каде што:

$$f(x) = \alpha x + \beta, \alpha, \beta \in \mathbb{R}$$

и ε е нормално распределена случајна променлива со математичко очекување 0 и дисперзија σ^2 .

Односно, моделот на проста линеарна регресија е зададен со формулата:

$$f(x) = \alpha x + \beta + \varepsilon,$$

при што ε има $N(0, \sigma^2)$ распределба. Понатаму, бидејќи $\alpha x + \beta$ не е случајна променлива од својствата на математичкото очекување и дисперзијата непосредно следува дека:

$$E(Y) = E(\alpha x + \beta + \varepsilon) = \alpha x + \beta + E(\varepsilon) = \alpha x + \beta + 0 = \alpha x + \beta$$

$$D(Y) = D(\alpha x + \beta + \varepsilon) = D(\varepsilon) = \sigma^2.$$

Нека разгледуваме случаен примерок (x_i, Y_i) , $i = 1, 2, \dots, n$ со реализација (x_i, y_i) , $i = 1, 2, \dots, n$. Тогаш важи:

$$Y_i = \alpha x_i + \beta + \varepsilon_i, \text{ за } i = 1, 2, \dots, n,$$

каде ε_i , $i = 1, 2, \dots, n$ се независни случајни променливи со распределби $N(0, \sigma^2)$. Затоа и Y_i , $i = 1, 2, \dots, n$ се независни случајни променливи кои имаат распределби од облик на $N(\alpha x_i + \beta, \sigma^2)$.

Оттука, следува дека во случај на линейна регресија определувањето на аналитички израз за случајната променлива Y се сведува на наоѓање на функционалната и стохастичката компонента на Y . Тоа значи дека врз основа на примерокот треба да се оценат параметрите α , β и σ^2 . Односно, врз основа на примерокот $(x_i, Y_i), i=1, 2, \dots, n$ треба да најдеме оценки a , b и s^2 на параметрите α , β и σ^2 , соодветно. Јасно, оценките a , b и s^2 зависат од примерокот $(x_i, Y_i), i=1, 2, \dots, n$, па затоа тие се статистики, односно случајни променливи, кои за конкретна реализација $(x_i, y_i), i=1, 2, \dots, n$ на примерокот се определени константи a' , b' и s'^2 , соодветно.

4.3.1. Точкести оценки на параметрите на проста линейна регресија и прави на линейната регресија

Нека е дадена реализацијата $(x_i, y_i), i=1, 2, \dots, n$; на примерокот $(x_i, Y_i), i=1, 2, \dots, n$. За да ги определиме оценките a и b на параметрите α и β , соодветно, врз основа на примерокот, во моделот на линейната регресија $Y = \alpha x + \beta + \varepsilon$ постапуваме на следниов начин. За секоја вредност x_i на параметарот x нека е пресметана вредноста $\hat{y}_i = \alpha x_i + \beta$. Тогаш, во точката (x_i, y_i) отстапувањето на пресметаната вредност $\hat{y}_i$ од реализираната вредност y_i е еднаква на реализираната вредност ε'_i на случајната променлива ε_i , односно:

$$\varepsilon'_i = y_i - \hat{y}_i = y_i - \alpha x_i - \beta.$$

Оттука, вкупното средноквадратно отстапување на реализираните вредности $y_i, i=1, 2, \dots, n$ од пресметаните вредности $\hat{y}_i, i=1, 2, \dots, n$ е дадено со:

$$s^2 = \frac{1}{n} \sum_i (y_i - \alpha x_i - \beta)^2 \quad (4.1)$$

Понатаму, оценките a и b на параметрите α и β ќе ги определиме така, што средноквадратното отстапување s^2 да биде

4. Проста линеарна регресија и корелација. Хи-квадрат тест

најмало, односно функцијата $s^2 = \frac{1}{n} \sum_i (y_i - \alpha x_i - \beta)^2$ да има минимум. За таа цел ги наоѓаме парцијалните изводи $\frac{\partial s^2}{\partial \alpha}$ и $\frac{\partial s^2}{\partial \beta}$ и нив ги изедначуваме на нула, со што го добиваме системот равенки:

$$\frac{\partial s^2}{\partial \alpha} = 0, \quad \frac{\partial s^2}{\partial \beta} = 0 \quad (4.2)$$

Бидејќи:

$$\begin{aligned} \frac{\partial s^2}{\partial \alpha} &= -2 \frac{1}{n} \sum_i x_i (y_i - \alpha x_i - \beta) = -2 \left(\frac{1}{n} \sum_i x_i y_i - \alpha \frac{1}{n} \sum_i x_i^2 - \beta \frac{1}{n} \sum_i x_i \right) = \\ &= -2 \left(s_{xy} + \bar{x}\bar{y} - \alpha (s_x^2 + \bar{x}^2) - \beta \bar{x} \right) \end{aligned}$$

и

$$\begin{aligned} \frac{\partial s^2}{\partial \beta} &= -2 \frac{1}{n} \sum_i (y_i - \alpha x_i - \beta) = -2 \left(\frac{1}{n} \sum_i y_i - \alpha \frac{1}{n} \sum_i x_i - \beta \right) = \\ &= -2 (\bar{y} - \alpha \bar{x} - \beta) \end{aligned}$$

каде што:

$$\bar{x} = \frac{1}{n} \sum_i x_i, \quad \bar{y} = \frac{1}{n} \sum_i y_i, \quad s_{xy} = \frac{1}{n} \sum_i x_i y_i - \bar{x}\bar{y}$$

и

$$s_x^2 = \frac{1}{n} \sum_i x_i^2 - \bar{x}^2, \quad s_y^2 = \frac{1}{n} \sum_i y_i^2 - \bar{y}^2$$

добиваме дека системот (4.2) е еквивалентен на системот:

$$\begin{cases} s_{xy} + \bar{x} \cdot \bar{y} - \alpha s_x^2 - \alpha \bar{x}^2 - \beta \bar{x} = 0 \\ \bar{y} - \alpha \bar{x} - \beta = 0 \end{cases}.$$

Решението на последниот систем е: $a' = \frac{s_{xy}}{s_x^2}$ и $b' = \bar{y} - \frac{s_{xy}}{s_x^2} \bar{x}$.

Ова значи дека за $\alpha_0 = a'$ и $\beta_0 = b'$ функцијата (4.1) достигнува минимум и тоа се бараните оценки a' и b' на параметрите α и β , соодветно. Понатаму, ако во равенката $y = \alpha x + \beta$ параметрите α и β ги замениме со најдените оценки a' и b' за равенката на линеарната регресија на примерокот наоѓаме:

$$y - \bar{y} = \frac{s_{xy}}{s_x^2}(x - \bar{x}) \quad (4.3).$$

За да ја определеме минималната вредност на средноквадратното отстапување доволно е во равенството (4.1) да ги замениме најдените оценки a' и b' на параметрите α и β . Притоа добиваме:

$$\begin{aligned} s'^2 &= \frac{1}{n} \sum_i \left(y_i - \bar{y} - \frac{s_{xy}}{s_x^2}(x_i - \bar{x}) \right)^2 = \\ &= \frac{1}{n} \sum_i \left[(y_i - \bar{y})^2 - 2 \frac{s_{xy}}{s_x^2}(x_i - \bar{x})(y_i - \bar{y}) + \frac{s_{xy}^2}{s_x^4}(x_i - \bar{x})^2 \right] = \\ &= \frac{1}{n} \sum_i (y_i - \bar{y})^2 - 2 \frac{s_{xy}}{s_x^2} \frac{1}{n} \sum_i (x_i - \bar{x})(y_i - \bar{y}) + \frac{s_{xy}^2}{s_x^4} \frac{1}{n} \sum_i (x_i - \bar{x})^2 = \\ &= s_y^2 - 2 \frac{s_{xy}^2}{s_x^2} + \frac{s_{xy}^2}{s_x^4} s_x^2 = s_y^2 - \frac{s_{xy}^2}{s_x^2} = s_y^2 (1 - r^2), \end{aligned}$$

каде што $r = \frac{s_{xy}}{s_x s_y}$. Од претходните разгледувања следи дека

$s'^2 = s_y^2 (1 - r^2)$ е оценка за дисперзијата σ^2 на стохастичката компонента ε .

