Богдан Гилев

Записки по вероятности и статистика

Учебник

София, 2025
ISBN

Учебникът е съобразен с програмите на бакалавърските курсове по „Приложна математика” и „Статистика” към ТФ на ВТУ „Т.Каблешков” и докторантския курс по “Статистика” на ВТУ „Т.Каблешков”. Също така може да бъде ползван за обучението по „Висша математика” във всички техническите вузове в България. Извън посочените по-горе програми в учебника са включени и няколко свързващи въпроса, които служат за последователно излагане на материала и осигуряване на логическа взаимовръзка на изложения материала от самото начало до самия край на учебника.

Авторът е дългогодишен преподавател в ТУ-София и ВТУ „Т.Каблешков”.

Рецензент: доц. д-р Г. Георгиев

Съдържание

ГЛАВА 1
ТЕОРИЯ НА ВЕРОЯТНОСТИТЕ

ТЕМА 1
Опит и случайни събития. Класическа и геометрична вероятност. Сигма алгебра и аксиоматична вероятност. Условна вероятност и независими събития.

1. Опит и случайни събития.

Теория на вероятностите е математическа теория. Както при всяка математическа теория, така и тук, основните понятия се дефинират интуитивно. В случая основните понятия са опит и елементарно събитие.

О1: Опит - осъществяването на определен комплекс от условия, който се характеризира с това, че: всички изходи са известни, при провеждане на опита се сбъдва само един от възможните изходи и всеки опит може да се повтаря многократно.

О2: Елементарни събития ω - са всички възможни взаимно изключващи се изходи (събития), реализиращи се при провеждането на опита.

На базата на вече въведените основни понятия се въвеждат и други понятия. Като това може да стане на база на познати математически теории, примерно математическа логика. Тук обаче следващите понятия са въведени на базата на добре познатата теория на множествата.

О3: Множеството Ω от всички възможни елементарни събития (изходи) ω се нарича пространство от всички елементарни събития.

О4: Всяко подмножеството A на Ω се нарича случайно събитие, т.е. $A \subset \Omega$.

О5: Ако A съдържа всички елементи на пространството Ω, то се нарича сигурното събитие и отново се бележи с Ω.

Забележка1: Очевидно това създава дублиране на означението Ω, понеже с този символ може да се означава: както пространство от всички елементарни събития така и сигурното събитие. Това обаче е стандартна практика, понеже в контекста на конкретен текст, винаги може да се разбере за кое от двете понятия става дума.

О6: Ако A не съдържа никой от елементи на пространството Ω, то се нарича невъзможното събитие и се бележи с Φ.

О7: Ако всички елементи на събитието A са елементи и на събитието B, то казваме, че събитието A влече събитието B, т.е. $A \subset B$.

О8: Събитието A или B $/A \cup B/$ се нарича събитието, чиито елементи принадлежат на поне едно от двете събития A или B.

О9: Събитието A и B $/A \cap B/$ се нарича събитието, чиито елементи принадлежат и на двете събития A и B.

О10: Противоположното събитие $\overline{A}$ на събитието A се състои от всички елементи на Ω, които не са елементи на A.

2. Класическа вероятност.

О11: При класическата вероятност: всички елементарни събития (изходи) са равно вероятни и броят им е крайно число.

О12: При горните ограничения вероятността $P(A)$ на събитието A се пресмята така

$$P(A)=\frac{m}{n},$$

където: m -броя на благоприятните изходи за A и n -броя на всички изходи.

Свойства: (следват директно от дефиницията)

а) $P(\Omega)=1$

б) $P(\Phi)=0$

г) $0\leq P(A)\leq 1$

Т1: (за събирането)

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

Доказателство:

Означаваме с

n -броя на всички изходи:

m -броя на благоприятните изходи за A

l -броя на благоприятните изходи за B

k -броя на благоприятните изходи за $A\cap B$.

Тогава

$$P(A\cup B)=\frac{m+l-k}{n}=\frac{m}{n}+\frac{l}{n}-\frac{k}{n}=P(A)+P(B)+P(A\cap B).$$

3. Геометрична вероятност.

Геометрична вероятност се прилага, когато на всяко елементарно събитие от Ω може да се съпостави еднозначно обратимо(биективно) една точка: от реалната ос, от равнината или от пространството. Очевидно това позволява елементарните изходи да се безброй много. Що се отнася до изискването елементарните изходи да се равновероятни, то се запазва.

За определеност ще приемем, че всяко елементарно събитие се изобразява биективно в равнинна точка. Нека тогава:

- пространство от всички елементарни събития Ω се изобразява в равнинното множество G;

- множеството от всички елементарни изходи, благоприятни за събитието A, се изобразява в равнинното множество g.

Тогава вероятността $P(A)$ на събитието A се пресмята така

$$P(A)=\frac{S(g)}{S(G)},$$

където $S(g)$ и $S(G)$ са лицата на равнините множества g и G.

Забележка2: Може да се каже, че в горната дефиниция сме използвали теория на мярката, която е неразделна част от теорията на множествата. Това е така, понеже лицето е основната мярка за равнинни геометрични множества. По същия начин дължината и обемът са основните мерки за едномерни и тримерни геометрични множества. В този дух мярката лице от горната дефиниция винаги може да се замени с мярката дължина или мярката обем.

4. Сигма алгебра и аксиоматична вероятност.

О13: Множеството F от всички възможни подмножества на събитието Ω се нарича алгебра, ако:

1. $\Omega\in F$

2. Ако $A \in F$, то $\overline{A} \in F$

3. Ако $A \in F$ и $B \in F$, то $A \cup B \in F$

О14: Ако към горните свойства на множеството F добавим свойството

4. Ако $A_1, A_2, \ldots, A_n, \ldots \in F$, то $\bigcup_{i=1}^{\infty} A_i \in F$,

то множеството F се нарича $\sigma -$алгебра(сигма алгебра).

А1: (на Колмогоров) На всяко събитие A се съпоставя число $P(A)$, наречено вероятност.

Забележка3: Ще обърнем внимание, че тази аксиома не дава правилото, по което на A се съпоставя $P(A)$. На пръв поглед това е недостатък, но именно тази свобода дава възможност за всеки отделен случай да се избере най-подходящото правило за съпоставяне, примерно класическа вероятност или геометрична вероятност.

А2: $P(\Omega) = 1$.

А3: Ако $A_1, A_2, \ldots, A_n, \ldots$ са несъвместими (не могат да се сбъднат едновременно),то

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

О15: Наредената тройка (Ω, F, P) се нарича вероятностно пространство

5. Условна вероятност и независими събития.

а) условна вероятност – първо ще въведем това понятия за класическа вероятност и след това, ще го обобщим за аксиоматична вероятност.

О16: (за класическа вероятност)

Вероятността на всяко събитието A, при условие че събитието B се е сбъднало е

$$P(A) = \frac{r}{m},$$

където:

r-броя на благоприятните изходи за A, при условие че B се е сбъднало /това всъщност са изходите, при които се е сбъднало както A така и B/.

m-броя на всички изходи, при условие че B се е сбъднало /това всъщност са изходите, при които се е сбъднало B/.

Т2: (за умножението)

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

Доказателство: (за класическа вероятност)

Нека:

- r-броят на изходите, при които се е сбъднало както A, така и B;
- m-броят изходите при които се е сбъднало B;
- n-броят на всички изходи.

Тогава

$$P(A/B) = \frac{r}{m} = \frac{r/n}{m/n} = \frac{P(A \cap B)}{P(B)}.$$

Забележка4: Ще обърнем внимание на факта, че ако вероятностна се въвежда аксиоматично, то горната теорема не може да се докаже. В този случай равенството

$$P(A/B) = \frac{P(A \cap B)}{P(B)}$$

играе ролята на дефиниционно равенство за условна вероятност.

б) независими събития.

О17: Събитията A и B са независими, когато вероятностите за сбъдване на едното не зависи от това дали другото се е сбъднало, или не се е сбъднало, т.е. $P(A)=P(A/B)$ и $P(B)=P(B/A)$.

Т3: (за умножението)

$$A \text{ и } B \text{ са независими} \Leftrightarrow P(A\cap B)=P(A)P(B).$$

Доказателство:

Следва директно от теоремата за условна вероятност $P(A/B)=\dfrac{P(A\cap B)}{P(B)}$ и от определението за независими събития $P(A)=P(A/B)$.

Забележка5: Ще обърнем внимание на факта, че ако вероятността се въвежда аксиоматично, обикновено Определение 17 се изпуска и Теорема 3 играе ролята на дефиниция за независими събития.

О18: Събитията $A_1,A_2,\ldots,A_n$ са две по две независими, ако

$$P(A_i\cap A_j)=P(A_i)P(A_j), \text{ за } i=1,\ldots,n \text{ и } j=1,\ldots,n \text{ и } i\neq j.$$

О19: Събитията $A_1,A_2,\ldots,A_n$ са независими в съвкупност, ако за кои да е индекси

$$1\le i_1<i_2<\ldots<i_k\le n$$

е изпълнено

$$P(A_{i_1}\cap\ldots\cap A_{i_k})=P(A_{i_1})\ldots P(A_{i_k}).$$

Забележка6: Може да се даде пример (на Бернщайн), където събитията са две по две независими, но са зависими в съвкупност.

Класическа и Геометрична Вероятност (задачи)

Зад.1 Лотария има 1000 билета. От тях един печели 200 лв., 10 билета печелят 50 лв., 50 печелят по 5 лв. и 100 печелят по 1 лв. Останалите билети не печелят. Гражданинът купува 1 билет. Да се намери вероятността той да е:

а) да спечели 5 лв.;

б) да спечели най-малко 5 лв.;

в) да спечели не повече от 5 лв.;

г) да вземе някоя от възможните печалби.

Решение:

а) P=$\dfrac{50}{1000}$;

б) P=$\dfrac{50+10+1}{1000}=\dfrac{61}{1000}$;

в) P=$\dfrac{50+100}{1000}=\dfrac{150}{1000}$;

г) P=$\dfrac{100+50+10+1}{1000}=\dfrac{161}{1000}$.

Зад.2 Гражданин набира телефонен номер обаче, е забравил последните две цифри. Помни само че са различни. Да се намери вероятността гражданинът да набере правилния номер от първия опит, ако набира последните две цифри по случаен начин.

Решение:

Първи начин:

Въвеждаме събитието, A - гражданинът набира правилния номер от първия опит.

Всички възможни комбинации, които може да набере гражданинът са: 00, 01, … ,98 и 99, общо 100 на брой. Понеже набраните цифри трябва да са различни, от тях трябва да махнем комбинациите: 00, 11, 22, … 88 и 99, общо 10 на брой. Следователно броят на всички възможни изходи за събитието A е $100 - 10 = 90$.

Очевидно за A има един благоприятен изход.

Следователно

$$P(A) = \frac{1}{90}.$$

Втори начин:

Въвеждаме събитията: B_i - гражданинът набира правилно i-тата цифра от първия опит ($i = 1,2$).

Очевидно $A = B_1 \cap B_2$, тогава

$$P(A) = P(B_1)P(B_2 / B_1) = \frac{1}{10}\frac{1}{9} = \frac{1}{90}.$$

Зад.3 Три урни съдържат по 20 топки. В първата урна има 8, във втората 11, а в третата 12 бели топки. От всяка урна по случаен начин се изважда по една топка. Да се намери вероятността и трите топки да се окажат бели.

Решение:

Въвеждаме събитията: A_i - от i-тата урна се изважда бяла топка ($i = 1,2,3$)

Тогава

$$P(A_1 \cap A_2 \cap A_3) = P(A_1)P(A_2)P(A_3) = \frac{8}{20}\frac{11}{20}\frac{12}{20} = \frac{33}{250}.$$

Зад.4 Хвърлят се едновременно два зара. Да се намери вероятността на поне един от тях да се паднат 6 точки.

Решение:

Първи начин:

От схемата на фиг.1 лесно се забелязва, че броя на всички елементарни изходи е 36, а броя на благоприятните такива (отбелязани с кръстче) е 11.

Зар-1 / Зар-2	1	2	3	4	5	6
1						X
2						X
3						X
4						X
5						X
6	X	X	X	X	X	X

Фиг. 1

Следователно

$$P = \frac{11}{36}.$$

Втори начин:

Въвеждаме събитията: B_i -пада се шестица на i -тия зар ($i = 1,2$). Тогава

$$P = P(B_1 \cup B_2) = P(B_1) + P(B_2) - P(B_1)P(B_2) = \frac{1}{6} + \frac{1}{6} - \frac{1}{6}\frac{1}{6} = \frac{11}{36}.$$

Трети начин:

$$P = P(B_1 \cup B_2) = 1 - P(\overline{B}_1 \cap \overline{B}_2) = 1 - P(\overline{B}_1)P(\overline{B}_2) = 1 - \frac{5}{6}\frac{5}{6} = \frac{11}{36}.$$

Зад.5 От 12 билета, номерирани от 1 до 12, произволно 1 след друг избираме 2. Намерете вероятността:

а) първия да е четен, а втория да е нечетен;

б) и двата четни;

в) единия четен, а другия нечетен.

Решение:

Въвеждаме събитията: A_i - i -тия по ред изтеглен билет е четен ($i = 1,2$).

а) $P(A_1 \cap \overline{A}_2) = P(A_1)P(\overline{A}_2 / A_1) = \frac{6}{12}\frac{6}{11} = \frac{3}{11}$;

б) $P(A_1 \cap A_2) = P(A_1)P(A_2 / A_1) = \frac{6}{12}\frac{5}{11} = \frac{5}{22}$;

в) $P((A_1 \cap \overline{A}_2) \cup (\overline{A}_1 \cap A_2)) = P(A_1)P(\overline{A}_2 / A_1) + P(\overline{A}_1)P(A_2 / \overline{A}_1) = \frac{6}{12}\frac{6}{11} + \frac{6}{12}\frac{6}{11} = \frac{6}{11}$.

Зад.6 Прекъсването на ел. верига настъпва в случай, че се повреди поне един от трите последователно свързани елемента. Да се намери вероятността за повреда на веригата, ако вероятностите за повреда на елементите са: 0.3, 0.4 и 0.6.

Решение:

Въвеждаме събитията: A_i - i -тия елемент се поврежда. Тогава

$$P = P(A_1 \cup A_2 \cup A_3) = 1 - P(\overline{A}_1 \cap \overline{A}_2 \cap \overline{A}_3) = 1 - 0.7\ 0.6\ 0.4 = 0.832.$$

Зад.7 Подхвърляме едновременно зар и монета. Определете вероятността да се паднат 2 точки на зара или герб на монетата.

Решение:

Въвеждаме събитията: A-падат се две точки на зара и B-пада се герб на монетата. Тогава

$$P = P(A \cup B) = P(A) + P(B) - P(A \cap B) = P(A) + P(B) - P(A)P(B) = \frac{1}{6} + \frac{1}{2} - \frac{1}{6}\frac{1}{2} = \frac{7}{12}.$$

Или

$$P = P(A \cup B) = 1 - P(\overline{A} \cap \overline{B}) = 1 - P(\overline{A})P(\overline{B}) = 1 - \frac{5}{6}\frac{1}{2} = \frac{7}{12}.$$

Зад.8 Три екипа A_1, A_2 и A_3 от физкултурно дружество A се състезават с три екипа B_1, B_2 и B_3 от физкултурно дружество B. Вероятностите за спечелване на мачовете от екипите A_1, A_2 и A_3 са съответно 0.8, 0.4 и 0.4. За победа е необходимо спечелване на най-малко два от три мача(равни мачове не може да има). Намерете вероятността за успех на дружеството A. Кое физкултурно дружество има по-голяма вероятност за победа?

Решение:

Означаваме с A_i събитието, победа на отбора A_i над отбора B_i ($i = 1,2,3$).