Меѓутоа, a' , b' и s'^2 треба да ги разгледуваме како реализирани вредности на статистиките a , b и s^2 , соодветно, за земиот примерок (x_i, y_i) . Ако ги искористиме ознаките:

$$\bar{Y} = \frac{1}{n} \sum_i Y_i, \quad s_{xY} = \frac{1}{n} \sum_i x_i Y_i - \bar{x} \bar{Y}, \quad s_Y^2 = \frac{1}{n} \sum_i Y_i^2 - \bar{Y}^2 \quad \text{и} \quad r = \frac{s_{xY}}{s_x s_Y}.$$

тогаш за статистиките a , b и s^2 добиваме:

$$a = \frac{s_{xY}}{s_x^2}, \quad b = \bar{Y} - \frac{s_{xY}}{s_x^2} \bar{x} \quad \text{и} \quad s^2 = s_Y^2 (1 - r^2).$$

соодветно. Притоа, за статистиките a и b важи:

$$E(a) = E\left(\frac{s_{xY}}{s_x^2}\right) = \frac{1}{s_x^2} E(s_{xY}) = \frac{1}{s_x^2} E\left(\frac{1}{n} \sum_i x_i Y_i - \bar{x} \cdot \bar{Y}\right) =$$

$$\begin{aligned} \frac{1}{s_x^2} E\left(\frac{1}{n} \sum_i x_i Y_i - \bar{x} \frac{1}{n} \sum_i Y_i\right) &= \frac{1}{s_x^2} E\left(\frac{1}{n} \sum_i (x_i - \bar{x}) Y_i\right) = \\ \frac{1}{s_x^2} \frac{1}{n} \sum_i (x_i - \bar{x}) E(Y_i) &= \frac{1}{s_x^2} \frac{1}{n} \sum_i (x_i - \bar{x}) (\alpha x_i + \beta) = \\ \frac{\alpha}{s_x^2} \left(\frac{1}{n} \sum_i (x_i^2 - \bar{x}^2) + \beta \left(\frac{1}{n} \sum_i x_i - \bar{x}\right)\right) &= \frac{\alpha s_x^2}{s_x^2} + \frac{0 \cdot \beta}{s_x^2} = \alpha; \end{aligned}$$

и

$$\begin{aligned} E(b) &= E(\bar{Y} - \alpha \bar{x}) = E(\bar{Y}) - \bar{x} E(\alpha) = \\ E\left(\frac{1}{n} \sum_i Y_i\right) - \alpha \bar{x} &= \frac{1}{n} \sum_i (\alpha x_i + \beta) - \alpha \bar{x} = \beta; \end{aligned}$$

што значи дека статистиките a и b се центрирани оценки за параметрите α и β . Но, статистиката s^2 не е центрирана оценка за дисперзијата σ^2 , меѓутоа статистиката $\bar{s}^2 = \frac{n}{n-2} s^2$ е центрирана оценка за дисперзијата σ^2 .

Од претходната дискусија можеме да кажеме дека равенката (моделот) на проста линеарна регресија е даден со изразот (4.3).

Притоа, $b = \bar{y} - \frac{S_{xy}}{S_x^2} \bar{x}$ се нарекува константен член, а $a = \frac{S_{xy}}{S_x^2}$ се

нарекува регресионен коефициент.

Регресиониот коефициент ја дава промената на регресионата вредност за единечно зголемување на независната променлива X , односно означува за колку во просек се менува вредноста на зависната променлива Y за единечна промена на независната променлива X .

Понатаму, вкупното средноквадратно отстапување на податоците од правата на регресија (4.3) изнесува $s^2 = s_y^2 (1 - r^2)$ и него го нарекуваме **дисперзија на линеарната регресија**, а квадратниот корен $s = s_y \sqrt{1 - r^2}$ го нарекуваме **стандардна девијација на линеарната регресија**.

Притоа, стандардната девијација на линеарната регресија покажува колкаво е просечното отстапување на емпириските вредности $y_i, i = 1, 2, \dots, n$ на зависната променлива Y од

регресиските вредности $\hat{y}_i, i=1,2,\dots,n$ во мерни единици на зависната променлива Y .

Количникот $\frac{100s}{\bar{y}} = V_s$ кој го нарекуваме **коэффициент на варијација на линейната регресија**, го дава релативното просечно отстапување (во %) на емпиriskите вредности од регресионите вредности.

Специфичен показател на репрезентативноста на линейната регресија е **коэффициентот на детерминација** c_d , кој е еднаков на количникот од средноквадратното отстапување на регресионите вредности $\hat{y}_i, i=1,2,\dots,n$; од аритметичката средина $\bar{y}$ и дисперзијата s_y^2 . Според тоа,

$$c_d = \frac{\frac{1}{n} \sum_{i=1} (\hat{y}_i - \bar{y})^2}{s_y^2} = \frac{\frac{1}{n} \sum_{i=1} \left(\frac{s_{xy}}{s_x^2} x_i + \bar{y} - \frac{s_{xy}}{s_x^2} \bar{x} - \bar{y} \right)^2}{s_y^2} = \frac{\frac{1}{n} \frac{s_{xy}^2}{s_x^4} \sum_{i=1} (x_i - \bar{x})^2}{s_y^2} =$$

$$\frac{\frac{s_{xy}^2}{s_x^4 s_y^2} s_x^2}{\frac{s_{xy}^2}{s_x^2 s_y^2}} = r^2.$$

Коефициентот на детерминација ни дава информација колкав дел од вкупната варијабилност на зависната променлива во однос на независната променливата е објаснета (детерминирана). Вредностите што ги прима коефициентот на детерминација се меѓу 0 и 1. Притоа, ако коефициентот на детерминација е близу 1, тогаш велиме дека регресионата права прилично добро ги репрезентира емпиriskите податоци.

Забелешка. Со слична постапка можеме да ја најдеме и другата права на регресија:

$$x - \bar{x} = \frac{s_{xy}}{s_y^2} (y - \bar{y}).$$

која ја дава зависноста на променливата X од променливата Y .

Задача 4.3.1: Врз основа на податоците за обележјата X и Y , дадени во следнава табела да се определат двете прави на регресија и средно-квадратното отстапување од нив.

4. Проста линейна регресија и корелација. Хи-квадрат тест

X	0	0	0	2	2	4	4	4	6	6	6	6	6	8	8	8	8	10	10	10
Y	1	1	2	2	4	3	4	5	4	5	6	7	5	6	7	8	7	8	9	7

Решение. За поедноставно пресметување ќе ја формираме табелата 4.3.1.

Од добиените податоци од табелата имаме:

$$\bar{x} = \frac{108}{20} = 5,4, \quad \bar{y} = \frac{101}{20} = 5,05, \quad s_x^2 = \frac{792}{20} - 5,4^2 = 10,44,$$

$$s_y^2 = \frac{619}{20} - 5,05^2 = 5,4475 \text{ и } s_{xy} = \frac{686}{20} - 5,4 \cdot 5,05 = 7,03.$$

Според тоа, коефициентот на корелација на примерокот е:
Правата на регресија на Y по X е дадена со:

$$y - 5,05 = \frac{7,03}{10,44}(x - 5,4),$$

а правата на регресија на X по Y со:

$$x - 5,4 = \frac{7,03}{5,4475}(y - 5,05).$$

$y_j \setminus x_i$	0	2	4	6	8	10	f_j	$f_j y_j$	$f_j y_j^2$	$\sum_i f_{ij} x_i$	$y_i \sum_i f_{ij} x_i$
1	2						2	2	2	0	0
2	1	1					2	4	8	2	4
3			1				1	3	9	4	12
4		1	1	1			3	12	48	12	48
5			1	2			3	15	75	16	80
6				1	1		2	12	72	14	84
7				1	2	1	4	28	196	32	224
8					1	1	2	16	128	18	144
9						1	1	9	81	10	90
f_i	3	2	3	5	4	3	Σ	101	619	108	686
			$f_i x_i$	0	4	12	30	32	30	108	
			$f_i x_i^2$	0	8	48	180	256	300	792	
			$\sum_j f_{ij} y_j$	4	6	12	27	28	24	101	

$x_i \sum_j f_{ij} y_j$	0	12	48	162	224	240	686
-------------------------	---	----	----	-----	-----	-----	-----

Табела 4.3.1.1.

Средноквадратното отстапување на регресијата на Y по X е:

$$s_1 = \sqrt{5,4475(1 - 0,9322^2)} = 0,84478,$$

а средноквадратното отстапување на регресијата на X по Y е:

$$s_2 = \sqrt{10,44(1 - 0,9322^2)} = 1,1695.$$

Коефициентот на варијација за регресијата на Y по X е:

$$V_{s_1} = \frac{100s_1}{\bar{y}} = \frac{100 \cdot 0,84478}{5,05} \% = 16,73\%,$$

а коефициентот на варијација за регресијата на Y по X изнесува:

$$V_{s_2} = \frac{100s_2}{\bar{x}} = \frac{100 \cdot 1,1695}{5,4} \% = 21,66\%.$$

И во двата случаја коефициентот на детерминација изнесува:

$$c_d r = \frac{s_{xy}^2}{s_x^2 s_y^2} = r^2 = 0,9322^2 = 0,869.$$

Оттука, заклучуваме дека 86,9 % од вкупната варијабилност на зависната променлива во однос на независната променливата е објасната (детерминирана). Остатокот, 13,1 % од вкупната варијабилност не е објаснета со регресионата линија, односно е под влијание на неидентификувани фактори. Бидејќи коефициентот на детерминација е близу 1, тогаш регресионата права прилично добро ги репрезентира емпириските податоци.

4.4. Проста линеарна корелација

Корелационата анализа е во многу блиска врска со регресионата анализа. Целта на корелационата анализа е да се утврди дали меѓу варијабилноста на разгледуваните појави постои квантитативно сложување (корелациона врска) и ако постои од кој степен е таа. Ако се разгледуваат две појави, тогаш зборуваме за проста корелација. Со проста корелација може да се испитува дали

меѓу појавите постои линеарна или криволиниска корелација. Ние ќе ја разгледаме само простата линеарна корелација.

4.4.1. Коефициент на проста линеарна корелација на примерок

За разлика од регресионата анализа, кај простата линеарна корелација двете разгледувани појави се разгледуваат како случајни променливи. Тука не се прави разлика меѓу зависните и независните променливи - и двете појави добиваат еднаков статус. Значи, целосно е сеедно која појава ќе ја означиме со X , а која со Y , бидејќи се добиваат идентични резултати.

Имено, ако за обележјата X и Y се разгледува следниов примерок:

$$(X_1, Y_1), (X_2, Y_2), (X_3, Y_3), \dots, (X_n, Y_n),$$

чија реализација изнесува $(x_1, y_1), (x_2, y_2), (x_3, y_3), \dots, (x_n, y_n)$, тогаш бројот:

$$s_{xy} = \frac{1}{n} \sum_{i=1}^n (x_i - \bar{x})(y_i - \bar{y})$$

се нарекува *коваријанса на примерокот*.

За поефективно пресметување на коваријансата на примерокот, се трансформира претходната формула и се добива следнава формула за пресметување на коваријансата на примерокот:

$$\begin{aligned} s_{xy} &= \frac{1}{n} \sum_{i=1}^n (x_i - \bar{x})(y_i - \bar{y}) = \frac{1}{n} \sum_{i=1}^n (x_i y_i - \bar{x} y_i - x_i \bar{y} + \bar{x} \cdot \bar{y}) = \\ &= \frac{1}{n} \sum_{i=1}^n x_i y_i - \frac{1}{n} \sum_{i=1}^n \bar{x} y_i - \frac{1}{n} \sum_{i=1}^n x_i \bar{y} + \frac{1}{n} \sum_{i=1}^n \bar{x} \cdot \bar{y} = \frac{1}{n} \sum_{i=1}^n x_i y_i - \bar{x} \cdot \bar{y} = \\ &= \frac{1}{n} \sum_{i=1}^n x_i y_i - \left(\frac{1}{n} \sum_{i=1}^n x_i \right) \left(\frac{1}{n} \sum_{i=1}^n y_i \right) \end{aligned}$$

Како мера на јачината на проста линеарна корелациона врска во примерокот се користи релативна мера, која се нарекува *коефициент на проста линеарна корелација*. Овој коефициент го покажува *степенот на праволиниското квантитативно сложување на две појави*. Се означува со r и се пресметува со формулата:

$$r = \frac{S_{xy}}{S_x S_y} ,$$

каде што S_{xy} е коваријансата на примерокот и:

$$\begin{aligned} s_x^2 &= \frac{1}{n} \sum_{i=1}^n (x_i - \bar{x})^2 = \frac{1}{n} \sum_{i=1}^n x_i^2 - \bar{x}^2 = \\ &= \frac{1}{n} \sum_{i=1}^n x_i^2 - \left(\frac{1}{n} \sum_{i=1}^n x_i \right)^2 , \\ s_y^2 &= \frac{1}{n} \sum_{i=1}^n (y_i - \bar{y})^2 = \frac{1}{n} \sum_{i=1}^n y_i^2 - \bar{y}^2 = \\ &= \frac{1}{n} \sum_{i=1}^n y_i^2 - \left(\frac{1}{n} \sum_{i=1}^n y_i \right)^2 . \end{aligned}$$

Коефициентот на проста линеарна корелација како релативна мерка прима вредности од -1 до +1. Притоа, ако прима позитивни вредности, корелацијата меѓу појавите е *директна или право-пропорционална*. Во случај кога $-1 \leq r < 0$ врската е инверзна или *обратно-пропорционална*.

Ако $r = \pm 1$, тогаш меѓу разгледуваните појави постои функционална врска (сите емпириски точки се наоѓаат на една права) и велиме дека е совршена корелација. Значи, што повеќе коефициентот на проста линеарна корелација по апсолутна вредност е поблиску до 1, сè повеќе е посилна корелационата врска меѓу појавите. Во спротивно, што повеќе коефициентот на проста линеарна корелација е поблиску до 0, корелационата врска меѓу појавите е послаба.

Во екстреман случај, кога коефициентот на проста линеарна корелација е еднаков на нула, се заклучува дека нема линеарна корелација меѓу појавите. Но, притоа погрешно е да се донесе заклучок дека не постои никаква врска меѓу појавите. Во тој случај меѓу појавите можеби постои некој облик на криволиниско сложување или воопшто не постои никаква квантитативна врска; но врз основа на нултиот коефициент на корелација не сме во состојба да заклучиме што е вистина.

4. Проста линеарна регресија и корелација. Хи-квадрат тест

Задача 4.4.1.1. Дадени се следниве податоци за две обележја X и Y :

x_i	2	4	4	6	6	8	8	10	10	10
y_j	3	4	6	6	7	7	8	9	10	10

Пресметајте го коефициентот на корелација r .

Решение. Прво ги добиваме пресметувањата дадени во следнава табела:

											Сума
x_i	2	4	4	6	6	8	8	10	10	12	70
y_j	3	4	6	6	7	7	8	9	10	10	70
x_i^2	4	16	16	36	36	64	64	100	100	144	580
y_i^2	9	16	36	36	49	49	64	81	100	100	540
$x_i y_i$	6	16	24	36	42	56	64	90	100	120	554

Табела 4.4.1.1.