Тогава вероятността за победа на дружеството A над дружеството B е

$$P = P((A_1 \cap A_2 \cap A_3) \cup (A_1 \cap A_2 \cap \overline{A}_3) \cup (A_1 \cap \overline{A}_2 \cap A_3) \cup (\overline{A}_1 \cap A_2 \cap A_3))$$
$$= 0.8\,0.4\,0.4 + 0.8\,0.4\,0.6 + 0.8\,0.6\,0.4 + 0.2\,0.4\,0.4 = 0.544.$$

Следователно дружеството A има по-голям шанс за победа.

Зад.8 Всеки избор на числата a и b, в интервалите $|a|<1$ и $|b|<1$ е равно вероятен. Намерете вероятността на следните събития :

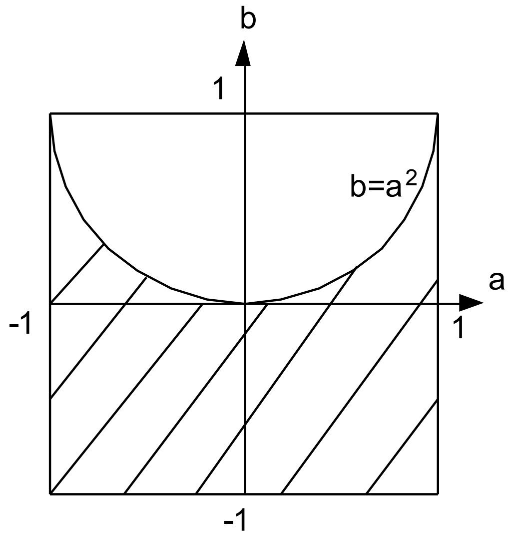
а) $A = \{\text{корените на } x^2 + 2ax + b = 0 \text{ са реални}\}$;

б) $B = \{\text{корените на } x^2 + 2ax + b = 0 \text{ са реални и пложителни}\}$.

Решение:

а) За да са корените реални е необходимо $D = a^2 - b > 0$, т.е $a^2 > b$. Като използваме и ограниченията $|a|<1$ и $|b|<1$, получаваме че

$$P(A) = \frac{\text{Лоцето на защрихованата част}}{\text{Лицето на квадрата}},$$ (фиг.2).

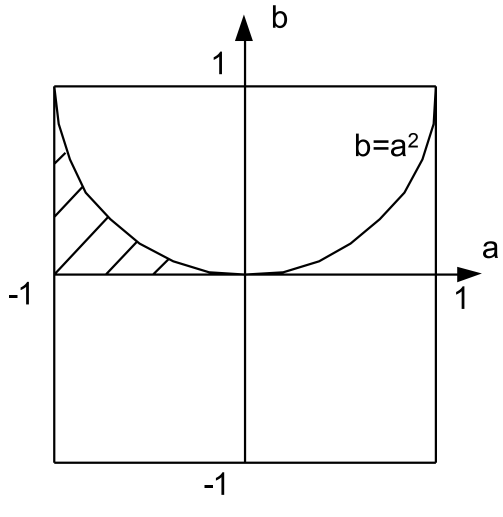


фиг.2

$$P(A)=\frac{2+\int\limits_{-1}^{1}x^2dx}{4}=\frac{2+\frac{2}{3}}{4}=\frac{4}{6}$$

б) За да са корените реални е необходимо: $D=a^2-b>0$ (т.е $a^2>b$), $x_1x_2>0$ (т.е. b>0) и $x_1+x_2>0$ (т.е. a<0). Като използваме и ограниченията $|a|<1$ и $|b|<1$, получаваме че

$$P(A)=\frac{\textit{Лоцето на защрихованата част}}{\textit{Лицето на квадрата}}$$, (фиг.3).



фиг.3

$$P(B)=\frac{\int\limits_{-1}^{0}x^2dx}{4}=\frac{\frac{1}{3}}{4}=\frac{1}{12}.$$

ТЕМА 2

Пълна група събития. Формула за пълната вероятност. Формула на Бейс. Схема на Бернули

1. Пълна група събития. Формула за пълната вероятност.

О1: Събитията $H_1, H_2, \ldots, H_n$ се наричат пълна група събития, ако

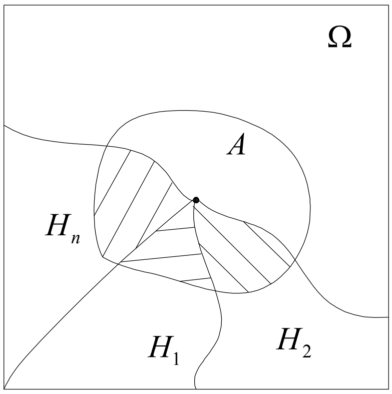
1. $H_i \cap H_j = \Phi$, за $i = 1, \ldots, n$ и $j = 1, \ldots, n$ и $i \neq j$.
2. $\bigcup_{i=1}^{n} H_i = \Omega$.

Т1: Ако събитията $H_1, H_2, \ldots, H_n$ са пълна група събития и $A \subset \Omega$, то

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

Доказателство:

От условията $H_i \cap H_j = \Phi$ и $\bigcup_{i=1}^{n} H_i = \Omega$ директно следва, че $A = \bigcup_{i=1}^{n} (A \cap H_i)$ /чертеж 1/.



Чертеж 1

Тогава

$P(A) = P\left(\bigcup_{i=1}^{n} (A \cap H_i) \right) =$ от несъвместимостта на събитията $A \cap H_1, A \cap H_2, \ldots, A \cap H_n$

$= \sum_{i=1}^{n} P(A \cap H_i) =$ от теоремата за умножение на вероятностите

$= \sum_{i=1}^{n} P(H_i) P(A / H_i)$

2. Формула на Бейс.

Т2: (на Бейс) Ако събитията $H_1, H_2, \ldots, H_n$ са пълна група събития и $A \subset \Omega$, то

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

Доказателство:

Съгласно теоремата за умножение на вероятностите

$$P(A) P(H_k / A) = P(A \cap H_k) = P(H_k) P(A / H_k)$$

$$P(H_k / A) = \frac{P(H_k)P(A/H_k)}{P(A)}.$$

В горния израз заместваме $P\left(A\right)$ с резултата от Теорема 1 и получаваме

$$P\left(H_k / A\right) = \frac{P(H_k)P(A/H_k)}{\sum_{i=1}^{n} P(H_i)P(A/H_i)}.$$

3. Схема на Бернули.

О2: Схема от n- опита или от n-повторения на един опит, се нарича схема на Бернули, ако:

1. Вероятността за сбъдване на единичното събитие A при всеки опит е една и съща, т.е. $P(A) = p = \mathrm{const}$.

2. Опитите са независими.

Задача: Да се намери вероятността при n- опита, провеждани по схемата на Бернули, единичното събитие A да се сбъдне точно k- пъти. Тази вероятност се означава с $P_{n,k}$ и се чете вероятността за k-успеха от n- опита.

Пример: Намерете вероятността при 10 хвърляния на монета "герб" да се падне точно 3 пъти.

Решение:

Първо, ще отбележим, че благоприятните за нас събития са тези при, които единичното събитие A се сбъдне точно k- пъти, а $\overline{A}$ се сбъдне точно $(n-k)$- пъти. Да разгледаме едно от тези благоприятни събития, нека това е събитието, при което A се сбъдва първите k- пъти, следователно $\overline{A}$ ще се сбъдва следващите $(n-k)$- пъти, т.е.

$$\underbrace{A \cap A \cap \ldots \cap A}_{k-пъти} \cap \underbrace{\overline{A} \cap \overline{A} \cap \ldots \cap \overline{A}}_{(n-k)-пъти}.$$

Вероятността на това събитие е

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

От друга страна, всички възможни начини на съставяне на благоприятни за нас събития са толкова, колкото са възможните начин на поставяне на A на n - места. Но това са точно комбинации от n-елемента k-клас, т.е. $C_n^k = \dfrac{n!}{k!(n-k)!}$.

От горните два резултата и от теоремата за събиране на вероятностите получаваме, че вероятността за k-успеха от n- опита е

$$P_{n,k} = C_n^k p^k (1-p)^{n-k}.$$

Забележка: Горната формула се нарича формула на Бернули**.**

Формула за пълната вероятност. Формула на Бейс (задачи)

Зад.1 Дадени са три партиди от радиолампи. Вероятностите една радио лампа да принадлежи на съответната партида са: 0.25, 0.25 и 0.5. Да се определи вероятността случайно взета лампа да работи безотказно, ако се знае че вероятностите за безотказна работа за лампа от всяка партида са съответно 0.1, 0.2 и 0.4.

Решение:

Въвеждаме събитието A-изтеглената лампа работи безотказно, както и хипотезите H_i-изтеглената лампа е от i-тата партида ($i = 1,2,3$). Тогава

$$P(A) = \sum_{i=1}^{3} P(H_i)P(A/H_i) = 0.25\ 0.1 + 0.25\ 0.2 + 0.5\ 0.4\,.$$

Зад.2 Нека са дадени трите партиди радио лампи от предната задача. Ако е известно, че изтеглената радио лампа работи безотказно, каква е вероятността тя да е била изтеглена от първата партида.

Решение:

Съгласно вече въведените означения в предната задача, търсим условната вероятност $P(H_1/A)$. Съгласно формулата на Бейс тя е

$$P(H_1/A) = \frac{P(H_1)P(A/H_1)}{P(A)} = \frac{0.25\ 0.1}{0.275} \approx 0.09\,.$$

Зад.3 Дадени са шест групи детайли със следния състав: в две групи има по 2 детайла с отклонение в плюс и по 6 детайла с отклонение в минус (от стандартните размери); в три групи има по 2 в плюс и по 8 в минус и в последната група има по 8 в плюс и 2 в минус. Да се определи вероятността от произволно избрана група да се изтегли детайл с отклонение в плюс.

Решение:

Въвеждаме събитието A-изтеглен е детайл с отклонение в плюс, както и хипотезите H_i- изтегления детайл е от i-тата група ($i = 1,2,3,4,5,6$). Тогава

$$P(A) = \sum_{i=1}^{6} P(H_i)P(A/H_i) = 2\frac{1}{6}\frac{2}{8} + 3\frac{1}{6}\frac{2}{10} + \frac{1}{6}\frac{8}{10} = \frac{19}{60}\,.$$

Зад.4 Нека са дадени шестте групи детайли от предната задача. Ако е известно, че изтегления детайл е с отклонение в плюс, каква е вероятността той да е била изтеглен от последната група детайли.

Решение:

Съгласно въведените означения в предната задача и използвайки формулата на Бейс получаваме

$$P(H_6/A) = \frac{P(H_6)P(A/H_6)}{P(A)} = \frac{\frac{1}{6}\frac{8}{10}}{\frac{19}{60}} = \frac{8}{19}\,.$$

Зад.5 В две кутии има по 12 и 10 еднотипни изделия, при това във всяка кутия има по едно дефектно изделие. Вземаме едно произволно изделие от първата кутия и го прехвърляме във втората, след това от втората кутия изтегляме едно изделие. Каква е вероятността последното изтеглено изделие да е дефектно.

Решение:

Въвеждаме събитията:

A-изтегляме дефектно изделие от втората кутия (след прехвърлянето)

H_1 (хипотеза едно)-от първата във втората кутия прехвърляме дефектно изделие

H_2 (хипотеза две)-от първата във втората кутия прехвърляме изправно изделие .
Тогава по формулата за пълната вероятност имаме

$$P(A)=\sum_{i=1}^{2}P(H_i)P(A/H_i)=\frac{1}{12}\frac{2}{11}+\frac{11}{12}\frac{1}{11}=\frac{13}{132}.$$

Зад.6 Три раздела на един доклад са били написани от три машинописки. Първата написала 10 стр., втората -8 стр. и третата - 27 стр. Първата машинописка не допуска нито една грешка на машинописна страница с вероятност 0,3. Тези вероятности за втората и третата машинописки са съответно 0.2 и 0.1. По произволен начин е взета една страница от доклада, на която се оказало че няма нито една грешка, каква е вероятността тя да е била написана от първата машинописка?

Решение:

Въвеждаме събитието A-изтеглената страница е без грешка, както и хипотезите H_i-изтеглената страница е написана от i-тата машинописка ($i=1,2,3$). Тогава съгласно формулата на Бейс имаме

$$P(H_1/A)=\frac{P(H_1)P(A/H_1)}{P(A)}=\frac{\frac{10}{45}0.3}{\frac{10}{45}0.3+\frac{8}{45}0.2+\frac{27}{45}0.1}=\frac{30}{73}.$$

Зад.6 От кутия съдържаща 3 бели и 2 черни топки, се прехвърлят две топки в друга кутия в която има 4 бели и 4 черни топки. След това от втората кутия се тегли една топка. Да се определи вероятността последната изтеглена топка да е бяла.

Решение:

Въвеждаме събитията:

A-изтегляме бяла топка от втората кутия (след прехвърлянето)

H_1 (хипотеза едно)-от първата във втората кутия прехвърляме две бели топки

H_2 (хипотеза две)-от първата във втората кутия прехвърляме бяла и черна топка

H_3 (хипотеза три)-от първата във втората кутия прехвърляме две черни топки.

За да пресметнем вероятностите на хипотезите, ще въведем следните помощни събития, B_i-i-тата по ред прехвърлена топка (от първата във втората урна) е бяла, където $i=1,2$. Тогава

$$P(H_1)=P(B_1)P(B_2/B_1)=\frac{3}{5}\frac{2}{4}=\frac{3}{10}$$

$$P(H_2)=P(B_1)P(\overline{B}_2/B_1)+P(\overline{B}_1)P(B_2/\overline{B}_1)=\frac{3}{5}\frac{2}{4}+\frac{2}{5}\frac{3}{4}=\frac{6}{10}$$

$$P(H_3)=P(\overline{B}_1)P(\overline{B}_2/\overline{B}_1)=\frac{2}{5}\frac{1}{4}=\frac{1}{10}.$$

Сега като използваме формулата за пълната вероятност получаваме

$$P(A)=\sum_{i=1}^{3}P(H_i)P(A/H_i)=\frac{3}{10}\frac{6}{10}+\frac{6}{10}\frac{5}{10}+\frac{1}{10}\frac{4}{10}=0.52.$$

Забележка: Вероятностите на хипотезите могат да се изчислят и като се използва формулата за хипергеометричната вероятност. С тази формула ще се запознаем по-късно.

Зад.7 Вероятността едно зенитно оръдие разполагащо с три снаряда да улучи самолет при първия изстрел е 0.5, при втория изстрел - 0.6 и при третия изстрел - 0.8. При три попадения свалянето на самолета е сигурно, при две попадения вероятността за свалянето му е 0.6 и при едно попадение – 0.3. Намерете вероятността след третия изстрел самолета да е свален.

Решение:

Въвеждаме събитията:

A-самолета е свален(след третия изстрел)

H_i-самолета е улучен i на брой пъти($i = 1,2,3$).

За да пресметнем вероятностите на хипотезите ще въведем следните помощни събития, B_i-самолета е улучен при i-тия поред изстрел($i = 1,2,3$). Тогава

$$P(H_1) = P((B_1 \cap \overline{B}_2 \cap \overline{B}_3) \cup (\overline{B}_1 \cap B_2 \cap \overline{B}_3) \cup (\overline{B}_1 \cap \overline{B}_2 \cap B_3))$$
$$= P(B_1)P(\overline{B}_2)P(\overline{B}_3) + P(\overline{B}_1)P(B_2)P(\overline{B}_3) + P(\overline{B}_1)P(\overline{B}_2)P(B_3)$$
$$= 0.5\ 0.4\ 0.2 + 0.5\ 0.6\ 0.2 + 0.5\ 0.4\ 0.6 = 0.26$$

$$P(H_2) = P((\overline{B}_1 \cap B_2 \cap B_3) \cup (B_1 \cap B_2 \cap \overline{B}_3) \cup (B_1 \cap \overline{B}_2 \cap B_3))$$
$$= P(\overline{B}_1)P(B_2)P(B_3) + P(B_1)P(B_2)P(\overline{B}_3) + P(B_1)P(\overline{B}_2)P(B_3)$$
$$= 0.5\ 0.6\ 0.8 + 0.5\ 0.4\ 0.8 + 0.5\ 0.6\ 0.2 = 0.46$$

$$P(H_3) = P(B_1 \cap B_2 \cap B_3) = P(B_1)P(B_2)P(B_3)$$
$$= 0.5\ 0.6\ 0.8 = 0.24$$

Сега като използваме формулата за пълната вероятност получаваме

$$P(A) = \sum_{i=1}^{3} P(H_i)P(A/H_i) = 0.26\ 0.3 + 0.46\ 0.6 + 0.24 = 0.594.$$

Схема на Бернули (задачи)

Зад.1 Вероятността за изработване на доброкачествено изделие е 0.9. Каква е вероятността от 5 произволни детайла:

а) точно 3 да бъдат доброкачествени:

б) поне едно да е доброкачествено.