Од податоците во претходната табела последователно наоѓаме

$$\begin{aligned}\bar{x} &= \frac{1}{n} \sum_{i=1}^n x_i = \frac{70}{10} = 7, \quad \bar{y} = \frac{1}{n} \sum_{i=1}^n y_i = \frac{70}{10} = 7, \\ s_x^2 &= \frac{1}{n} \sum_{i=1}^n x_i^2 - \left(\frac{1}{n} \sum_{i=1}^n x_i \right)^2 = \frac{580}{10} - 7^2 = 58 - 49 = 9, \\ s_y^2 &= \frac{1}{n} \sum_{i=1}^n y_i^2 - \left(\frac{1}{n} \sum_{i=1}^n y_i \right)^2 = \frac{540}{10} - 7^2 = 54 - 49 = 5\end{aligned}$$

и

$$s_{xy} = \frac{1}{n} \sum_{i=1}^n x_i y_i - \left(\frac{1}{n} \sum_{i=1}^n x_i \right) \left(\frac{1}{n} \sum_{i=1}^n y_i \right) = \frac{554}{10} - 7 \cdot 7 = 55,4 - 49 = 6,4.$$

Ако замениме во формулата за коефициентот на корелација на примерокот:

$$r = \frac{s_{xy}}{s_x s_y} = \frac{6,4}{\sqrt{9} \sqrt{5}} = 0,95408.$$

Да се вратиме на формулата за пресметувањето на коефициентот на корелација на примерокот r . Нека подредениот пар (x_i, y_j) има честота f_{ij} , при што важи:

$$\sum_i \sum_j f_{ij} = n$$

Од формулата за коваријансата на примерокот имаме:

$$\begin{aligned} s_{xy} &= \frac{1}{n} \sum_i \sum_j f_{ij} (x_i - \bar{x})(y_j - \bar{y}) = \\ &= \frac{1}{n} \sum_i \sum_j f_{ij} x_i y_j - \frac{1}{n} \bar{y} \sum_i \sum_j f_{ij} x_i - \frac{1}{n} \bar{x} \sum_i \sum_j f_{ij} y_j + \frac{1}{n} \bar{x} \cdot \bar{y} \sum_i \sum_j f_{ij} = \\ &= \frac{1}{n} \sum_i \sum_j f_{ij} x_i y_j - \frac{1}{n} \bar{y} \sum_i f_i x_i - \frac{1}{n} \bar{x} \sum_j f_j y_j + \bar{x} \cdot \bar{y} = \\ &= \frac{1}{n} \sum_i \sum_j f_{ij} x_i y_j - \left(\frac{1}{n} \sum_i f_i x_i \right) \left(\frac{1}{n} \sum_j f_j y_j \right) \end{aligned}$$

и како:

$$\begin{aligned} s_x^2 &= \frac{1}{n} \sum_i f_i (x_i - \bar{x})^2 = \frac{1}{n} \sum_i f_i x_i^2 - \bar{x}^2 = \\ &= \frac{1}{n} \sum_i f_i x_i^2 - \left(\frac{1}{n} \sum_i f_i x_i \right)^2 \end{aligned}$$

и

$$\begin{aligned} s_y^2 &= \frac{1}{n} \sum_j f_j (y_j - \bar{y})^2 = \frac{1}{n} \sum_j f_j y_j^2 - \bar{y}^2 = \\ &= \frac{1}{n} \sum_j f_j y_j^2 - \left(\frac{1}{n} \sum_j f_j y_j \right)^2. \end{aligned}$$

тогаш формулата за пресметувањето на коефициентот на корелација на примерокот се трансформира во следниов облик:

$$r = \frac{s_{xy}}{s_x s_y} = \frac{\frac{1}{n} \sum_i \sum_j f_{ij} x_i y_j - \left(\frac{1}{n} \sum_i f_i x_i \right) \left(\frac{1}{n} \sum_j f_j y_j \right)}{\sqrt{\frac{1}{n} \sum_i f_i x_i^2 - \left(\frac{1}{n} \sum_i f_i x_i \right)^2} \cdot \sqrt{\frac{1}{n} \sum_j f_j y_j^2 - \left(\frac{1}{n} \sum_j f_j y_j \right)^2}}.$$

За пресметување на коефициентот на корелација на примерокот последната формула е посоодветна од претходната, што може да се види од следниов пример.

Задача 4.4.1.2. Дадени се следниве податоци за две обележја X и Y :

(4,3) (6,1) (2,4) (3,3) (6,2) (4,3) (4,2) (3,2) (5,3) (5,2)
 (4,4) (3,4) (4,3) (4,2) (2,3) (3,4) (5,2) (4,2) (6,1) (2,5)
 (2,5) (3,3) (4,3) (5,1) (2,4) (6,3) (3,4) (5,1) (4,3) (4,4)
 (5,2) (3,3) (4,3) (5,1) (2,4) (6,3) (3,4) (5,1) (4,3) (5,2)
 (2,4) (5,1) (4,2) (3,2) (5,3) (6,2) (4,3) (3,3) (5,2) (3,3)
 (4,3) (6,1) (3,4) (5,2) (4,3) (2,4) (3,3) (6,2) (4,4) (4,3)

при што во секоја заграда првиот број се однесува на обележјето X , а вториот на обележјето Y .

Пресметајте го коефициентот на корелација r на обележјата X и Y .

Решение. Прво ќе ја најдеме распределбата на честотите, за што ја составуваме табелата:

y_j/x_i	2	3	4	5	6
1					
2					
3			 		
4					
5					

Табела 4.4.1.2.1.

Од горната табела се гледа колку пати одреден подреден пар се јавува во множеството податоци. На пример, подредениот пар (4,3) се јавува 9 пати, а подредениот пар (3,3) се јавува 5 пати.

Потоа, составуваме табела во која наместо цртите ги ставаме соодветните честоти f_{ij} на подредените парови (x_i, y_i) , и неа ја дополнуваме со соодветни колони во кои ги вршиме потребните пресметувања. Ја добиваме табелата 4.4.1.2.

Да забележиме дека редот $f_i x_i$ го добиваме со множење на вредностите x_i со соодветните честоти x_i и збирот на овој ред $\sum_i f_i x_i = 199$. Ако добиените вредности од редот $\sum_i f_i x_i$ ги

4. Проста линеарна регресија и корелација. Хи-квадрат тест

помножине со соодветните вредности x_i , го добиваме редот $\sum_i f_i x_i^2$.

Збирот на овој ред е $\sum_i f_i x_i^2 = 869$. Аналогно, наоѓаме:

$$\sum_j f_j y_j = 138 \text{ и } \sum_j f_j y_j^2 = 434.$$

Редот $\sum_j f_{ij} y_j$ го добиваме како збир на производите на честотите f_{ij} во една колона и соодветните вредности y_j . Така, на пример:

$$29 = 1 \cdot 3 + 4 \cdot 4 + 2 \cdot 5, \quad 35 = 2 \cdot 2 + 3 \cdot 5 + 4 \cdot 4 \dots$$

Ако редот $\sum_j f_{ij} y_j$ го помножине со соодветните вредности x_i , го добиваме редот $x_i \sum_j f_{ij} y_j$, чиј збир е:

$$\sum_i \sum_j x_i f_{ij} y_j = \sum_i x_i \sum_j f_{ij} y_j = 502.$$

$y_j \backslash x_i$	2	3	4	5	6	f_j	$f_j y_j$	$f_j y_j^2$	$\sum_i f_{ij} x_i$	$y_i \sum_i f_{ij} x_i$
1				3	3	6	6	6	33	33
2		2	4	5	3	14	28	56	65	130
3	1	5	9	2	1	18	54	162	69	207
4	4	4	2			10	40	160	28	112
5	2					2	10	50	4	20
f_i	7	11	15	10	7	Σ	138	434	199	502
$f_i x_i$	14	33	60	50	42	199				
$f_i x_i^2$	28	99	240	250	252	869				
$\sum_j f_{ij} y_j$	29	35	43	19	12	138				
$x_i \sum_j f_{ij} y_j$	58	105	172	95	72	502				

Табела 4.4.1.2.2.

Конечно, ако земеме предвид дека:

$$50 = n = \sum_i \sum_j f_{ij} ,$$

за коефициентот на корелација добиваме:

$$r = \frac{\frac{1}{n} \sum_i \sum_j f_{ij} x_i y_j - \left(\frac{1}{n} \sum_i f_i x_i \right) \left(\frac{1}{n} \sum_j f_j y_j \right)}{\sqrt{\frac{1}{n} \sum_i f_i x_i^2 - \left(\frac{1}{n} \sum_i f_i x_i \right)^2} \cdot \sqrt{\frac{1}{n} \sum_j f_j y_j^2 - \left(\frac{1}{n} \sum_j f_j y_j \right)^2}} =$$
$$\frac{\frac{502}{50} - \frac{199}{50} \frac{138}{50}}{\sqrt{\frac{869}{50} - \left(\frac{199}{50} \right)^2} \sqrt{\frac{434}{50} - \left(\frac{138}{50} \right)^2}} =$$
$$\approx -0,74 .$$

Како што забележуваме, апсолутната вредност на коефициентот на корелација е доста голема, па затоа можеме да заклучиме дека меѓу обележјата X и Y постои доста силна линейна корелациона врска. Бидејќи коефициентот на корелација е негативен, заклучуваме дека со растењето на вредностите на обележјето X опаѓаат вредностите на обележјето Y и обратно, односно линейната корелација е обратно пропорционална.

4.4.2. Тестирање хипотези за коефициентот на проста линейна корелација

Бидејќи X_i и Y_i $i=1,2,\dots,n$ се случајни променливи, заклучуваме дека и коефициентот на корелација на популацијата кој се означува со ρ е случајна променлива. Се покажува дека функцијата на веројатностите на случајната променлива ρ зависи само од параметрите r и n , но таа е доста сложена. Затоа вредноста на коефициентот на корелација на примерокот нема да се совпадне со вредноста на коефициентот на корелација на дводимензионалната случајна променлива од која потекнува примерокот.

4. Проста линеарна регресија и корелација. Хи-квадрат тест

Коефициентот на корелација на популацијата го покажува интензитетот на линеарната врска меѓу двете разгледувани променливи во популацијата. За точно пресметување на неговата вредност, потребни се сите податоци од популацијата. Но, бидејќи во пракса најчесто се работи со примерок, реалната вредност на популацискиот корелационен коефициент останува непозната. Затоа се поставува прашањето како врз основа на коефициентот на корелација на примерокот може да се донесе сигурен заклучок за присуството на корелација во популацијата.

Може да се докаже дека, ако популацијата од која е земен примерокот има нормална распределба, тогаш коефициентот на корелација на примерокот r ја дава максимално веродостојната оценка на коефициентот на корелација ρ на случајните променливи X и Y .

Во следнава табела 4.4.2.1 се претставени условите за примена на t -тестот и постапката за тестирање хипотези за коефициентот на корелација.

Табела 4.4.2.1. Тестирање на хипотези за коефициентот на корелација со примена на t - тестот

<i>Услови за примена на t - тестот:</i>
Случајните променливи имаат нормална распределба. Коефициентот на корелација: $\rho(X, Y) = 0$.
Избор на обликот на нултата и алтернативната хипотеза: $H_0 : \rho(X, Y) = 0, \quad H_1 : \rho(X, Y) \neq 0$.
Статистика за тестот: $t = \frac{r\sqrt{n-2}}{\sqrt{1-r^2}}$ <i>Со ниво на значајност α.</i>
Пресметуваме: $\bar{t} = \frac{ r \sqrt{n-2}}{\sqrt{1-r^2}}$, каде што r е кофициент на корелација на примерокот. Читаеме вредност $t_\alpha(n-2)$ од таблицата за Студентова распределба двостран тест и заклучуваме: - ако $\bar{t} \leq t_\alpha(k)$, не ја отфрламе хипотезата H_0 ; - ако $\bar{t} > t_\alpha(k)$, ја отфрламе хипотезата H_0 со веројатност на

грешка од прв тип помала од α .

Пример 4.4.2.2. Врз основа на примерок со големина $n = 30$, пресметан е коефициент на корелација $r_{xy} = 0,48$. Проверете ја хипотезата дека коефициентот на корелација на обележјата X и Y е еднаков на нула, при ниво на значајност $\alpha = 0,01$.

Решение. Во таблицата за t -распределбата наоѓаме:

$$t_{0,01}(30-2) = 2,763;$$

а од примерокот пресметуваме:

$$\bar{t} = \frac{0,48}{\sqrt{1-0,48^2}} \sqrt{30-2} = 2,8961.$$

Бидејќи:

$$\bar{t} = 2,8961 > 2,763 = t_{0,01}(28),$$

со веројатност на грешка од прв вид помала од $\alpha = 0,01$, хипотезата H_0 ја отфрламе.

Нека претпоставиме дека коефициентот на корелација на дводимензионалната распределба е различен од нула. Во овој случај претходната статистика нема t -распределба. Меѓутоа, за $\rho \neq 0$ статистиката:

$$Z = \frac{1}{2} \ln \frac{1+r}{1-r},$$

при зголемување на големината на примерокот многу брзо се приближува кон нормална распределба со параметри:

$$m_z = \frac{1}{2} \ln \frac{1+\rho}{1-\rho} + \frac{\rho}{2(n-1)} \text{ и } \sigma_z^2 = \frac{1}{n-3}.$$

Тогаш, статистиката $u = \frac{Z - m_z}{\sigma_z}$ има $N(0,1)$ распределба.

Во следната табела 4.4.2.3 се претставени условите за примена на Z -тестот и постапката за тестирање хипотези за коефициентот на корелација.

Табела 4.4.2.3. Тестирање хипотези за коефициентот на корелација со примена на Z - тестот

<i>Услови за примена на Z - тестот:</i>
Случајните променливи имаат нормална распределба. Коефициентот на корелација $\rho(X, Y) = \rho_1 \neq 0$. Примерокот има големина $n > 30$.
Избор на обликот на нултата и алтернативната хипотеза: $H_0 : \rho(X, Y) = \rho_0 \neq 0, \quad H_1 : \rho(X, Y) = \rho_1 \neq \rho_0$ каде што ρ_0 е одредена нумеричка вредност.
Статистика за тестот: $u = \frac{Z - m_z}{\sigma_z}$ Со ниво на значајност α .
i) Пресметуваме $\bar{u} = \frac{ z - m_z }{\sigma_z}$, каде што: $z = \frac{1}{2} \ln \frac{1+r}{1-r}$, r е коефициент на корелација на примерокот, $m_z = \frac{1}{2} \ln \frac{1+\rho}{1-\rho} + \frac{\rho}{2(n-1)} \text{ и } \sigma_z^2 = \frac{1}{n-3}.$ ii) одредуваме реален број u_α таков што $P_{H_0} \{ \bar{u} \geq u_{1-\alpha} \} = \alpha$, односно $\Phi_0(u_\alpha) = 1 - \frac{\alpha}{2}$ и заклучуваме - ако $\bar{u} \leq u_\alpha$, не ја отфрламе хипотезата H_0, - ако $\bar{u} > u_\alpha$, ја отфрламе хипотезата H_0 со веројатност на грешка од прв тип помала од α.