Решение:

а) $P_{5,3} = C_5^3\ 0.9^3\ 0.1^2 = 0.0729$;

б) $P = \sum_{i=1}^{5} P_{5,i} = 1 - P_{5,0} = 1 - 0.1^5 = 0.99999$.

Зад.2 Правят се 6 изстрела по цел. Вероятността за попадение в целта при всеки изстрел е 0.4. Намерете вероятността да имаме:

а) точно едно попадение;

б) не по-малко от 5 попадения;

в) поне едно попадение.

Решение:

а) $P_{6,1} = C_6^1 \; 0.4 \; 0.6^5 \approx 0.187$;

б) $P_{6,5} + P_{6,6} = C_6^5 \; 0.4^5 \; 0.6 + C_6^6 \; 0.4^6 \approx 0.041$;

в) $P = \sum_{i=1}^{6} P_{6,0} = 1 - P_{6,0} = 1 - C_6^0 \; 0.6^6 \approx 0.953$.

Зад.3 От кутия в която има 20 изправни и 6 дефектни детайла, правим n на брой тегления с връщане. Да се намери най-малката стойност на числото n, при която вероятността P за изтегляне на поне един дефектен детайл е по-голяма от 0.5.

Решение:

Търсим най-малкото цяло n, което удовлетворява неравенството.

$$1 - P_{n,0} \geq \frac{1}{2} \qquad P_{n,0} \leq \frac{1}{2} \qquad \left(\frac{10}{11}\right)^n \leq \frac{1}{2} \quad n \geq \log_{10/11} 0.5 \qquad n \geq 7.27$$

$$n = 8$$

Зад.4 Установено е, че едно лице влязло в един магазин прави покупки с вероятност $p = 0.2$. В момента в магазина се намират 20 души. Да се намери най-вероятният брой души, които ще направят покупки, както и вероятността да бъдат направени този брой покупки.

Решение:

Ако m е най-вероятния брой покупки, то за това m е в сила зависимостта $P_{20,m-1} \leq P_{20,m} \geq P_{20,m+1}$. След преработване на горната зависимост получаваме

$$(20+1)p - 1 \leq m \leq (20+1)p$$

$$3.2 \leq m \leq 4.2$$

$$m = 4$$

Най-вероятният брой покупки е четири. Вероятността да се направят точно този брой покупки е

$$P_{20,4} = C_{20}^4 \; 0.2^4 \; 0.8^4 \approx 0.218\,.$$

Зад.5 Стрелец стреля по цел 3 пъти. Вероятността той да улучи поне един път е 0.992. Да се намери вероятността за улучване при един изстрел.

Решение:

$$0.992 = 1 - P_{3,0} \qquad 0.992 = 1 - (1-p)^3 \qquad 0.008 = (1-p)^3 \quad 0.2 = 1 - p$$

$$p = 0.8\,.$$

Зад.6 Колко независими изпитания трябва да се направят, в които събитието A настъпва с една и съща вероятност равна на 0.2, така че най-вероятното число появявания на това събитие да е 20.

Решение:

Нека n е този брой независими изпитания, тогава за това n е в сила зависимост аналогична на тази от зад.4, т.е.

$$(n+1)p - 1 \leq 20 \leq (n+1)p$$

$$(n+1)\, 0.2 - 1 \leq 20 \leq (n+1)\, 0.2$$

$$100 \leq n+1 \leq 105$$

$$99 \leq n \leq 104.$$

Зад.7 Стрелец стреля 49 пъти по мишена. Каква е вероятността за попадение в мишената, ако най-вероятният брой на попаденията е 30?

Решение:

Нека p е единичната вероятност за попадение в мишената, тогава за това p е в сила зависимост, аналогична на тази от зад.4 и зад.6, т.е.

$$(49+1)p-1 \leq 20 \leq (49+1)p$$

$$30 \leq 50p \leq 31$$

$$0.6 \leq p \leq 0.62$$

ТЕМА 3

Случайни величини – определения. Функция на разпределение. Дискретни и непрекъснати случайни величини. Числови характеристики– математическо очакване и дисперсия

1. Случайни величини – определения.

а) нестрого определение

О1: Числовата функция $X:\Omega \xrightarrow{X} \mathbb{R}$, заемаща стойности случайни числа, се нарича случайна величина.

б) аксиоматично определение – в този случай ще използваме дефинираното във Въпрос 1 понятие вероятностно пространство (Ω, F, P).

О2: Случайна величина е числовата функция $X(\omega)$ или $X:\Omega \xrightarrow{X} \mathbb{R}$ такава, че за всяко $x \in \mathbb{R}$ имаме $\{\omega : X(\omega) < x\} \in F$.

Забележка 1:

Условието

$$\{\omega : X(\omega) < x\} \in F$$

означава, че множеството от $\{\omega\}$, за което $X(\omega) < x$ е случайно събитие от (Ω, F, P).

Забележка 2:

С други думи, определението за случайна величина гласи, че ако $X(\omega)$ случайна величина, то множеството $\{\omega\}$, за което $X(\omega) < x$ е случайно събитие от (Ω, F, P).

Не строго казано - определението за случайна величина ни казва, че за всеки избор на $x \in \mathbb{R}$, дефиниционната стойност на $X(\omega)$, /при която е изпълнено условието $X(\omega) < x$/ е случайно събитие. Това гарантира, че винаги може да се пресметне вероятността на събитието $X(\omega) < x$, т.е. винаги може да се пресметне $P(X(\omega) < x)$.

в) независими случайни величини

О3: Случайните величини $X(\omega)$ и $Y(\omega)$ са независими, ако за всяко $x, y \in \mathbb{R}$ събитията $\{X(\omega) < x\}$ и $\{Y(\omega) < y\}$ са независими.

Забележка 3:

Като използваме теоремата/дефиницията за независими събития от Въпрос 1, можем да преформулираме горната дефиниция, така - Случайните величини $X(\omega)$ и $Y(\omega)$ са независими, ако за всяко $x, y \in \mathbb{R}$ имаме

$$P(\{X(\omega) < x\} \cap \{Y(\omega) < y\}) = P(\{X(\omega) < x\}) P(\{Y(\omega) < y\}).$$

2. Функция на разпределение.

О4: Функция на разпределение $F(x)$ /$x \in \mathbb{R}$/ на случайната величина $X(\omega)$ се задава с равенството

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

Забележка 4:

Следователно функцията на разпределение $F(x)$ на случайната величина $X(\omega)$ е вероятността на събитието "случайната величина $X(\omega)$ заема стойност по-малка от числото x".

Свойства: (следват директно от дефиницията)

а) $0 \le F(x) \le 1$;

б) $F(x)$ е монотонно растяща;

г) $\lim\limits_{x\to-\infty} F(x) = P(X(\omega) < -\infty) = 0$ и $\lim\limits_{x\to\infty} F(x) = P(X(\omega) < \infty) = 1$.

3. Дискретни и непрекъснати случайни величини.

а) дискретна

О5: Случайната величина $X(\omega)$ е дискретна, ако стойностите, които тя заема, са крайно или изброимо множество.

В такъв случай, най-удобният начин за задаване на такава величина е чрез двойката

$$\left(x_i, p_i = P(X = x_i)\right),$$

където x_i пробягва всички стойности, които случайната величина заема.

Тази двойка най-често се записва в табличен вид

X	x_1	x_2	$\dots$	x_n	$\dots$
P	p_1	p_2	$\dots$	p_n	$\dots$

и се нарича ред на разпределение.

Ако случайната величина е зададена с ред на разпределение , то нейната функция на разпределение се пресмята по формулата

$$F(x) = \sum_{i:\, x_i < x} p_i \,.$$

б) непрекъсната

О5: Случайната величина $X(\omega)$ е непрекъсната с функция на разпределение $F(x)$, ако съществува функция $f(x) \ge 0$ такава, че

$$F(x) = \int_{-\infty}^{x} f(x)dx \text{, за всяко } x \in \mathbb{R}.$$

Функцията $f(x)$ се нарича плътност на разпределение на $X(\omega)$.

Свойства:

а) $\int_{-\infty}^{+\infty} f(x)dx = 1$ - следват от горната дефиницията и третото свойство на $F(x)$.

б) $P(a \le X < b) = P(X < b) - P(X < a) == \int_{-\infty}^{b} f(x)dx - \int_{-\infty}^{a} f(x)dx = \int_{a}^{b} f(x)dx$

4. Числови характеристики –математическо очакване и дисперсия.

а) математическо очакване

О6: Математическо очакване EX на дискретната случайна величина X се нарича сумата

$$EX = \sum_i x_i \; p_i \,.$$

Забележка 3:

С други думи, математическото очакване на една дискретна случайна величина е сумата от стойностите, които тя заема по вероятностите, с които тези стойности се заемат.

О7: Математическо очакване EX на непрекъсната случайната величина X се нарича интеграла

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

Забележка 4:

В горното определение се предполага, че $EX < +\infty$, т.е. EX е крайно число.

б) свойства на математическото очакване

Свойство 1: $E(c) = c$, за $c = \text{const}$

Доказателство:

Нека $X = c = \text{const}$, следователно $P(X = c) = 1$, следователно от определението за математическо очакване на дискретна случайна величина имаме $EX = C.P(X = \text{const}) = C$.

Свойство 2: $E(cX) = cE(X)$, за $c = \text{const}$

Доказателство: (за дискретна случайна величина)
$E(cX) = \sum_i (cx_i)p_i = c\sum_i x_i p_i = cE(X)$.

Свойство 3: $E(X+Y) = E(X) + E(Y)$.

Доказателство: (за дискретна случайна величина)
Означаваме с $p_{ij} = P\left(\{X = x_i\} \cap \{Y = y_j\}\right)$, тогава

$$E(X+Y) = \sum_i \left(\sum_j (x_i + y_j) p_{ij} \right) = \sum_i x_i \left(\sum_j p_{ij} \right) + \sum_j y_j \left(\sum_i p_{ij} \right) =$$

$$= \sum_i x_i \left(\sum_j P\left(\{X = x_i\} \cap \{Y = y_j\}\right) \right) + \sum_j y_j \left(\sum_i P\left(\{X = x_i\} \cap \{Y = y_j\}\right) \right) =$$

$$= \sum_i x_i\, P\left(\{X = x_i\}\right) + \sum_j y_j P\left(\{Y = y_j\}\right) = \sum_i x_i p_i + \sum_j y_j q_j = E(X) + E(Y)$$

Свойство 4: Ако X и Y са независими случайни величини, то $E(X.Y) = E(X).E(Y)$.

Доказателство: (за дискретна случайна величина)

Ако X и Y са независими случайни величини, то $p_{ij} = P\left(\{X = x_i\} \cap \{Y = y_j\}\right) = P\left(\{X = x_i\}\right) P\left(\{Y = y_j\}\right) = p_i q_j$. Тогава

$$E(X.Y) = \sum_i \left(\sum_j (x_i . y_j) p_{ij} \right) = \sum_i \left(\sum_j x_i y_j p_i q_j \right) = \sum_i x_i p_i \left(\sum_j y_j q_j \right) =$$

$$= \sum_i x_i p_i \left(EY\right) = EY \sum_i x_i p_i = EY.EX$$

в) дисперсия и средно квадратично отклонение

О8: Дисперсия DX на случайната величина X се задава с дефиниционното равенство

$$DX = E\left((X - EX)^2\right).$$

Коренът на дисперсията се нарича средно квадратично отклонени σ, т.е.

$$\sigma = \sqrt{DX}\,.$$

Забележка 5:

Като се използват дефинициите за математическо очакване на дискретна и на непрекъсната случайна величина, дефиницията за дисперсия може да се преформулира, както следва.

Дисперсия DX на случайната величина X е:

$DX = \sum_i (x_i - EX)^2 p_i$ - за дискретна случайна величина

$DX = \int_{-\infty}^{+\infty} (x - EX)^2 f(x) dx$ - за непрекъсната случайна величина

г) свойства на дисперсията

Свойство 5: $D(X) = E(X^2) - (EX)^2$.

Доказателство: (с използване на свойствата на математическото очакване)

$$D(X) = E\left((X - EX)^2\right) = E\left(X^2 - 2X\ EX + (EX)^2\right) = E(X^2) - 2E(X\ EX) + E\left((EX)^2\right) =$$
$$= E(X^2) - 2EX.EX + E\left((EX)^2\right) = E(X^2) - 2(EX)^2 + (EX)^2 = E(X^2) - (EX)^2$$

Свойство 6: $D(cX) = c^2\ D(X)$, за $c = const$.

Доказателство: (с използване на свойствата на математическото очакване)

$$D(cX) = E\left((cX - E(cX))^2\right) = E\left(c^2(X - E(X))^2\right) = c^2 E\left((X - E(X))^2\right) = c^2\ D(X)$$

Свойство 7: Ако X и Y са независими случайни величини, то $D(X+Y) = D(X) + D(Y)$.