Пример 4.4.2.4. Врз основа на примерок со големина $n = 30$, пресметан е коефициент на корелација $r = 0,45$. Проверете ја хипотезата дека коефициентот на корелација на обележјата X и Y е еднаков на $\rho_0 = 0,50$, при ниво на значајност $\alpha = 0,01$.

Решение. Од дадените податоци пресметуваме:

4. Проста линейна регресија и корелација. Хи-квадрат тест

$$z = \frac{1}{2} \ln \frac{1+r}{1-r} = \frac{1}{2} \ln \frac{1+0,45}{1-0,45} = 0,484700278,$$

$$\sigma_z = \frac{1}{n-3} = 0,037037037$$

и

$$m_z = \frac{1}{2} \ln \frac{1+\rho}{1-\rho} + \frac{\rho}{2(n-1)} = \frac{1}{2} \ln \frac{1+0,5}{1-0,5} + \frac{0,5}{2(30-1)} =$$
$$0,549306144 + 0,008620689 = 0,557926833.$$

Оттука:

$$\bar{u} = \left| \frac{z - m_z}{\sigma_z} \right| = \left| \frac{0,484700278 - 0,557926833}{0,037037033} \right| =$$
$$|-1,977157404| = 1,977157404$$

Од таблицата на нормалната распределба за нивото на значајност $\alpha = 0,01$, наоѓаме $u_{0,01} = 3,29$ и како:

$$u_{0,01} = 3,29 > 1,977157404 = \bar{u}$$

заклучуваме дека хипотезата H_0 не ја отфрламе, т. е. $r = 0,45$ суштински не се разликува од претпоставениот коефициент на корелација $\rho_0 = 0,50$.

4.5. χ^2 -тест на независност

Хи-квадрат тестот за независност се користи за да се утврди дали постои статистички значајна релација меѓу две категориски променливи. Основната идеја е да се спореди распределбата на набљудуваните податоци со она што би се очекувало ако двете променливи се независни.

Главните причини за употреба на Хи-квадрат тестот за независност се:

- Тестирање на независност: Овој тест овозможува да се утврди дали променливите се независни или имаат некаква поврзаност. На пример, можеме да тестираме дали има врска меѓу полот и навиките за пушење, односно дали едната променлива (пол) влијае на другата (пушење).

4. Проста линеарна регресија и корелација. Хи-квадрат тест

- Едноставна анализа на категориски податоци: Тестот е особено погоден за категориски податоци (номинални или редни податоци) кои не може лесно да се анализираат со други методи како корелација или регресија.
- Објективност: Хи-квадрат тестот користи математички пристап за споредување на набљудуваните и очекуваните броеви, обезбедувајќи објективен начин за донесување заклучок за поврзаноста.
- Флексибилност: Тестот може да се користи за различни големини на табели (на пример, 2×2 , 3×4 , итн.), што го прави погоден за широк спектар на истражувања и податочни структури.

Со помош на овој тест, ако се утврди дека релацијата е статистички значајна, тоа значи дека постои голема веројатност за една категорија да зависи од друга категорија, што може да има значајни импликации во разни истражувачки полиња.

Претпоставете дека анализираме две категориски променливи во една популација. Првата променлива има r можни исходи, а втората s исходи. Можеме да ги сумираме податоците со табела, т.н. контингентната табела која има r редови и s колони.

Контингентната табела ги прикажува честотите на секој пар во секоја комбинација на категориите на двете променливи. Секое поле (вкупно $r \times s$) претставува една комбинација на категории на двете променливи.

Притоа со овој тест ги тестираме хипотезите во следната форма:

- H_0 : Двете променливи не се поврзани.

- H_1 : Двете променливи се поврзани.

Две различни експериментални ситуации водат до контингентни табели:

1. Ако имаме две популации кои ги набљудуваме, при што и двете имаат одредено обележје како категориска променлива. Во овој случај, нултата хипотеза е изјава за хомогеност меѓу двете популации.

2. Ако имаме една популација која ја набљудуваме; и сме заинтересирани да ја провериме поврзаноста меѓу две категориски променливи, односно две обележја на популацијата. Во овој случај, нултата хипотеза е изјава за независност меѓу двете променливи.

4. Проста линейна регресија и корелација. Хи-квадрат тест

За доволно големи примероци, истиот тест е применлив за двете ситуации. Во продолжение ќе ги разгледаме чекорите за тестирање на поврзаноста меѓу две променливи.

Забелешка. За $i=1$ до r (редови) и $j=1$ до c (колони), означуваме:

- R_i = вкупниот број набљудувања/фреквенција во i -от ред.
- C_j = вкупниот број набљудувања/фреквенција во j -тата колона.
- O_{ij} = набљудуваната бројка/фреквенција во полето во i -от ред и j -тата колона.
- E_{ij} = очекуваната бројка во полето во i -от ред и j -тата колона, ако променливите се независни, односно ако е точна нултата хипотеза. Овие вредности се пресметуваат со формулата:

$$E_{ij} = \frac{R_i \cdot C_j}{n}, i=1, \dots, r \text{ и } j=1, \dots, c;$$

Статистиката на тестот ја мери разликата помеѓу набљудуваните фреквенции и очекуваните фреквенции, под претпоставка за независност и се пресметува со формулата:

$$\chi^2 = \sum \frac{(O_{ij} - E_{ij})^2}{E_{ij}}.$$

Ако χ^2 -статистиката е голема, тоа значи дека набљудуваните фреквенции не се блиску до фреквенциите што би ги очекувале доколку двете променливи се независни. Затоа, „големи“ вредности на χ^2 -статистиката според распределбата на хи-квадрат покажуваат докази против нултата хипотеза и ја поддржуваат алтернативната хипотеза.

p -вредноста на хи-квадрат тестот е веројатноста χ^2 -статистиката да биде толку голема или поголема од добиената вредност, ако H_0 е вистинита. Исто така, ако H_0 е вистинита, χ^2 -статистиката има хи-квадрат распределба со $(r-1) \times (c-1)$ степени на слобода.

Затоа, p -вредноста за хи-квадрат тест секогаш е областа десно од статистиката на тестот под графикот на густината на χ^2 -распределбата со $(r-1) \times (c-1)$ степени на слобода, односно $P(X > \chi^2) = p$, каде што X има хи-квадрат распределба со $(r-1) \times (c-1)$ степени на слобода. За да ја најдеме оваа веројатност, треба да користиме хи-квадрат распределба со $(r-1) \times (c-1)$ степени на слобода. Со помош на некој статистички софтвер, може да се добие p -вредноста од резултатите и доколку таа е помала од нивото на значајност α нултата хипотеза се отфрла. Користејќи ја табелата за χ^2 -распределбата (бидејќи обично нема да може да се најде точната p -вредност во табелата), може да се одреди опсег за p -вредноста.

Постои и друг начин во согласност со кој се одлучува дали нултата хипотеза се отфрла или не постојат доволно статистички значајни докази дека може да се отфрли нултата хипотеза, со утврдување на критичната вредност на тестот $\chi^2_{(r-1) \times (c-1), \alpha}$ која се чита од табелата за χ^2 -распределбата за $(r-1) \times (c-1)$ степени на слобода и веројатност p еднаква на нивото на значајност на тестот α , односно $p = \alpha$. Ако вредноста на тест статистиката е поголема од критичната вредност тогаш се отфрла нултата хипотеза, односно се заклучува дека двете променливи се зависни со веројатност на грешка од прв тип помала од α .

Пред да се примени тестот мора да се проверат следниве услови:

- Сите очекувани вредности треба да се поголеми од 1.
- Најмалку 80 % од очекуваните вредности треба да се поголеми од 5.