Доказателство: (с използване на свойствата на математическото очакване и първото свойство на дисперсията)

$$D(X+Y) = E\left((X+Y)^2\right) - \left(E(X+Y)\right)^2 = E\left((X+Y)^2\right) - \left(EX + EY\right)^2 =$$
$$= E\left(X^2 + 2XY + Y^2\right) - \left((EX)^2 + 2EX\ EY + (EY)^2\right) = E(X^2) + 2E(XY) + E(Y^2) -$$
$-(EX)^2 - 2EX\ EY - (EY)^2 =$ но X и Y са независими и $E(XY) = E(X)E(Y)$ тогава
$$= E(X^2) + 2EX\ EY + E(Y^2) - (EX)^2 - 2EX\ EY - (EY)^2 = E(X^2) - (EX)^2 + E(Y^2) - (EY)^2 =$$
$$= D(X) + D(Y)$$

Дискретна Случайна Величина (задачи)

Зад.1 Дадена е случайна величина с ред на разпределение

X	-1	0	2	4
P(X)	a	4a	a	a

Намерете: *a, EX* и *DX.*

Решение:

$$\sum p_i = 1 \qquad a + 4a + a + a = 1 \qquad a = \frac{1}{7}$$
$$EX = \sum p_i x_i = -a + 2a + 4a = \frac{5}{7}$$
$$E(X^2) \sum x_i^2 p_i = a + 4a + 16a = \frac{21}{7}$$
$$DX = E(X^2) + (EX)^2 = \frac{21}{7} - \frac{25}{49} = \frac{126}{49}.$$

Зад.2 Ловец, който има 4 патрона, стреля по цел до първо попадение или до изразходване на куршумите. Да се намери реда на разпределението на случайната величина *X*-задаваща броя на изразходваните патрони, ако вероятността за попадение при всеки изстрел е една и съща $\frac{1}{4}$. Пресметнете *EX* и *DX.*

Решение:

Въвеждаме събитието *A*-целта се улучва при единичен изстрел. Тогава

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

$P(X=2)=P(\overline{A})P(A)=\frac{3}{4}\frac{1}{4}=\frac{3}{16}$

$P(X=3)=P(\overline{A})P(\overline{A})P(A)=\frac{3}{4}\frac{3}{4}\frac{1}{4}=\frac{9}{64}$

$P(X=3)=P(\overline{A})P(\overline{A})P(\overline{A})=\frac{3}{4}\frac{3}{4}\frac{3}{4}=\frac{27}{64}$

Следователно реда на разпределение на *X* е

X	1	2	3	4
P(X)	$\frac{1}{4}$	$\frac{3}{16}$	$\frac{9}{64}$	$\frac{27}{64}$

$EX=\frac{1}{4}+2\frac{3}{16}+3\frac{9}{64}+4\frac{27}{64}=\frac{175}{64}$

$EX^2=\frac{1}{4}+4\frac{3}{16}+9\frac{9}{64}+16\frac{27}{64}=\frac{577}{64}$

$DX=EX^2-(EX)^2=\frac{175}{64}+\frac{577}{64}\approx 1.54$

Зад.3 Няколко души играят при следните условия. От урна в която има 1 зелена 5 червени и 2 бели топки, се изваждат последователно една след друга с връщане две топки. При изваждане на: зелена топка играчът печели 10 точки, при червена печели 2 и при бяла губи 2 точки. Да се намери реда на разпределението на случайната величина *X*-задаваща броя спечелените точки от играч. Пресметнете *EX* и *DX.*

Решение:

Въвеждаме събитието A_i, B_i и C_i-играча изважда съответно: зелена, червена или бяла топка на *i*-тото теглене, където $i=1,2$. Тогава

$P(X=20)=P(A_1\cap A_2)=\frac{1}{8}\frac{1}{8}=\frac{1}{64}$

$P(X=12)=P((A_1\cap B_2)\cup(B_1\cap A_2))=\frac{1}{8}\frac{5}{8}+\frac{5}{8}\frac{1}{8}=\frac{6}{64}$

$P(X=8)=P((A_1\cap C_2)\cup(C_1\cap A_2))=\frac{1}{8}\frac{2}{8}+\frac{2}{8}\frac{1}{8}=\frac{4}{64}$

$P(X=4)=P(B_1\cap B_2)=\frac{5}{8}\frac{5}{8}=\frac{25}{64}$

$P(X=0)=P((B_1\cap C_2)\cup(C_1\cap B_2))=\frac{5}{8}\frac{2}{8}+\frac{2}{8}\frac{5}{8}=\frac{20}{64}$

$P(X=-4)=P(C_1\cap C_2)=\frac{2}{8}\frac{2}{8}=\frac{4}{64}$

X	20	12	8	4	0	-4
P(X)	$\frac{1}{64}$	$\frac{10}{64}$	$\frac{4}{64}$	$\frac{25}{64}$	$\frac{20}{64}$	$\frac{4}{64}$

$EX=20\frac{1}{64}+12\frac{10}{64}+8\frac{4}{64}+4\frac{25}{64}-4\frac{4}{64}=4$

$$EX^2 = 400\frac{1}{64} + 144\frac{10}{64} + 8\frac{4}{64} + 16\frac{25}{64} + 16\frac{4}{64} = 38$$

$$DX = EX^2 - (EX)^2 = 38 - 16 = 22.$$

Зад.4 Да се намери математическото очакване на сл. величина X- даваща сбора от точките които могат да се паднат при хвърляне на два зара.

Решение:

Първи начин:

Като се използва следното графично представяне на пространството от ел. събития при опита хвърляне на два зара

Зар-1 / Зар-2	1	2	3	4	5	6
1	×					
2		×				
3			×			
4				×		
5					×	
6						×

Лесно може да се съобрази че X има следния ред на разпределение

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

Тогава

$$EX = 2\frac{1}{36} + 3\frac{2}{36} + 4\frac{3}{36} + 5\frac{4}{36} + 6\frac{5}{36} + 7\frac{6}{36} + 8\frac{5}{36} + 9\frac{4}{36} + 10\frac{3}{36} + 11\frac{2}{36} + 12\frac{1}{36} = 7$$

Втори начин:

Въвеждаме случайната величина Y_i - броя на точките паднали се на i-тия зар($i = 1,2$). Тогава очевидно

$$X = Y_1 + Y_2.$$

Използвайки свойствата на математическото очакване получаваме

$$EX = EY_1 + EY_2.$$

Очевидно реда на разпределение на Y_i е

Y_i	1	2	3	4	5	6
$P(Y_i)$	$\frac{1}{6}$	$\frac{1}{6}$	$\frac{1}{6}$	$\frac{1}{6}$	$\frac{1}{6}$	$\frac{1}{6}$

Следователно

$$EY_i = 1\frac{1}{6} + 2\frac{1}{6} + 3\frac{1}{6} + 4\frac{1}{6} + 5\frac{1}{6} + 6\frac{1}{6} = \frac{7}{2}.$$

Окончателно получаваме

$$EX = EY_1 + EY_2 = \frac{7}{2} + \frac{7}{2} = 7 .$$

Зад.5 Подхвърлят се едновременно два зара- бял и черен. Боря на точките на белия зар се повдигат на квадрат и от тях се вади утроения брой точки от черния зар. Да се намери математическото очакване на сл. величина X- даваща броя на точките от едно хвърляне на двата зара.

Решение:

Въвеждаме случайната величина Y_i - броя на точките паднали се на i-тия зар($i = 1,2$). Тогава

$$X = Y_1^2 - 3Y_2 .$$

Използвайки свойствата на математическото очакване получаваме

$$EX = EY_1^2 + 3EY_2 .$$

От предната задача знам, че $EY_2 = \frac{7}{2}$. За да намерим EY_1^2 ще съставим реда на Y_1^2.

Той е

Y_1^2	1	4	6	16	25	36
$P(Y_1^2)$	$\frac{1}{6}$	$\frac{1}{6}$	$\frac{1}{6}$	$\frac{1}{6}$	$\frac{1}{6}$	$\frac{1}{6}$

Следователно

$$EY_1^2 = 1\frac{1}{6} + 4\frac{1}{6} + 9\frac{1}{6} + 16\frac{1}{6} + 25\frac{1}{6} + 36\frac{1}{6} = \frac{91}{6} .$$

Окончателно получаваме

$$EX = EY_1^2 + 3EY_2 = \frac{91}{6} + 3\frac{7}{2} = \frac{14}{3} .$$

Забележка: Очевидно $EY_1^2 \neq EY_1\ EY_1$ понеже за да имаме равенство, случайните величини трябва да са независими, докато сл. величина Y_1 очевидно е зависима от себе си.

Зад.6 Да се намери реда на разпределение на сл. величина *X*- даваща броя на гербовете от 3 хвърляния на монета. Да се намерят *EX* и *DX.*

Решение:

Като използване формулата на Бернули получаваме

$$P(X = 0) = C_3^0 p^0 q^3 = \frac{1}{8}$$

$$P(X = 1) = C_3^1 p^1 q^2 = \frac{3}{8}$$

$$P(X = 0) = C_3^2 p^2 q^1 = \frac{3}{8}$$

$$P(X = 0) = C_3^3 p^3 q^0 = \frac{1}{8} .$$

В горните формули *p* и *q* са съответно вероятностите за успех и неуспех при единичен опит (в случая пада се или не се пада герб, при еднократно хвърляне на монета).

Очевидно $p = q = 0.5$.

Следователно реда на разпределение на *X* е

X	0	1	2	3
P(X)	$\frac{1}{8}$	$\frac{3}{8}$	$\frac{3}{8}$	$\frac{3}{8}$

Числовите характеристики *EX* и *DX* могат да се пресметнат като се използват формулите, използвани до сега. Този начин на пресмятаме ще оставим за упражнение на читателя. Ние ще пресметнем тези характеристики като се позовем на факта, че разглежданата в случая случайна величина е биномно разпределена, т.е. вероятностите с които заема стойностите: 0,1,…*n*, се пресмятат по формулата на Бернули $P(X=k)=C_n^k p^k q^{n-k}$. За биномно разпределена случайна величина са в сила формулите

$$EX = n\,p = 3\frac{1}{2} = \frac{3}{2}$$

$$DX = n\,p\,q = 3\frac{1}{2}\frac{1}{2} = \frac{3}{4}.$$

Непрекъсната Случайна Величина (задачи)

Зад.1 Дадена е непрекъсната случайна величина *X* с плътност на разпределение

$$f(x) = \begin{cases} 0, \text{ за } x < 0 \\ \lambda\, x, \text{ за } 0 \le x \le 2 \\ 0, \text{ за } x > 2 \end{cases}.$$

Намерете:

а) при какво значение на λ функцията $f(x)$ може да бъде приета за плътност на разпределение на сл. величина *X.*

б) числовите характеристики *EX* и *DX*;

в) вероятността $P(1 \le X \le 3/2)$;

г) функцията на разпределение $F(x)$.

Решение:

а) Използваме нормиращото условие

$$P(X < +\infty) = 1$$

$$1 = \int_{-\infty}^{+\infty} f(x)dx = \int_0^2 \lambda\, xdx = \lambda \left.\frac{x^2}{2}\right|_0^2 = 2\lambda$$

$$\lambda = \frac{1}{2}$$

б) Използвайки дефинициите за първия и втория начален момент получаваме

$$EX = \int_{-\infty}^{+\infty} xf(x)dx = \int_0^2 x\frac{x}{2}dx = \frac{4}{3}$$

$$EX^2 = \int_{-\infty}^{+\infty} x^2 f(x)dx = \int_0^2 x^2\frac{x}{2}dx = 2.$$

Използвайки познатата формула получаваме

$$DX = EX^2 - (EX)^2 = 2 - \frac{16}{9} = \frac{2}{9}.$$

в) Търсената вероятност има представянето

$P(1 \le X \le \frac{3}{2}) = P(X \le \frac{3}{2}) - P(1 < X) = \int_{-\infty}^{3/2} f(x)dx - \int_{-\infty}^{1} f(x)dx$, т.е.

$P(1 \le X \le \frac{3}{2}) = \int_{1}^{3/2} f(x)dx = \int_{1}^{3/2} x/2\ dx = \frac{9}{8} - \frac{1}{4} = \frac{1}{8}$.

г) За пресмятане на функцията на разпределение ще използваме дефиниционната зависимост $F(x) = \int_{-\infty}^{x} f(t)dt$, където *x* се разглежда като параметър. Като отчетем че *x* може да попадне в всеки един от интервалите: $(-\infty,0)$, $(0,2)$ и $(2,+\infty)$, то пресмятането на *F(x)* трябва да стане на три случая.

Случай първи: $x \in (-\infty,0)$

$F(x) = \int_{-\infty}^{x} f(t)dt = \int_{-\infty}^{x} 0\ dt = 0$.

Случай втори: $x \in (0,2)$

$F(x) = \int_{-\infty}^{x} f(t)dt = \int_{0}^{x} t/2\ \ dt = \frac{x^2}{4}$.

Случай трети: $x \in (2,+\infty)$

$F(x) = \int_{-\infty}^{x} f(t)dt = \int_{0}^{2} t/2\ \ dt = 1$.

Окончателно получаваме

$$F(x) = \begin{cases} 0,\ за\ x < 0 \\ x^2/4,\ за\ 0 \le x \le 2 \\ 1,\ за\ x > 2 \end{cases}.$$

Зад.2 Дадена е непрекъсната случайна величина *X* с плътност на разпределение

$$f(x) = \begin{cases} 0,\ за\ x < 0 \\ \lambda \cos x,\ за\ 0 \le x \le \pi/2 \\ 0,\ за\ x > 2 \end{cases}.$$

Намерете:

а) при какво значение на λ функцията $f(x)$ може да бъде приета за плътност на разпределение на сл. величина *X.*

б) числовите характеристики *EX* и *DX*;

в) вероятността $P(1 \le X \le 3/2)$.

Решение: (Аналогично на зад.1)

а)

$$1 = \int_{-\infty}^{+\infty} f(x)dx = \int_{0}^{\pi/2} \lambda \cos x dx = \lambda$$

$$\lambda = 1$$

б)

$$EX=\int_{-\infty}^{+\infty} xf(x)dx=\int_{0}^{\pi/2} x\cos xdx = x\sin x\Big|_{0}^{\pi/2}-\int_{0}^{\pi/2}\sin xdx=\frac{\pi}{2}+\cos x\Big|_{0}^{\pi/2}=\frac{\pi}{2}-1$$

$$EX^2=\int_{-\infty}^{+\infty} x^2 f(x)dx=\int_{0}^{\pi/2} x^2\cos xdx = x^2\sin x\Big|_{0}^{\pi/2}-\int_{0}^{\pi/2}\sin xdx^2=\frac{\pi^2}{4}+2\int_{0}^{\pi/2} xd\cos x$$

$$=\frac{\pi^2}{4}+2x\cos x\Big|_{0}^{\pi/2}-2\int_{0}^{\pi/2}\cos xdx=\frac{\pi^2}{4}-2\sin x\Big|_{0}^{\pi/2}=\frac{\pi^2}{4}-2$$

$$DX=EX^2-(EX)^2=\frac{\pi^2}{4}-2-\left(\frac{\pi}{2}-1\right)^2=\pi-3\,.$$

в)

$$P(1\le X\le \frac{3}{2})=\int_{1}^{3/2} f(x)dx=\int_{1}^{3/2}\cos x\,dx=\sin x\Big|_{1}^{3/2}=\sin(3/2)-\sin 1\,.$$

Зад.3 Дадена е непрекъсната случайна величина *X* с функция на разпределение

$$F(x)=\begin{cases}0,\ за\ x<1\\ \dfrac{\lambda}{13}(x^4-x),\ за\ 1\le x\le 3\\ 1,\ за\ x>3\end{cases}\,.$$

Намерете:

а) при какво значение на λ функцията $F(x)$ може да бъде приета за функция на разпределение на сл. величина *X* ;

б) плътността на разпределение $f(x)$;

в) вероятността $P(1\le X\le 2)$.

Решение:

а) Очевидно плътността *f(x)* е нула в интервалите $(-\infty,1)$ и $(3,+\infty)$. Следователно нормиращото условие придобива вида

$$1=\int_{1}^{3} f(x)dx \qquad F(3)-F(1)=1 \qquad \frac{\lambda}{13}(3^4-3)=1$$

$$\lambda=\frac{1}{6}\,.$$

б) С непосредствено диференциране получаваме

$$f(x)=\begin{cases}0,\ за\ x<1\\ \dfrac{\lambda}{13}(4x^3-1),\ за\ 1\le x\le 3\\ 0,\ за\ x>3\end{cases}\,.$$

в)

$$P(1\le X\le 2)=\int_{1}^{2} f(x)dx=F(2)-F(1)=\frac{1}{78}(2^4-2)-\frac{1}{78}(1^4-1)=\frac{7}{39}\,.$$

Зад.4 Дадена е непрекъсната случайна величина *X* с функция на разпределение

$$F(x) = \begin{cases} 1 - \dfrac{a^3}{x^3}, \text{ за } x \geq a \\ 0, \quad \text{за} \quad a < x \end{cases}$$, където *a>0*.

Намерете:

а) плътността на разпределение $f(x)$;

б) при какво значение на *a* функцията $F(x)$ може да бъде приета за функция на разпределение на сл. величина *X* ;

в) числовите характеристики *EX* и *DX*;.

Решение:

а)

$$f(x) = \begin{cases} 3\dfrac{a^3}{x^4}, \text{ за } x \geq a \\ 0, \quad \text{за} \quad a < x \end{cases} .$$

б)

$$\int_{-\infty}^{+\infty} f(x)dx = \int_{a}^{+\infty} 3\frac{a^3}{x^4}dx = a^3 x^{-3}\Big|_a^{+\infty} = 1$$, за всяко *a>0*.

Това показва, че функцията F(x) може да бъде приета за функция на разпределение на сл. величина *X* за всяко *a>0*.