Пример 4.5.1. Илија сака да дознае дали географската област во која пораснал студентот е поврзана со тоа дали студентот конзумира алкохол. Подолу се резултатите што ги добил од случаен примерок на студенти од Факултетот за информатички технологии:

	Не консумира алкохол	Консумира алкохол	Вкупно
Голем град	21	65	86
Рурална средина	11	130	141
Мало градче	18	198	216
Субурбана средина	37	345	382
Вкупно	87	738	825

Табела 4.5.1.1.

Ниво на значајност $\alpha = 0,001$.

Решение. Ја тестираме следнава хипотеза:

H_0 : Нема релација помеѓу географската област во која пораснал студентот и тоа дали студентот консумира алкохол.

H_1 : Постои релација помеѓу географската област во која пораснал студентот и тоа дали студентот консумира алкохол.

За да ги провериме условите, потребно е да ги пресметаме очекуваните фреквенции за секоја ќелија.

$$E_{11} = \frac{R_1 \cdot C_1}{n} = \frac{86 \cdot 87}{825} = 9,07$$

$$E_{12} = \frac{R_1 \cdot C_2}{n} = \frac{86 \cdot 738}{825} = 76,93$$

$$E_{21} = \frac{R_2 \cdot C_1}{n} = \frac{141 \cdot 87}{825} = 14,87$$

$$E_{22} = \frac{R_2 \cdot C_2}{n} = \frac{141 \cdot 738}{825} = 126,13$$

$$E_{31} = \frac{R_3 \cdot C_1}{n} = \frac{216 \cdot 87}{825} = 22,78$$

$$E_{32} = \frac{R_3 \cdot C_2}{n} = \frac{216 \cdot 738}{825} = 193,22$$

$$E_{41} = \frac{R_4 \cdot C_1}{n} = \frac{382 \cdot 87}{825} = 40,28$$

$$E_{42} = \frac{R_4 \cdot C_2}{n} = \frac{382 \cdot 738}{825} = 341,72$$

Очигледно условите за да се примени тестот се исполнети и ја формираме следнава табела:

	Не консумира алкохол	Консумира алкохол	Вкупно
Голем град	$O_{11} = 21$ $E_{11} = 9,07$	$O_{12} = 65$ $E_{12} = 76,93$	86
Рурална средина	$O_{21} = 11$ $E_{21} = 14,87$	$O_{22} = 130$ $E_{22} = 126,13$	141
Мало гратче	$O_{31} = 18$ $E_{31} = 22,78$	$O_{32} = 198$ $E_{32} = 193,22$	216
Субурбана средина	$O_{41} = 37$ $E_{41} = 40,28$	$O_{42} = 345$ $E_{42} = 341,72$	382
Вкупно	87	738	825

Табела 4.5.1.2.

Тогаш, χ^2 - статистиката за примерокот изнесува:

$$\begin{aligned}\chi^2 = \sum_{i=1}^4 \sum_{j=1}^2 \frac{(O_{ij} - E_{ij})^2}{E_{ij}} = & \frac{(21 - 9,07)^2}{9,07} + \frac{(65 - 76,93)^2}{76,93} + \frac{(11 - 14,87)^2}{14,87} \\ & + \frac{(130 - 126,13)^2}{126,13} + \frac{(18 - 22,78)^2}{22,78} + \frac{(198 - 193,22)^2}{193,22} + \frac{(37 - 40,28)^2}{40,28} \\ & + \frac{(345 - 341,72)^2}{341,72} = 20,091\end{aligned}$$

Критичната вредност на тестот која се чита од табелата за χ^2 - распределбата за $(r-1) \cdot (c-1) = (2-1) \cdot (4-1) = 3$ степени на слобода и нивото на значајност на тестот $\alpha = 0,001$ е:

$$\chi^2_{3;0,001} = 16,27$$

Бидејќи $\chi^2 = 20,091 > 16,27 = \chi^2_{3;0,001}$ ја отфрламе нултата хипотеза, односно заклучуваме дека постои релација помеѓу географската област во која пораснал студентот и тоа дали студентот консумира алкохол со ниво на грешка од 0,001.

1. ЛИТЕРАТУРА

[1] Зоран Мисајлески, **Векторска и линеарна алгебра**, Е-издание на УКИМ, Скопје, 2018;

[2] Зоран Мисајлески, **Решени задачи по диференцијално и интегрално сметање I (на функции од една променлива)**, Е-издание на УКИМ, Скопје, 2019;

[3] Зоран Мисајлески, **Решени задачи по векторска и линеарна алгебра**, Е-издание на УКИМ, Скопје, 2019;

[4] Зоран Мисајлески, **Решени задачи по диференцијално и интегрално сметање I (на функции од една променлива)**, Е-издание на УКИМ, Скопје, 2023;

[5] Зоран Мисајлески, **Решени задачи по диференцијално и интегрално сметање II (на функции од повеќе променливи)**, Е-издание на УКИМ, Скопје, 2024;

[6] Зоран Мисајлески, Б. Андоновиќ, Т. Димовски, **Збирка задачи по математика I**, за студентите на Технолошко-металургискиот факултет, Е-издание на УКИМ, Скопје, 2014;

[7] Даниел Велинов, **Теорија на веројатност за инженери**, Е-издание на УКИМ, Скопје, 2021;

[8] Душан Чакмаков, **Теорија на веројатност за инженери**, Е-издание на УКИМ, Скопје, 2015;

[9] Димитри П. Берцекас, Џон Н. Цициклис, **Вовед во веројатност**, Арс Ламина 2012;

[10] Пол Њуболд, Вилијам Л. Карлсон, Бети М. Торн, **Статистика за бизнис и економија**, Магор, Скопје, 2010;

[11] Џон А. Рајс, **Математичка статистика и анализа на податоци**, Арс Ламина, 2014;

[12] Тонко Тонков, **Теория на вероятностите и математическа статистика**, Издателска фирма „Тонко тонков“, София, 2002;

[13] Vepa Čuljak, **Verojatnost i statistika**, Građevinski fakultet, Sveučilište u Zagrebu, radni materijali, 2011;

[14] Бранко Трпеновски, **Елементарен увод во теоријата на вбројатноста**, УКИМ, Скопје, 1969 (Студентски збор, Скопје, 1982, Просветно дело, Скопје, 1994);

[15] Магдалена Георгиева, **Веројатност и нејзина примена во теоријата на случајни грешки** (Прв дел), Математичка школа - 1982, Скопје, 1982;

[16] М. Георгиева, **Теорија на веројатност со елементи на статистика** (Втор дел), Математичка школа -1982, Скопје, 1982;

[17] Магдалена Георгиева, **Случајни грешки** (трет дел), Математичка школа - 1982, Скопје, 1982;

[18] Магдалена Георгиева, Катерина Чундева, **Веројатност и статистика** за III година на математичко – информатичката струка, Просветно дело, Скопје, 1988;

[19] Магдалена Георгиева, Радојко Секуловски, Катерина Чундева, **Практична математика**, за IV година природно – математичка струка, Просветно дело, Скопје, 1992;

[20] Верица Бакева, **Веројатност**, е-издание на УКИМ, Скопје 2016;

[21] Анета Бучковска, Катерина Х. Санева, Сања Атанасова, **Вовед во веројатност за инженери**, ФЕИТ, Скопје, 2018;

[22] Ристо Малчески, **Статистика за бизнис**, Скопје, 2005;

[23] Ристо Малчески, **Вовед во теоријата на веројатност**, Скопје, 2006;

[24] Никола Речкоски, **Веројатност**, Институт за истражување на туризмот, Охрид, 1998;

[25] Катерина Х. Санева, Сања Атанасова, Анета Бучковска, **Збирка решени задачи од веројатност**, Скопје, 2016;

[26] Ivana Kostić Kovačević, **Verovatnoća i statistika** sa zbirkom zadataka, Univerzitet Singidunum, Beograd, 2022;

[27] Jovan Mališić, **Zbirka zadataka iz teorije verovatnoće** sa primenama, Građevinska knjiga, Beograd, 1982;

[28] Zoran Clišić, Predrag Peruničić, **Zbirka rešenih zadataka iz verovatnoće i matematičke statistike**, Naučna knjiga, Beograd, 1989;

[29] Испитни задачи зададени на испитите на Градежниот факултет на предметите: Веројатност и статистика (1+1), (2+2), (3+2), (3+3); Одбрани поглавја од математиката (4+4) и Одбрани делови од математиката (2+2).

[30] Светозар Вукадиновић, Ј. Поповић, **Вероватноћа и статистика**, Универзитет Београду – Саобраћајни факултет, Београд, 2019;

[31] Зоран Пуцановић, Матеја Кнежевић, Марко Пешовић, Линеарна алгебра, аналитичка геометрија, елементи вероватноће и статистике, **Збирка решених задатака**, Академска мисао, Београд 2022;

[32] Milan Merkle, **Verovatnoća i statistika** za inženere i studente tehnike, Академска мисао, Београд 2021;

[33] Vesna Jevremović, **Verovatnoća i statistika**, Matematički fakultet, Beograd, 2014;

[34] Morris H. DeGroot, Mark J. Schervish, **Probability and Statistics, Fourth Edition**, Pearson Education, Inc., 2012;

[35] Murray R. Spiegel, John Schiller, R. Alu Srinivasa, **Probability and Statistics, Fourth Edition**, Mc Graw Hill, 2013.

[36] <https://medium.com/@ruiminpan62/the-broken-stick-problem-c70d19423397>, 5.11.2024.4.34AM

[37] Капка Димитрова, Пламен Паскалев, **Методическо ръководство** за решавање на задачи по висша математика, том 3, Архимед 2000, ЕОД-Софија 2008

[38] Венера Сотирова Димова-Нанчева, Александър Михайлов Витанов, Георги Иванов Караджов, Иван Михайлов Михов, Васил Борисов Попов, Стефанка Стефанова Тодорова, **Методическо ръководство** за решавање на задачи по висша математика, том 5, Техника, София, 1975.

[39] William Feller, **An Introduction to Probability Theory and Its Applications, Third Edition**, John Wiley & Sons, Inc, 1968.

[40] Williams, D., **Probability with Martingales**, Cambridge University Press, 1991.

[41] Cochran, W. G., **The distribution of quadratic forms in a normal system, with applications to the analysis of covariance**,

Mathematical Proceedings of the Cambridge Philosophical Society. 30 (2): 178–191, 1934.

[42] Fisher R. A., **Applications of 'Student's' distribution**, Metron. 5: 90–104, 1925.

2. СОДРЖИНА

I Дел: Веројатност

Предговор	3
1.1-3. Комбинаторика	5
1.0. Комбинаторика	5
1.1. Варијации	5
1.2. Пермутации	9
1.3. Комбинации	12
2.1-5. Простор на веројатност	18
2.1. Настани	18
2.2. Статистичка веројатност	25
2.3. Класична веројатност	26
2.4. Геометриска веројатност	31
2.5. Аксиоматска дефиниција на веројатноста	40
3.1-4. Условна веројатност. Независни настани. Тотална веројатност. Бејесови формули.	48
3.1. Условна веројатност	48
3.2. Независни настани	53
3.3. Тотална веројатност	56
3.4. Бејесови формули	59
4.1-5. Серија независни експерименти (Бернулиева шема)	62
4.1. Бернулиева шема	62
4.2. Најверојатен број	65
4.3-5. Формули за приближно пресметување на веројатностите во Бернулиевата шема	67
4.3. Локална теорема на Моавр-лаплас	68
4.4. Интегрална теорема на Моавр-лаплас	69
4.5. Теорема на Пуасон	73
5.1-7. Случајни променливи	76
5.1. Случајни променливи	76
5.1. Случајни променливи од дискретен тип	82
5.2. Случајни променливи од апсолутно непрекинат тип	85

6.1-3. Математичко очекување и дисперзија	92
6.1. Математичко очекување	92
6.2. Дисперзија	97
6.3. Задачи со случајни променливи	100
7.1-2. Примери на распределби	107
7.1. Распределби од дискретен тип	107
7.1.1. Индикатор на настан	107
7.1.2. Биномна распределба	108
7.1.3. Рамномерна распределба	109
7.1.4. Геометриска распределба	110
7.1.5. Пуасонова распределба	111
7.2. Распределби од апсолутно-непрекинат тип	112
7.2.1. Рамномерна распределба	113
7.2.2. Експоненцијална распределба	114
7.2.3. Гама распределба	115
7.2.4. Нормална (Гаусова) распределба	116
7.2.5. Кошиева распределба	119
7.2.6. χ^2 -распределба	120
7.2.7. Студентова распределба	123
8.1-7. Дводимензионални случајни променливи	128
8.1. Дводимензионални случајни променливи	128
8.2. Дискретни дводимензионални случајни променливи	130
8.3. Непрекинати дводимензионални случајни променливи	132
8.4. Маргинални распределби	135
8.5. Условна распределба	140
8.6. Независност на случајни променливи	144
8.7. Коваријанса и коефициент на корелација	150
9.1-2. Функции од случајни променливи	153
9.1. Функции од една случајна променлива	153
9.2. Функции од две случајни променливи	161
10.1-2. Закон за големи броеви. Централна гранична теорема	167
10.1. Закон за големи броеви	167
10.2. Централна гранична теорема	172

II Дел: Статистика

1.1-5. Статистика	178
1.1. Популација и примерок	165
1.2-5. Дескриптивна статистика	180
1.2. Претставување на статистички податоци	180
1.3. Бројни карактеристики на примерокот	187
1.4. Параметри кои го мерат расејување на случајната променлива околу центарот на растурање	193
1.5. Нумерички карактеристики на аритметичките средини на примерокот	196
1.6. Задачи со популација и примерок	197
2.1-4. Статистички оценки	207
2.1. Оценки и нивни карактеристики	207
2.2. Метод на максимална веродостојност	209
2.3. Метод на моменти	216
2.4. Оценување на математичкото очекување на популација	220
2.4.1. Интервал на доверба за математичко очекување μ , кога е позната дисперзијата на популацијата σ^2	221
2.4.2. Интервал на доверба за математичкото очекување μ , кога е непозната дисперзијата на популацијата	224
3.1-4. Тестирање хипотези	231
3.1. Метод за тестирање статистички хипотези	231
3.1.1. Нулта и алтернативна хипотеза	232
3.1.2. Избор на статистика на тестот и одредување на нејзината распределба	233
3.2. Тестирање хипотези врз основа на еден примерок	237
3.2.1. Тестови за математичко очекување	237
3.2.2. Тестови за непозната веројатност	251
3.3. Тестирање на хипотези врз основа на два примерока	256
3.3.1. Тестови за еднаквост на математичките очекувања при независни примероци	257
3.3.2. Тестови за еднаквост на математичките очекувања при зависни примероци	274
3.4. Задачи	282

4.1-5. Проста линеарна регресија и корелација. Хи квадрат тест	285
4.0. Проста линеарна регресија и корелација	285
4.1. Детерминистичка и случајна зависност	285
4.2. Цели на регресионата и корелационата анализа	286
4.3. Проста линеарна регресија	
4.3.1. Точкести оценки на параметрите на проста линеарна регресија и прави на линеарната регресија	290
4.4. Проста линеарна корелација	295
4.4.1. Коефициент на проста линеарна корелација на примерок	296
4.4.2. Тестирање хипотези за коефициентот на простата линеарна корелација	301
4.5. Хи на квадрат тест	305

Литература. Содржина

1. Литература	313
2. Содржина	317

Ниту еден дел од оваа публикација не смее да биде репродуциран на кој било начин без претходна писмена согласност на авторот.

Е-издание: <https://ukim.edu.mk/e-book/00161>