в)

$$EX = \int_{-\infty}^{+\infty} xf(x)dx = \int_{a}^{+\infty} x\frac{3a^3}{x^4}dx = -\frac{3}{2}a^3 x^{-2}\Big|_a^{+\infty} = \frac{3}{2}a$$

$$EX^2 = \int_{-\infty}^{+\infty} x^2 f(x)dx = \int_{0}^{+\infty} x^2 \frac{3a^3}{x^4 a}dx = -3a^3 x^{-1}\Big|_a^{+\infty} = 3a^2$$

$$DX = EX^2 - (EX)^2 = 3a^2 - \left(\frac{3a}{2}\right)^2 = \frac{3}{4}a^2 .$$

Зад.5 Дадена е непрекъсната случайна величина *X* с функция на разпределение

$$F(x) = a + b\, arctg\frac{x}{a}$$, за $-\infty < x < \infty$.

Намерете:

а) при какво значение на *a* и *b* функцията $F(x)$ може да бъде приета за функция на разпределение на сл. величина *X* ;

б) плътността на разпределение $f(x)$;

в) вероятността $P(-\frac{1}{2} \leq X \leq \frac{1}{2})$.

Решение:

а) Като използваме зависимостите

$$F(-\infty) = 0$$
$$F(+\infty) = 1$$

получаваме

$$\begin{cases} a - \dfrac{\pi}{2} b = 0 \\ a + \dfrac{\pi}{2} b = 1 \end{cases} \text{,т.е.} \quad \begin{cases} a = \dfrac{1}{2} \\ b = \dfrac{1}{\pi} \end{cases}.$$

б)

$$f(x) = F'(x) = \frac{1}{\pi} \frac{2}{1+4x^2}.$$

в) $P(-1/2 \le X \le 1/2) = F(1/2) - F(-1/2) = \dfrac{1}{2} + \dfrac{1}{\pi} arctg1 - \dfrac{1}{2} + \dfrac{1}{\pi} arctg(-1) = \dfrac{2}{\pi} arctg1 = \dfrac{1}{2}$

Забележка: Оставяме на читателя сам да се увери че тази случайна величина няма математическо очакване. За целта е достатъчно да се пресметне интеграла с които се дефинира *EX* и да се установи, че той е разходящ.

ТЕМА 4

Някои често използвани разпределения в теория на вероятностите - биномно, поасоново, равномерно, нормално и експоненциално

1. Биномно разпределение.

а) определение

О1: Дискретната случайна величина X е биномно разпределена с параметри n и p / $X \in B_i(n,p)$ / ако

$$P(X=k)=C_n^k p^k (1-p)^{n-k} \text{ за } k=0,1,\ldots,n.$$

Забележка 1: От Въпрос 2 (Схема на Бернули) директно следва, че при $X \in B_i(n,p)$ вероятността $P(X=k)$ е точно вероятността за k-успеха от n- опита, където p е вероятността за успех при еднократно провеждане на опита.

б) числови характеристики

За да пресметнем числовите характеристики на биномно разпределената случайна величина X, ще обърнем внимание на факта, че

$$X=X_1+\ldots+X_n,$$

където X_i е случайна величина, заемаща стойност 1, ако единичното събитие се сбъдне при i-опит и стойност 0, ако не се сбъдне, т.е.

$$P(X_i=1)=p \text{ и } P(X_i=0)=1-p=q.$$

При това

$$E(X_i)=0.p+1.q=q$$

$$D(X_i)=E(X_i^2)-E(X_i)^2=(0.p+1.q)-(0.p+1.q)^2=q-q^2=q(1-q)=q\,p.$$

Тогава

$$E(X)=E\left(\sum_{i=1}^n X_i\right)=\sum_{i=1}^n E(X_i)=nq$$

$D(X)=D\left(\sum_{i=1}^n X_i\right)=$ понеже X_i са независими случайна величина

$$=\sum_{i=1}^n D(X_i)=npq.$$

2. Поасоново разпределение.

а) определение

О2: Дискретната случайна величина X е поасоново разпределена с параметър λ / $X \in P_o(\lambda)$ /, ако

$$P(X=k)=\frac{\lambda^k}{k!}e^{-\lambda} \text{ за } k=0,1,\ldots,\infty.$$

Забележка 2: Може да се покаже, че ако $X \in B_i(n,p)$ и $Y \in P_o(\lambda)$, където за параметрите е в сила: $n \to \infty$ и $n\,p \to \lambda$ (в практически задачи тези изисквания се заменят с: n-“голямо” и $n\,p$-едноцифрено число), тогава

$$C_n^k p^k (1-p)^{n-k} \to \frac{\lambda^k}{k!}e^{-\lambda}$$

т.е.

$$P(X=k)\approx P(Y=k).$$

б) числови характеристики

$$E(X)=\sum_{k=0}^{\infty}k\,P(X=k)=\sum_{k=0}^{\infty}k\,\frac{\lambda^k}{k!}e^{-\lambda}=\sum_{k=1}^{\infty}\frac{\lambda^k}{(k-1)!}e^{-\lambda}=\lambda e^{-\lambda}\sum_{k=1}^{\infty}\frac{\lambda^{k-1}}{(k-1)!}=\lambda e^{-\lambda}e^{\lambda}=\lambda$$

$$E(X^2)=\sum_{k=0}^{\infty}k^2\,P(X=k)=\sum_{k=0}^{\infty}k^2\,\frac{\lambda^k}{k!}e^{-\lambda}=\sum_{k=1}^{\infty}k\,\frac{\lambda^k}{(k-1)!}e^{-\lambda}=\sum_{k=1}^{\infty}(k-1+1)\,\frac{\lambda^k}{(k-1)!}=$$

$$=\sum_{k=1}^{\infty}(k-1)\,\frac{\lambda^k}{(k-1)!}e^{-\lambda}+\sum_{k=1}^{\infty}\frac{\lambda^k}{(k-1)!}e^{-\lambda}=\sum_{k=2}^{\infty}\frac{\lambda^k}{(k-2)!}e^{-\lambda}+\sum_{k=1}^{\infty}\frac{\lambda^k}{(k-1)!}e^{-\lambda}=$$

$$=\lambda^2e^{-\lambda}\sum_{k=2}^{\infty}\frac{\lambda^{k-2}}{(k-2)!}+\lambda e^{-\lambda}\sum_{k=1}^{\infty}\frac{\lambda^{k-1}}{(k-1)!}=\lambda^2e^{-\lambda}e^{\lambda}+\lambda e^{-\lambda}e^{\lambda}=\lambda^2+\lambda$$

$$D(X)=E(X^2)-E(X)^2=\lambda^2+\lambda-(\lambda)^2=\lambda.$$

3. Равномерно разпределение.

а) определение

О3: Непрекъснатата случайна величина X е равномерно разпределена в интервала $\{a,b]$ / $X\in U(a,b)$/ ,ако плътността й е

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

б) числови характеристики

$$E(X)=\int_{-\infty}^{+\infty}x\,f(x)dx=\int_a^b x\frac{1}{b-a}dx=\frac{1}{b-a}\left.\frac{x^2}{2}\right|_{x=a}^{x=b}=\frac{1}{b-a}\frac{b^2-a^2}{2}=\frac{b+a}{2}$$

$$E(X^2)=\int_{-\infty}^{+\infty}x^2\,f(x)dx=\int_a^b x^2\frac{1}{b-a}dx=\frac{1}{b-a}\left.\frac{x^3}{3}\right|_{x=a}^{x=b}=\frac{1}{b-a}\frac{b^3-a^3}{3}=\frac{b^2+ba+a^2}{3}$$

$$D(X)=E(X^2)-E(X)^2=\frac{b^2+ba+a^2}{3}-\left(\frac{b+a}{2}\right)^2=\frac{4b^2+4ba+4a^2-3b^2-6ba-3a^2}{12}=$$

$$=\frac{b^2-2ba+a^2}{12}=\frac{(b-a)^2}{12}$$

4. Нормално разпределение.

а) определение

О4: Непрекъснатата случайна величина X е равномерно разпределена с параметри μ и σ / $X\in N(\mu,\sigma)$/, ако плътността й е

$$f(x)=\frac{1}{\sigma\sqrt{2\pi}}e^{-\frac{1}{2}\left(\frac{x-\mu}{\sigma}\right)^2},\ \text{за } -\infty<x<+\infty.$$

Забележка 3: Може да се покаже, че ако $X\in B_i(n,p)$ и $Y\in N(\mu,\sigma)$, където за параметрите е в сила: $n\to\infty$, $np\to\mu$ и $\sqrt{npq}\to\sigma$ (в практически задачи тези изисквания се заменят с: n-“голямо” и np-“голямо”), тогава

$$C_n^k p^k(1-p)^{n-k}\to\frac{1}{\sigma\sqrt{2\pi}}e^{-\frac{1}{2}\left(\frac{k-\mu}{\sigma}\right)^2}$$

и

$$\sum_{k=k_1}^{k_2} C_n^k p^k (1-p)^{n-k} \to \frac{1}{\sigma\sqrt{2\pi}} \int_{(k_1-\mu)/\sigma}^{(k_2-\mu)/\sigma} e^{-\frac{1}{2}\left(\frac{k-\mu}{\sigma}\right)^2} dk$$

т.е.

$$P(k_1 \le X \le k_2) \approx P(k_1 \le Y \le k_2).$$

б) числови характеристики

$$E(X) = \int_{-\infty}^{+\infty} x\, f(x)dx = \frac{1}{\sigma\sqrt{2\pi}} \int_{-\infty}^{+\infty} x\, e^{-\frac{1}{2}\left(\frac{x-\mu}{\sigma}\right)^2} dx = \frac{1}{\sigma\sqrt{2\pi}} \int_{-\infty}^{+\infty} (x-\mu+\mu)\, e^{-\frac{1}{2}\left(\frac{x-\mu}{\sigma}\right)^2} dx =$$

$= \frac{1}{\sigma\sqrt{2\pi}} \int_{-\infty}^{+\infty} (x-\mu)\, e^{-\frac{1}{2}\left(\frac{x-\mu}{\sigma}\right)^2} d(x-\mu) + \frac{1}{\sigma\sqrt{2\pi}} \int_{-\infty}^{+\infty} \mu\, e^{-\frac{1}{2}\left(\frac{x-\mu}{\sigma}\right)^2} dx =$ първият интеграл е в симетрични граници от нечетна подинтегрална функция, следователно стойността му е нула

$= \mu \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{+\infty} e^{-\frac{1}{2}\left(\frac{x-\mu}{\sigma}\right)^2} d\left(\frac{x-\mu}{\sigma}\right) = \mu \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{+\infty} e^{-\frac{1}{2}t^2} dt =$ този интеграл е известен като интеграл на Поасон и е доказано, че стойността му е $\sqrt{2\pi}$, следователно

$$= \mu \frac{1}{\sqrt{2\pi}} \sqrt{2\pi} = \mu$$

$$D(X) = \int_{-\infty}^{+\infty} (x-EX)^2\, f(x)dx = \frac{1}{\sigma\sqrt{2\pi}} \int_{-\infty}^{+\infty} (x-\mu)^2\, e^{-\frac{1}{2}\left(\frac{x-\mu}{\sigma}\right)^2} dx = \frac{\sigma^2}{\sqrt{2\pi}} \int_{-\infty}^{+\infty} \frac{(x-\mu)^2}{\sigma^2} e^{-\frac{1}{2}\left(\frac{x-\mu}{\sigma}\right)^2} d\left(\frac{x-\mu}{\sigma}\right) =$$

$$= \frac{\sigma^2}{\sqrt{2\pi}} \int_{-\infty}^{+\infty} t^2 e^{-\frac{1}{2}t^2} dt = \frac{\sigma^2}{\sqrt{2\pi}} \int_{-\infty}^{+\infty} t\, e^{-\frac{1}{2}t^2} d\left(\frac{t^2}{2}\right) = -\frac{\sigma^2}{\sqrt{2\pi}} \int_{-\infty}^{+\infty} t\, d\left(e^{-\frac{1}{2}t^2}\right) = -\frac{\sigma^2}{\sqrt{2\pi}} te^{-\frac{1}{2}t^2}\Big|_{-\infty}^{+\infty} + \frac{\sigma^2}{\sqrt{2\pi}} \int_{-\infty}^{+\infty} e^{-\frac{1}{2}t^2} dt =$$

$= 0 + \frac{\sigma^2}{\sqrt{2\pi}} \int_{-\infty}^{+\infty} e^{-\frac{1}{2}t^2} dt =$ това отново е интегралът на Поасон, чиято стойност е $\sqrt{2\pi}$, следователно $= \frac{\sigma^2}{\sqrt{2\pi}} \sqrt{2\pi} = \sigma^2$.

5. Експоненциално разпределени.

а) определение

О2: Дискретната случайна величина X е експоненциално разпределена с параметър λ / $X \in Exp(\lambda)$ /, ако

$$f(x) = \begin{cases} \lambda e^{-\lambda x}, \text{ за } x \ge 0 \\ 0, \text{ за } x < 0 \end{cases}$$

б) числови характеристики

$$E(X) = \int_{-\infty}^{+\infty} x\, f(x)dx = \int_{0}^{+\infty} \lambda\, x e^{-\lambda x} dx = -\int_{0}^{+\infty} x\, d(e^{-\lambda x}) = -x e^{-\lambda x}\Big|_0^{\infty} + \int_{0}^{+\infty} e^{-\lambda x} dx = -\frac{e^{-\lambda x}}{\lambda}\Big|_0^{\infty} = \frac{1}{\lambda}$$

$$E(X^2) = \int_{-\infty}^{+\infty} x^2\, f(x)dx = \lambda \int_{0}^{+\infty} x^2\, e^{-\lambda x} dx = -\int_{0}^{+\infty} x^2 d(e^{-\lambda x}) = -x^2\, e^{-\lambda x}\Big|_0^{\infty} + \int_{0}^{+\infty} e^{-\lambda x} d(x^2) =$$

$$= 2\int_{0}^{+\infty} x\, e^{-\lambda x} dx = \frac{2}{\lambda} \int_{0}^{+\infty} \lambda\, x\, e^{-\lambda x} dx = \frac{2}{\lambda} E(X) = \frac{2}{\lambda}\frac{1}{\lambda} = \frac{2}{\lambda^2}$$

$$D(X) = E(X^2) - E(X)^2 = \frac{2}{\lambda^2} - \left(\frac{1}{\lambda}\right)^2 = \frac{1}{\lambda^2}$$

Нормално и Биномно Разпределения (задачи)

Зад.1 Случайната величина *X* даваща дължината на изготвен от автомат детайл, е нормално разпределена. Стандартната дължина на тези детайли е 10 см., а стандартното отклонение е 0.07см. Намерете вероятността за брак, ако допустимите размери на детайлите трябва да са 10 ± 0.05см.

Решение:

В тази задача търсим вероятността

$$P(X < 9.95 \cup X > 10.05) = 1 - P(9.95 < X < 10.05).$$

Случайната величина *X* е нормално разпределена. Това означава, че

$$(1) \quad P(x_1 < X < x_2) = \frac{1}{\sigma\sqrt{2\pi}} \int_{x_1}^{x_2} e^{-\frac{(t-\mu)^2}{2\sigma^2}} dt .$$

Въвеждаме стандартното означение

$$(2) \qquad \Phi(z) = \frac{1}{\sqrt{2\pi}} \int_0^z e^{-\frac{t^2}{2}} dt \quad и \quad \Phi(-z) = -\Phi(z), \; за \; z > 0$$

Горния интеграл не е решим в елементарни функции, но той се решава числено. Стойностите получени от това решение се нанасят в таблица.

Очевидно ако положим

$$(3) \qquad z = \frac{x-\mu}{\sigma}$$

в интеграла (1), то той може да се сведе към разлика от два интеграла от вида (2),т.е

$$\frac{1}{\sigma\sqrt{2\pi}} \int_{x_1}^{x_2} e^{-\frac{(t-\mu)^2}{2\sigma^2}} dt = \frac{1}{\sqrt{2\pi}} \int_{z_1}^{z_2} e^{-\frac{t^2}{2}} dt = \Phi(z_2) - \Phi(z_1) .$$

Следователно ако използваме численото решение на интеграла (2), то ние можем да намерим численото решение и на интеграла (1).

Забележка: Често се използва полагането $z = \frac{x-\mu}{\sigma\sqrt{2}}$, което води до модификации във вида на $\Phi(z)$. Срещат се и други модификации в начина на дефиниране на $\Phi(z)$, породени от други съображения.

Забележка: Полагането (3) се нарича свеждане на нормално разпределената сл. величина *X* с параметри μ и σ, към стандартно нормално разпределената сл. величина *Z* с параметри $\mu = 0$ и $\sigma = 1$.

Като използваме полагането (3), за търсената вероятност получаваме

$$1 - P(9.95 < X < 10.05) = 1 - P(\frac{9.95-\mu}{\sigma} < Z < \frac{10.05-\mu}{\sigma}) .$$

От условието имаме $\mu = 10$ и $\sigma = 0.07$, следователно

$$1 - P(\frac{9.95-\mu}{\sigma} < Z < \frac{10.05-\mu}{\sigma}) = 1 - P(\frac{9.95-10}{0.07} < Z < \frac{10.05-10}{0.07})$$

$$= 1 - P(-0.72 < Z < 0.72) = 1 - 2P(0 < Z < 0.72) == 1 - 2\;0.2624 = 0.4716$$

Зад.2 Едно събитие има единична вероятност p=0.4. Каква е вероятността то да се появи не по-малко от 35 пъти и не повече от 50 пъти при 100 независими изпитания?
Упътване: Да се използва нормално приближение на биномното разпределение.
Решение:
Въвеждаме случайната величина *X*-броя на появяванията на единичното събитие при 100 опита.
Очевидно търсим вероятността е $P(35 \le X \le 50)$.
Като използваме формулата на Бернули получаваме

$$P(35 \le X \le 50) = \sum_{i=35}^{50} C_{100}^{i}\, 0.4^{i}\ 0.6^{100-i}\ .$$

На практика горната формула е безполезна понеже съдържа твърде големи факториели. За това използваме интегралната теорема на Моавар-Лаплас и получаваме

$$P(35 \le X \le 50) = \sum_{i=35}^{50} C_{100}^{i}\, 0.4^{i}\ 0.6^{100-i} \approx \frac{1}{\sigma\sqrt{2\pi}} \int_{35}^{50} e^{-\frac{(t-\mu)^2}{2\sigma^2}}\, dt\,,$$

където $\mu = np = 100\ 0.4 = 40$ и $\sigma = npq = \sqrt{100\ 0.4\ 0.6} \approx 4.9$.

Забележка: За да е в сила горното приблизително равенство е необходимо *n* и *np* да са относително големи числа.
Използвайки полагането (3), аналогично на зад.1 получаваме

$$P(35 \le X \le 50) = P(\frac{35-40}{4.9} < Z < \frac{50-40}{4.9}) = P(-1.02 < Z < 2.04)$$
$$= P(0 < Z < 2.04) + P(0 < Z < 1.02) = 0.4793 + 0.3461 = 0.8254\,.$$

Зад.3 Колко независими изпитания трябва да се проведат, така че вероятността единичното събитие да се появи не по-малко от 5 пъти, да е 0.8. Вероятността на единичното събитие е 0.05.
Забележка: Вероятността на единичното събитие е една и съща във всеки опит.
Решение:
Въвеждаме случайната величина *X*-броя на появяванията на единичното събитие при *n* опита.
От условието следва

$$P(5 \le X \le n) = 0.8\,.$$

Аналогично на зад.2 получаваме

$$0.8 = P(5 \le X \le n) = P(\frac{5-n\ 0.05}{\sqrt{n\ 0.05\ 0.95}} < Z < \frac{n-n\ 0.05}{\sqrt{n\ 0.05\ 0.95}}) = P(\frac{5-n\ 0.05}{0.22\sqrt{n}} < Z < 4.35\sqrt{n})$$

Понеже *n>5*, то $P(\frac{5-n\ 0.05}{0.22\sqrt{n}} < Z < 4.35\sqrt{n}) = P(\frac{5-n\ 0.05}{0.22\sqrt{n}} < Z < \infty)$, а $\frac{5-n\ 0.05}{0.22\sqrt{n}}$ е отрицателно.
Следователно

$$0.8 = P(\frac{5-n\ 0.05}{0.22\sqrt{n}} < Z < 4.35\sqrt{n}) = P(\frac{5-n\ 0.05}{0.22\sqrt{n}} < Z < \infty) =$$

$$= P(0 < Z < \frac{5 - n\,0.05}{0.22\sqrt{n}}) + P(0 < Z < \infty) == P(0 < Z < \frac{5 - n\,0.05}{0.22\sqrt{n}}) + 0.5$$

Тогава

$$0.3 == P(0 < Z < \frac{5 - n\,0.05}{0.22\sqrt{n}})$$

$$\frac{5 - n\,0.05}{0.22\sqrt{n}} = -0.84$$

$$-0.05n + 0.183\sqrt{n} + 5 = 0$$

Решаваме горното квадратно уравнение (относно $\sqrt{n}$) и получаваме

$$\sqrt{n} \approx 12 \text{ или } \sqrt{n} \approx -8.5\,.$$

Очевидно второто решение няма физически смисъл и следователно получаваме

$$n \geq 144\,.$$

Получихме, че са необходими поне 144 изпитания, така че вероятността за поне 5 появявания на единичното събитие да е по-голяма или равна на 0.8.

Зад.4 Нормално в ловен патрон се поставят 2.3 гр. барут. Знае се, че стандартното отклонение на случайните грешки правени при отмерването на това количество е 0.15 гр. Да се намери вероятността за повреда на оръжието, ако максимално допустимия заряд е 2,5 гр.

Решение:

Въвеждаме случайната величина *X*-количеството барут поставен в патрона.

Следователно в задачата се търси вероятността

$$P(2.5 < X < +\infty).$$

Използваме полагането (3) и получаваме

$$P(2.5 < X < +\infty) = P(\frac{2.5 - 2.3}{0.15} < Z < \frac{\infty - 2.3}{0.15}) = P(1.33 < Z < +\infty)$$
$$= P(0 < Z < \infty) - P(0 < Z < 1.33) = 0.5 - 0.4082 = 0.0918$$

Зад.5 Нека средно на 10000 души един има рядка кръвна група. Каква е вероятността в град с 10000 души население да няма нито един човек с такава кръвна група.

Решение:

В тази задача се търси вероятността от 10000 опита да имаме нула успеха, където под успех разбираме наблюдавания човек да има рядка кръвна група.

По формулата на Бернули тя е

$$P_{0,10000} = C_{10000}^{0}\, p^0\, q^{10000}.$$

Понеже средно на 10000 души един е с рядка кръвна група, то единичната вероятност $p = \frac{1}{10000}$. Директното пресмятане на горния израз е трудно, за това ние ще използваме поасоновото приближение на биномното разпределение, т.е.

$$P_{0,10000} = C_{10000}^{0}\, p^0\, q^{10000} \approx \frac{(np)^0}{0!} e^{-np},$$

където $n = 10000$ и $p = \frac{1}{10000}$. Следователно

$$P_{0,10000} = e^{-1} \approx 0.368\,.$$

Забележка: Поасоновото приближение на биномното разпределение се прилага в случаите, когато *n*-голямо и *np*-няколко еденици.

Зад.6 Апаратура съдържа 2000 еднакво надеждни елемента. Вероятността за повреда на всеки от тях е *p=0.05*. Каква е вероятността за отказ на апаратурата, ако апаратурата се поврежда при повреда на поне един елемент.

Решение:

В случая се търси вероятността от 2000 опита да имаме поне един успеха, където под успех разбираме повреда на даден елемент.

Следователно търсената вероятност е

$$1-P_{0,2000}=1-C_{2000}^{0}\,p^{0}\,q^{2000}.$$

В случая ще приближаваме биномното разпределение с помощта на локалната теорема на Моавар-Лаплас, т.е.

$$P_{0,10000}=C_{10000}^{0}\,p^{0}\,q^{10000}\approx\frac{1}{\sigma\sqrt{2\pi}}e^{-\frac{(o-\mu)^2}{2\sigma^2}},$$

където $\mu=np=100$ и $\sigma=\sqrt{npq}=\sqrt{2000\;0.05\;0.95}=\sqrt{95}$.

Окончателно получаваме

$$P_{0,10000}\approx\frac{1}{\sqrt{2\pi\,95}}e^{-\frac{100^2}{190}}\approx 0.024.$$

Забележка: Приближението на Моавар-Лаплас на биномното разпределение се прилага в случаите, когато *n* и *np* са относително големи.

ТЕМА 5
Сума, произведение и частно на случайни величини. Функция от случайна величина

1. Въведение

Ще разгледаме сума, произведение и частно на две независими случайни величини X и Y, имащи плътности на разпределение $f_X(x)$ и $f_Y(y)$. Тогава тяхната съвместна функция на разпределение е

$$F(x,y) = \iint_D f_X(x) f_Y(y) dxdy \,,$$

където областта D е такава, че за $\forall(x,y) \in D$ е изпълнено едно от условията:

1 сл. (за събиране): $D \equiv \{(x,y) : x + y \le z\}$;

2 сл. (за умножение): $D \equiv \{(x,y) : x.y \le z\}$;

3 сл. (за деление): $D \equiv \{(x,y) : x / y \le z\}$;

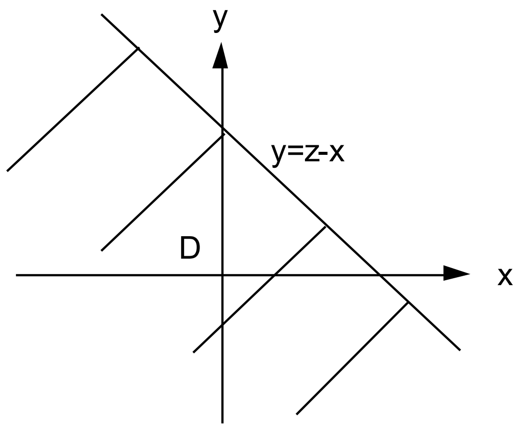
където z е реален параметър (фиксирано реално число).

2. Сума на случайни величини.

Разглеждаме случайната величина $Z = X + Y$.

$$F(z) = P(X + Y < z) = \iint_D f_X(x) f_Y(y) dxdy \,,$$

където $D : \begin{cases} -\infty < x < \infty \\ -\infty < y < z - x \end{cases}$



Фиг.1

Тогава

$$F(z) = \int_{-\infty}^{\infty} \left(\int_{-\infty}^{z-x} f_X(x) f_Y(y) dy \right) dx = \int_{-\infty}^{\infty} f_X(x) F_Y(z-x) dx = \int_{-\infty}^{\infty} F_Y(z-x) dF_X(x)$$

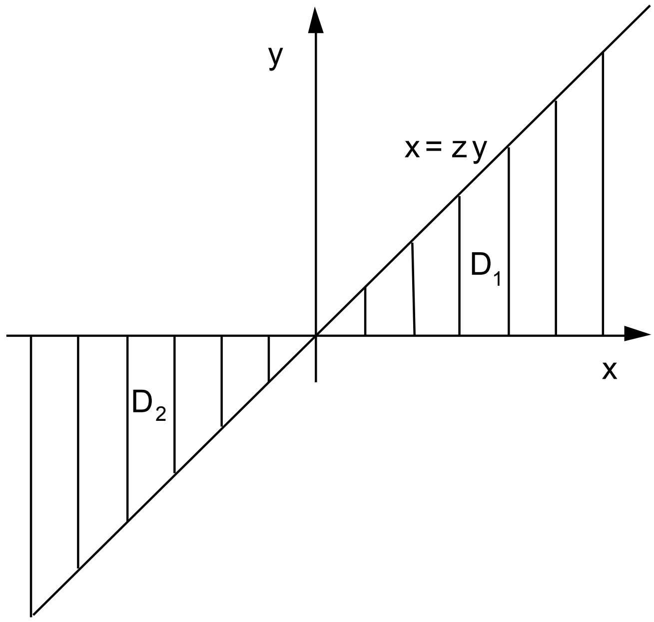
$$f(z) = \frac{dF(z)}{dz} = \int_{-\infty}^{\infty} f_X(x) f_Y(z-x) dx \,.$$

3. Частно на случайни величини.

Разглеждаме случайната величина $Z = X / Y$.

$$F(z) = P(X / Y < z) = \iint_D f_X(x) f_Y(y) dxdy \,,$$

където $D = D_1 \cup D_2$, $D_1 : \begin{cases} -\infty < y < \infty \\ -\infty < x < z\,y \end{cases}$ и $D_2 : \begin{cases} -\infty < y < 0 \\ z\,y < x < \infty \end{cases}$



Фиг.2

Тогава

$$F(z) = \int_{-\infty}^{0}\left(\int_{z\,y}^{\infty} f_X(x) f_Y(y)dx\right)dy + \int_{0}^{\infty}\left(\int_{-\infty}^{z\,y} f_X(x) f_Y(y)dx\right)dy =$$

$$= -\int_{-\infty}^{0} f_Y(y) F_X(z\,y)dy + \int_{0}^{\infty} f_Y(y) F_X(z\,y)dy$$

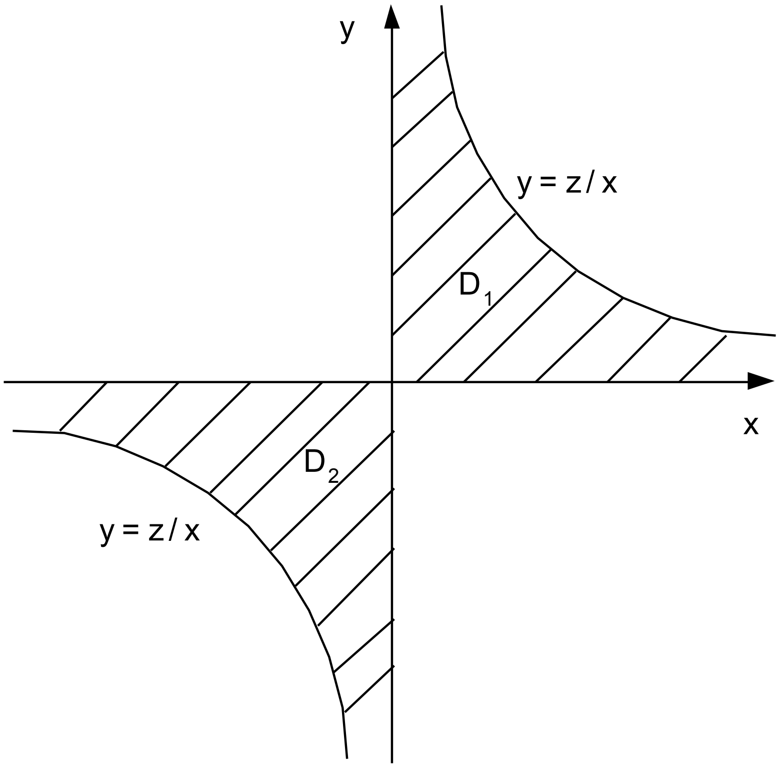
$$f(z) = \frac{dF(z)}{dz} = -\int_{-\infty}^{0} y\; f_Y(y) f_X(z\,y)dy + \int_{0}^{\infty} y\; f_Y(y) f_X(z\,y)dy = \int_{-\infty}^{\infty} |\,y\,|\; f_Y(y) f_X(z\,y)dy\,.$$

3. Произведение на случайни величини.

Разглеждаме случайната величина $Z = X.Y$.

$$F(z) = P(X.Y < z) = \iint_D f_X(x) f_Y(y)dxdy\,,$$

където $D = D_1 \cup D_2$, $D_1 : \begin{cases} -\infty < x < \infty \\ -\infty < u < z/x \end{cases}$ и $D_2 : \begin{cases} -\infty < x < 0 \\ z/x < y < \infty \end{cases}$



Фиг.3

Тогава

$$F(z) = \int_{-\infty}^{0}\left(\int_{z/x}^{\infty} f_X(x) f_Y(y)dy\right)dx + \int_{0}^{\infty}\left(\int_{-\infty}^{z/x} f_X(x) f_Y(y)dy\right)dx =$$

$$= -\int_{-\infty}^{0} f_Y(x)F_X(z/x)dx + \int_{0}^{\infty} f_X(y)F_Y(z/x)dx$$

$$f(z) = \frac{dF(z)}{dz} = -\int_{-\infty}^{0} y\ f_Y(x) f_X(z\ y)dy + \int_{0}^{\infty} y\ f_Y(y) f_X(z\ y)dy = \int_{-\infty}^{\infty} |\ y\ |\ f_Y(y) f_X(z\ y)dy\,.$$

4. Функция от случайна величина.

Разглеждаме функцията $\varphi(\cdot)$ от случайната величина X, т.е. $\varphi(X)$

1 случай. Функцията $\varphi(\cdot)$ е еднозначно обратима, т.е.

$$y = \varphi(x) \Leftrightarrow x = \psi(y)$$

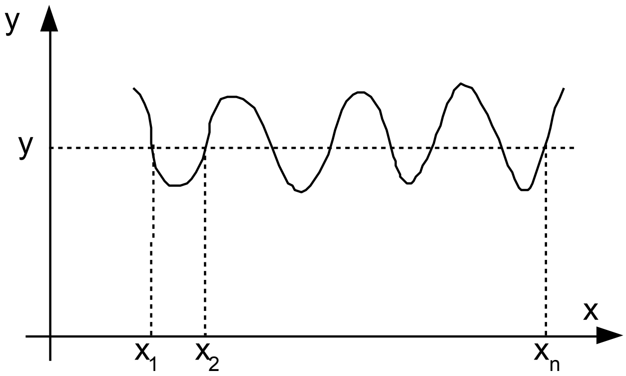
Тогава

$$F_Y(y) = P(\varphi(X) < y) = P(X < \psi(y)) < y) = \int_{-\infty}^{\psi(y)} f_X(t)dt$$

$$f_Y(y) = \frac{dF_Y(y)}{dy} = f_X(\psi(y))\,|\,\psi'(y)\,|$$

2 случай. Функцията $\varphi(\cdot)$ е такава, че на едно y съпоставяме n различни аргумента $x_1, \ldots, x_n$, т.е. обръщането на функцията става по елементно, т.е.

$$x_1 = \psi_1(y),\ x_2 = \psi_2(y), \cdots, x_н = \psi_n(y)$$



Фиг.4

Тогава

$$f_Y(y) = \sum_{i=1}^{n} f_X(\psi_i(y))\,|\,\psi_i'(y)\,|$$

Сума, разлика, произведение и частно на случайни величини (задачи)

Зад.1 Да се намери плътността на разпределение на случайната величина $Z = \dfrac{X}{Y}$, ако сл. величини *X* и *Y* са независими и експоненциално разпределени с параметър λ,т.е. $X, Y \in Exp(\lambda)$.

Решение:

По условие $f(x) = \begin{cases} \lambda e^{-\lambda x},\ за\ x > 0 \\ 0,\ за\ x \le 0 \end{cases}$ и $f(y) = \begin{cases} \lambda e^{-\lambda y},\ за\ y > 0 \\ 0,\ за\ y \le 0 \end{cases}$.

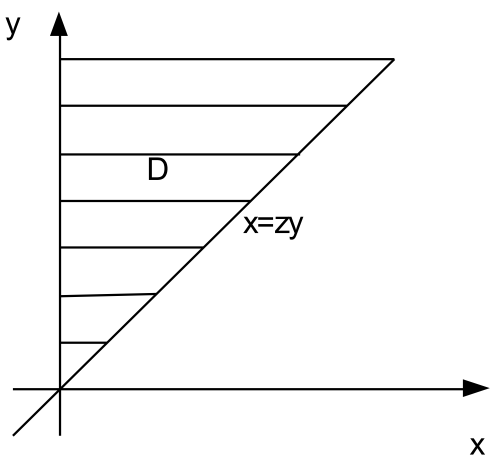
Ще намерим функцията на разпределение

$$F(z) = P(Z < z) = P(X < zY) = \iint_D f(x,y)\,dxdy\,.$$

Където за областта *D* имаме следните два случая

1 случай: *z>0*

В този случай областта *D* изглежда, така



Следователно

$$F(z)=\int_0^{+\infty}dy\int_0^{zy}f(x,y)dxdy=\int_0^{+\infty}\lambda e^{-\lambda y}dy\int_0^{zy}\lambda e^{-\lambda x}dx$$

$$=-\int_0^{+\infty}\lambda e^{-\lambda y}\,\lambda e^{-\lambda x}\Big|_{x=0}^{x=zy}dy=-\int_0^{+\infty}\lambda e^{-\lambda y}(\lvert\lambda e^{-\lambda zy}-1)dy$$

$$=-\int_0^{+\infty}e^{-\lambda(z+1)y}d(\lambda y)+\int_0^{+\infty}e^{-\lambda y}d(\lambda y)$$

$$=\frac{e^{-\lambda(z+1)y}}{z+1}\Bigg|_{y=0}^{y=+\infty}-e^{-\lambda y}\Big|_{y=0}^{y=+\infty}$$

$$=-\frac{1}{z+1}+1$$

2 случай: z<0

В този случай $F(z)=P(X<zY)=0$, понеже експоненциално разпределените сл. величини *X* и *Y*, могат да приемат само положителни стойности.

За плътността на *Z* имаме

$$f(z)=\begin{cases}\dfrac{1}{(1+z)^2},\ за\ z>0\\ 0,\ за\ z\le 0\end{cases}.$$

Зад.2 Да се намери плътността на разпределение на случайната величина $Z=XY$, ако сл. величини *X* и *Y* са независими и равномерно разпределени в интервала (0,1),т.е. $X,Y\in U(0,1)$.

Решение:

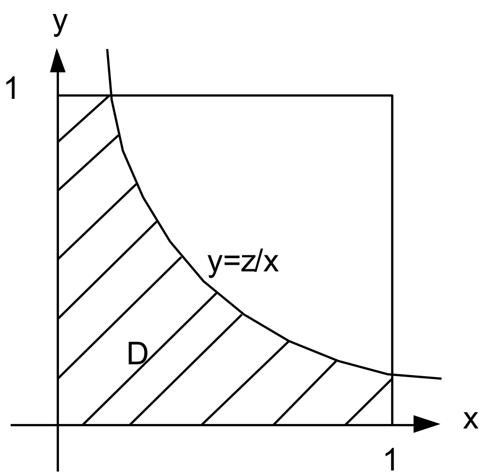
По условие $f(x)=\begin{cases}1,\ за\ x\in(0,1)\\0,\ за\ x\notin(0,1)\end{cases}$ и $f(y)=\begin{cases}1,\ за\ y\in(0,1)\\0,\ за\ y\notin(0,1)\end{cases}$.

Ще пресметнем функцията на разпределение

$$F(z)=P(Z<z)=P(X<\frac{z}{Y})=\iint_D f(x,y)dxdy.$$

Където за областта *D* имаме следните два случая

1 случай: *0<z<1*



Следователно

$$F(z)=\int_0^z dx\int_0^1 1dy+\int_z^1 dx\int_0^1 1dy=\int_0^z y\Big|_0^1 dx+\int_z^1 y\Big|_0^{\frac{z}{x}} dx$$

$$=\int_0^z dx+\int_z^1 z/x\,dx=z+z\,\ln x\Big|_z^1=z-z\,\ln z$$

2 случай: $z\notin(0,1)$

В този случай *F(z)=const* , понеже вероятността *P(Z<z)* не зависи от стойността на *z*, когато $z\notin(0,1)$.

Следователно плътността на *Z* е

$$f(z)=\begin{cases}-\ln z,\ за\ z\in(0,1)\\ 0,\ за\ z\notin(0,1)\end{cases}.$$

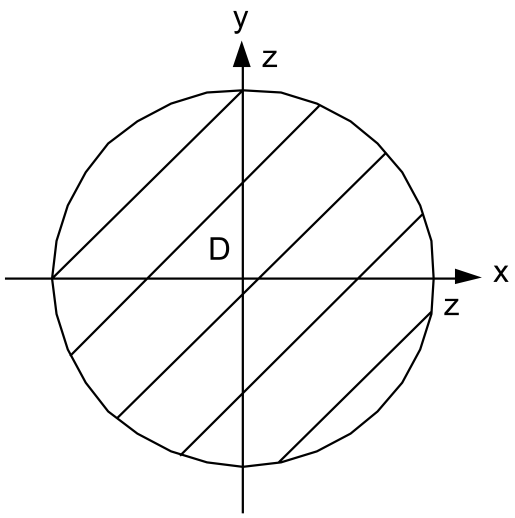
Зад.3 Да се намери плътността на разпределение на случайната величина $Z=\sqrt{X^2+Y^2}$, ако сл. величини *X* и *Y* са независими и нормално разпределени с параметри $\mu=0$ и σ,т.е $X,Y\in N(0,\sigma)$.

Решение:

Пресмятаме

$$F(z)=P(Z<z)=P(\sqrt{X^2+Y^2}<z)=\iint_D f(x,y)dxdy\,,$$

където областта *D* има вида



За да пресметнем горния интеграл полагаме $\begin{matrix}x=\rho\cos\vartheta\\ y=\rho\sin\theta\end{matrix}$.

Ако $z\geq 0$, то

$$F(z)=\frac{1}{2\pi\sigma^2}\int_0^{2\pi}d\vartheta\int_0^r e^{-\frac{\rho^2}{2\sigma^2}}\rho d\rho=\int_0^r e^{-\frac{\rho^2}{2\sigma^2}}d\,\frac{\rho^2}{2\sigma^2}$$

$$=-e^{-\frac{\rho^2}{2\sigma^2}}\Bigg|_0^r=1-e^{-\frac{r^2}{2\sigma^2}}\,.$$

Ако *z<0*, очевидно $F(z)=P(\sqrt{X^2+Y^2}<z)=0$.
Следователно плътността на *Z* е

$$f(z)=\begin{cases}\dfrac{z}{\sigma^2}e^{-\frac{z^2}{2\sigma^2}}, \; за \; z\ge 0\\ 0, \; за \; z<0\end{cases}.$$

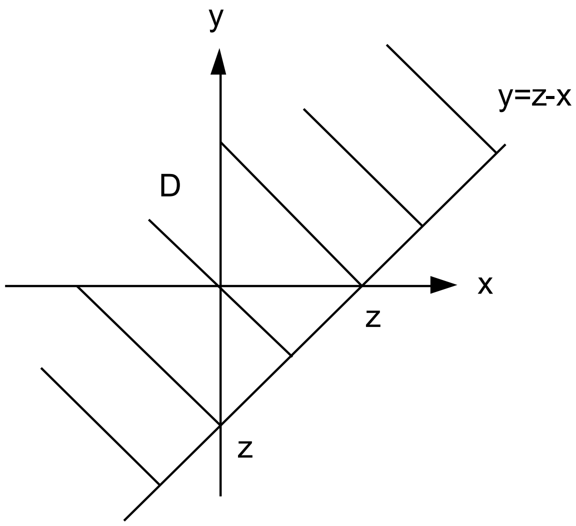
Зад.4 Да се намери плътността на разпределение на случайната величина $Z=X-Y$, ако сл. величини *X* и *Y* са независими и нормално разпределени с параметри $\mu=0$ и σ, т.е $X,Y\in N(0,\sigma)$.

Решение:

Пресмятаме

$$F(z)=P(Z<z)=P(X-Y<z)=\iint_D f(x,y)dxdy,$$

където областта *D* има вида



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

Понеже пресмятането на горния интеграл се оказва невъзможно в елементарни функции, то ще пресметнем плътността на разпределение. За целта диференцираме под знака на интеграла по параметъра *z* и получаваме

$$f(z)=F'(z)=\int_{-\infty}^{+\infty}0-f(x,x-z)(-1)dx.$$

Забележка: За да имаме право да диференцираме под знака на интеграла, то подинтегралната функция, както и граничните функции, трябва да изпълняват някои условия за регулярност. На тези условия няма да се спираме.

Подинтегралния израз се получава по следния начин.

Ако означим с $\Phi(x,y)=\int f(x,y)dy$, то вътрешният интеграл в *F(z)* придобива вида

$$\int_{x-z}^{+\infty}f(x,y)dy=\Phi(x,y)\Big|_{y=x-z}^{y=+\infty}=\Phi(x,+\infty)-\Phi(x,x-z).$$

Сега като диференцираме горния израз получаваме точно подинтегралния израз на плътността $f(z)$.

Като използваме, че *f(x,y)* е съвместната плътност на независимите сл. величини *X* и *Y,* за които $X, Y \in N(0,\sigma)$ получаваме

$$f(x,y) = \frac{1}{2\pi\sigma^2} e^{-\frac{x^2+y^2}{2\sigma^2}}.$$

Използвайки горния израз, както и израза за *f(z),* получаваме

$$f(z) = \frac{1}{2\pi\sigma^2}\int_{-\infty}^{+\infty} e^{-\frac{x^2+(x-z)^2}{2\sigma^2}} dx = \frac{1}{2\pi\sigma^2}\int_{-\infty}^{+\infty} e^{-\frac{(x-z/2)^2+z^2/4}{\sigma^2}} dx = \frac{1}{2\pi\sigma} e^{-\frac{z^2}{4\sigma^2}} \int_{-\infty}^{+\infty} e^{-\frac{(x-z/2)^2}{\sigma^2}} d(x/\sigma)$$

$$= \frac{1}{2\pi\sigma} e^{-\frac{z^2}{4\sigma^2}} \int_{-\infty}^{+\infty} e^{-t^2} dt = \frac{1}{2\pi\sigma} e^{-\frac{z^2}{4\sigma^2}} \sqrt{\pi} = \frac{1}{\sqrt{2\pi}\sqrt{2}\sigma} e^{-\frac{z^2}{2(\sqrt{2}\sigma)^2}}.$$

В горния израз сме използвали известния факт, че интеграла на Поасон е

$$\int_{-\infty}^{+\infty} e^{-t^2} dt = \sqrt{\pi}.$$

Получихме, че

$$f(z) = \frac{1}{\sqrt{2\pi}\sqrt{2}\sigma} e^{-\frac{z^2}{2(\sqrt{2}\sigma)^2}}.$$

Забележка: Горния резултат показва, че разликата на две сл. величини *X* и *Y* принадлежащи на нормалното разпределение $N(0,\sigma)$ е също нормално разпределена сл. величина *Z*, принадлежаща на $N(0,\sqrt{2}\sigma)$, т.е. дисперсиите на сл. величини *X* и *Y* се събират и се получава дисперсията на сл. величина *Z*.

Функция от случайна величина (задачи)

Зад.1 Да се намери плътността на разпределение на сл. величина даваща обема на куб, ако дължината на ръба му е сл. величина разпределена равномерно в интервала *(0,a)*.

Решение:

Нека *X* е дължината на ръба на куба. По условие *X* е разпределена равномерно в интервала *(0,a)*,т.е.

$$f_X(x) = \begin{cases} \frac{1}{a}, \ за \ x \in (0,a) \\ 0, \ за \ x \notin (0,a) \end{cases}$$

Ако *Y* е сл.величина даваща обема на куба, то $Y=X^3$.

Когато функцията $y=\varphi(x)$ свързваща случайните величини *X* и *Y* е еднозначно обратима, т.е. $x=\psi(y)$. Лесно се извежда връзката (1).

Наистина

$$F_Y(y) = P(Y<y) = P(\varphi(X)<y) = P(X<\psi(y)) = \int_{-\infty}^{\psi(y)} f_X(x)dx.$$

Диференцираме горното и получаваме

$$(1) \quad f_Y(y) = f_X(\psi(y))\,|\,\psi'(y)\,|.$$

Понеже $y=x^3$ е еднозначно обратима функция, като използваме връзката (1) получаваме

$$f_Y(y) = f_Y(\sqrt[3]{y})(\sqrt[3]{y})'.$$

Окончателно получаваме

$$f_Y(y) = \begin{cases} \frac{1}{a}(\sqrt[3]{y})' = \frac{1}{3a\sqrt[3]{y^2}}, \ за\ y \in (0, a^3) \\ 0, \ за\ y \notin (0, a^3) \end{cases}.$$

Зад.2 Да се намери плътността на разпределение на сл.величина *Y=|X|*, ако сл. величина *X* е разпределена равномерно, с параметри $\mu = 0$ и σ ,т.е. $X \in N(0, \sigma)$.

Решение:

Когато функцията $y = \varphi(x)$ свързваща случайните величини *X* и *Y* не е еднозначно обратима, т.е. имаме *n* на брой стойности за който $x_1 = \psi_1(y), x_2 = \psi_2(y), \ldots, x_n = \psi_n(x)$, то е в сила връзката

$$(2) \quad f_Y(y) = \sum_{i=1}^{n} f_X(\psi_i(y)) \, | \, \psi'_i(y) \, |.$$

Тъй като $X \in N(0, \sigma)$, то плътността й е

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

Сега използвайки (2) получаваме

$$f_Y(y) = \begin{cases} \frac{1}{\sigma\sqrt{2\pi}} e^{-\frac{y^2}{2\sigma^2}} \, | \, y' | + \frac{1}{\sigma\sqrt{2\pi}} e^{-\frac{y^2}{2\sigma^2}} \, | (-y)' |, \quad за \quad y > 0 \\ 0 \quad , \quad за \quad y \leq 0 \end{cases}$$

$$f_Y(y) = \begin{cases} \frac{2}{\sigma\sqrt{2\pi}} e^{-\frac{y^2}{2\sigma^2}} \quad , \quad за \quad y > 0 \\ 0 \quad , \quad за \quad y \leq 0 \end{cases}.$$

ТЕМА 6
Някои често използвани разпределения в статистиката – χ^2 (хи-квадрат), t, и F разпределения

1. χ2 (хи-квадрат) разпределение.

Задача 1.

Нека $\eta_1, \dots, \eta_n$ са независими, нормално разпределени случайни величини с параметри μ и σ, т.е. $\eta_i \in N(\mu, \sigma)$. Да се намери плътността на разпределение не случайната величина

$$\chi = \sqrt{\sum_{i=1}^{n} \left(\frac{\eta_i - \mu}{\sigma} \right)^2} .$$

За горната случайна величина, казваме, че е с n степени на свобода.

Решение:

Като използваме, че плътността на случайната величина χ е произведение от плътностите на случайните величини η_i, то за функцията на разпределение на χ получаваме

$$F_\chi(x) = \int\limits_{\sqrt{\sum_{i=1}^{n}\left(\frac{\eta_i-\mu}{\sigma}\right)^2}<x} \cdots \int (2\pi)^{-n/2} \sigma^{-n} e^{-\frac{1}{2}\sum_{i=1}^{n}\left(\frac{\eta_i-\mu}{\sigma}\right)^2} d\eta_1 \cdots d\eta_n .$$

Полагаме $x_i = \dfrac{\eta_i - \mu}{\sigma}$ и получаваме

$$F_\chi(x) = \int\limits_{\sqrt{\sum_{i=1}^{n} x_i^2}<x} \cdots \int (2\pi)^{-n/2} e^{-\frac{1}{2}\sum_{i=1}^{n} x_i^2} dx_1 \cdots dx_n .$$

Използваме полагане (аналогично на полагане в полярни координати за троен интеграл), т.е.

$$\left| \begin{array}{l} x_1 = \rho \cos(\theta_1) \cdots \cos(\theta_{n-2}) \cos(\theta_{n-1}) \\ x_2 = \rho \cos(\theta_1) \cdots \cos(\theta_{n-2}) \sin(\theta_{n-1}) \\ x_3 = \rho \cos(\theta_1) \cdots \cos(\theta_{n-3}) \sin(\theta_{n-2}) \\ \vdots \\ x_{n-1} = \rho \cos(\theta_1) \sin(\theta_2) \\ x_n = \rho \sin(\theta_1) \end{array} \right.$$

и като използваме теоремата за интегриране с полагане на многократен интеграл, получаваме

$$F_\chi(x) = \int_{-\pi/2}^{\pi/2} \cdots \int_{-\pi/2}^{\pi/2} \int_0^x (2\pi)^{-n/2} e^{-\rho^2/2} J(\rho, \theta_{n-1}, \cdots, \theta_1)\ d\rho d\theta_{n-1} \cdots d\theta_1 \text{, за } x > 0,$$

където $J(\rho, \theta_{n-1}, \cdots, \theta_1)$ е Якобиана на полагането, т.е.

$$J(\rho, \theta_{n-1}, \cdots, \theta_1) = \frac{\partial(x_1, \dots, x_n)}{\partial(\rho, \theta_{n-1}, \cdots, \theta_1)} = \rho^{n-1} D(\theta_{n-1}, \cdots, \theta_1)$$

Последното равенство се получава като се съобрази, че Якобиана е детерминанта от ред n, чиито редове с изключение на първия са умножени с ρ, а $D(\theta_{n-1},\cdots,\theta_1)$ е някаква тригонометрична функция, която за сега няма да търсим в явен вид.

Понеже границите на горния повторен интеграл са константи (с изключение на последното интегриране), то можем да сменим реда на интегриране и получаваме

$$F_\chi(x)=\int_0^x (2\pi)^{-n/2}e^{-\rho^2/2}\rho^{n-1}\kappa\, d\rho \text{, за } x>0,$$

където $\kappa=\int_{-\pi/2}^{\pi/2}\cdots\int_{-\pi/2}^{\pi/2} D(\theta_{n-1},\cdots,\theta_1)\ d\theta_{n-1}\cdots d\theta_1$ е някаква константа.

Може да се покаже, че

$$\kappa=\frac{(2\pi)^{n/2}}{2^{(n-2)/2}\Gamma(n/2)},$$

където $\Gamma(\kappa)=\int_0^\infty x^{k-1}e^{-x}dx$ е гама функцията.

Окончателно за функцията на разпределение на χ получаваме

$$F_\chi(x)=\begin{cases}\dfrac{1}{2^{(n-2)/2}\Gamma(n/2)}\displaystyle\int_0^x e^{-\rho^2/2}\rho^{n-1}d\rho, \ за\ x>0\\ 0,\ за\ x\le 0\end{cases}.$$

Понеже плътността на χ разпределението е $f_\chi(x)=\frac{d}{dx}\big(F_\chi(x)\big)$, тогава от горното равенство следва, че

$$f_\chi(x)=\begin{cases}\dfrac{e^{-x^2/2}x^{n-1}}{2^{(n-2)/2}\Gamma(n/2)},\ за\ x>0\\ 0,\ за\ x\le 0\end{cases}.$$

Забележка 1:

Чрез аналогични пресмятания се получава, че случайната величина

$$\frac{\chi}{\sqrt{n}}=\sqrt{\frac{1}{n}\sum_{i=1}^n\left(\frac{\eta_i-\mu}{\sigma}\right)^2}$$

има функцията на разпределение

$$F_{\chi/\sqrt{n}}(x)=\frac{1}{2^{(n-2)/2}\Gamma(n/2)}\int_0^{x\sqrt{n}} e^{-\rho^2/2}\rho^{n-1}d\rho.$$

След полагането $\rho=t\sqrt{n}$ и сменяна на границите

ρ	0	$x\sqrt{n}$
t	0	x

, получаваме

$$F_{\chi/\sqrt{n}}(x)=\frac{1}{2^{(n-2)/2}\Gamma(n/2)}\int_0^x e^{-(t\sqrt{n})^2/2}(t\sqrt{n})^{n-1}d(t\sqrt{n}).$$

От където за плътността получаваме

$$f_{\chi/\sqrt{n}}(x)=\begin{cases}\dfrac{e^{-n x^2/2}\ x^{n-1}(\sqrt{n}\,)^n}{2^{(n-2)/2}\Gamma(n/2)}=\dfrac{2\,(n/2)^{n/2}\ e^{-n x^2/2}\ x^{n-1}}{\Gamma(n/2)},\ за\ x>0\\ 0,\ за\ x\le 0\end{cases}.$$

Забележка 2:

Като се използва факта, че $\chi^2 < x$ е еквивалентно на $\chi < \sqrt{x}$ за случайната величина

$$\chi^2 = \frac{1}{n}\sum_{i=1}^{n}\left(\frac{\eta_i - \mu}{\sigma}\right)^2$$

намираме, че функцията й на разпределение е

$$F_{\chi^2}(x) = \frac{1}{2^{(n-2)/2}\Gamma(n/2)}\int_0^{\sqrt{x}} e^{-\rho^2/2}\rho^{n-1}d\rho .$$

Тази случайна величина наричаме "хи-квадрат" разпределение с n степени на свобода, записваме $\chi^2(n)$.

След полагането $\rho = \sqrt{t}$ и сменяна на границите $\begin{array}{|c|c|c|} \rho & 0 & \sqrt{x} \\ \hline t & 0 & x \end{array}$, получаваме

$$F_{\chi^2}(x) = \frac{1}{2^{(n-2)/2}\Gamma(n/2)}\int_0^{x} e^{-t/2}(\sqrt{t})^{n-1}d(\sqrt{t}) .$$

От където за плътността получаваме

$$f_{\chi^2}(x) = \begin{cases} \dfrac{e^{-x/2}x^{(n-1)/2}}{2^{(n-2)/2}\Gamma(n/2)2\sqrt{x}} = \dfrac{e^{-x/2}x^{n/2-1}}{2^{n/2}\Gamma(n/2)}, \text{за } x > 0 \\ 0, \text{за } x \le 0 \end{cases} .$$

2. F разпределение.

Задача 2.

Нека X е "хи-квадрат" разпределена случайна величина с m степени на свобода, т.е. $X \in \chi^2(m)$, а Y е "хи-квадрат" разпределена случайна величина с n степени на свобода, т.е. $Y \in \chi^2(n)$. Да се намери плътността на разпределение на случайната величина

$$Z = \frac{X}{Y} .$$

За горната случайна величина, казваме, че е с (m,n) степени на свобода.

Решение:

От предната точка следва, че плътностите на разпределение на X и Y са

$$f_X(x) = \begin{cases} \dfrac{e^{-x/2}x^{m/2-1}}{2^{m/2}\Gamma(m/2)}, \text{за } x > 0 \\ 0, \text{за } x \le 0 \end{cases} \quad \text{и} \quad f_Y(y) = \begin{cases} \dfrac{e^{-y/2}y^{n/2-1}}{2^{n/2}\Gamma(n/2)}, \text{за } y > 0 \\ 0, \text{за } y \le 0 \end{cases} .$$

Тогава от Въпрос 5 следва, че плътността на F се пресмята по формулата

$$f_Z(x) = \int_0^{\infty} y\, f_X(xy)\, f_Y(y)\, dy$$

Заместваме и получаваме

$$f_Z(x) = \frac{x^{m/2-1}}{2^{(m+n)/2}\Gamma(m/2)\Gamma(n/2)}\int_0^{\infty} y^{(m+n)/2-1} \; e^{-(xy+y)/2}\, dy$$

Полагаме $t = (xy + y)/2$ от където получаваме

$$y = \frac{2}{x+1} t\,,\quad dy = \frac{2}{x+1} dt \text{ и } y^{(m+n)/2-1} = \left(\frac{2}{x+1}\right)^{(m+n)/2-1} t^{(m+n)/2-1}\,.$$

Тогава

$$f_Z(x) = \frac{x^{m/2-1}}{2^{(m+n)/2}\Gamma(m/2)\Gamma(n/2)} \int_0^\infty \left(\frac{2}{x+1}\right)^{(m+n)/2-1} t^{(m+n)/2-1}\, e^{-t}\, \frac{2}{x+1} dt$$

$$f_Z(x) = \frac{x^{m/2-1}}{\Gamma(m/2)\Gamma(n/2)(x+1)^{(m+n)/2}} \int_0^\infty t^{(m+n)/2-1}\, e^{-t}\, dt$$

$$f_Z(x) = \frac{x^{m/2-1}}{\Gamma(m/2)\Gamma(n/2)(x+1)^{(m+n)/2}} \Gamma\big((m+n)/2\big).$$

Забележка 3:

Чрез пресмятания аналогични на тези от първа точка могат да се пресметнат плътностите на X/m и Y/n. След това с пресмятания аналогични на тези от втора точка може да се пресметне плътността на

$$F = \frac{X/m}{Y/n}\,.$$

Тази плътност е

$$f_F(x) = \begin{cases} \dfrac{m^{m/2}\, n^{n/2}\, x^{m/2-1}\, \Gamma\big((m+n)/2\big)}{\Gamma(m/2)\Gamma(n/2)(mx+n)^{(m+n)/2}}, \text{за } x > 0 \\ 0, \text{за } x \le 0 \end{cases}.$$

Тази случайна величина наричаме F разпределение с (m,n) степени на свобода.

2. t разпределение.

Задача 3.

Нека X е нормално разпределени случайни величини с параметри $\mu = 0$ и $\sigma = 1$, т.е. $X \in N(0,1)$, а Y е "хи-квадрат" разпределена случайна величина с n степени на свобода, т.е. $Y \in \chi^2(n)$. Да се намери плътността на разпределение не случайната величина

$$T = \frac{X}{\sqrt{Y/n}}\,.$$

Тази случайна величина наричаме t разпределение с n степени на свобода.

Решение:

От предния и от този въпрос знаем, че плътностите на разпределение на X и $Z = \sqrt{Y/n}$ са

$$f_X(x) = \frac{1}{\sqrt{2\pi}} e^{-x^2/2}, \text{за} -\infty < x < +\infty.$$

$$f_Z(z) = \begin{cases} \dfrac{2\,(n/2)^{n/2}\, e^{-n\,z^2/2}\, z^{n-1}}{\Gamma(n/2)}, \text{за } z > 0 \\ 0, \text{за } z \le 0 \end{cases}.$$

Тогава от Въпрос 5 следва, че плътността на F се пресмята по формулата

$$f_T(x) = \int_0^\infty z\, f_X(xz)\, f_Y(z)\, dz$$

Заместваме и получаваме

$$f_T(x) = \frac{2\,(n/2)^{n/2}}{\sqrt{2\pi}\;\Gamma(n/2)} \int\limits_0^{\infty} z^{n-1}\; e^{-(x^2 z^2 + n\,z^2)/2}\; zdz$$

Полагаме $t = \dfrac{x^2 + n}{2}\, z^2$ от където получаваме

$$dt = \frac{x^2 + n}{2} 2zdz\,,\;\; \frac{1}{x^2 + n} dt = zdz\,,$$

$$t\frac{2}{x^2 + n} = z^2\,,\;\; t^{1/2}\left(\frac{2}{x^2 + n}\right)^{1/2} = z \text{ и } t^{(n-1)/2}\left(\frac{2}{x^2 + n}\right)^{(n-1)/2} = z^{n-1}.$$

Тогава

$$f_T(x) = \frac{2\,(n/2)^{n/2}}{\sqrt{2\pi}\;\Gamma(n/2)} \int\limits_0^{\infty} t^{(n-1)/2}\left(\frac{2}{x^2 + n}\right)^{(n-1)/2} e^{-t}\; \frac{1}{x^2 + n} dt$$

$$f_T(x) = \frac{2\,(n/2)^{n/2}\, 2^{(n-1)/2}}{\Gamma(n/2)\big(n(x^2/n + 1)\big)^{(n+1)/2}} \int\limits_0^{\infty} t^{(n-1)/2}\; e^{-t}\; dt$$

$$f_T(x) = \frac{1}{\sqrt{\pi\, n}\;\Gamma(n/2)(x^2/n + 1)^{(n+1)/2}} \Gamma\big((n+1)/2\big),\;\; \text{за} -\infty < x < +\infty.$$

ТЕМА 7
Неравенство на Чебишев. Закон за големите числа. Централна гранична теорема.

1. Неравенство на Чебишев.

Т1: Ако случайната величина X има математическо очакване EX и дисперсия DX, то

$$P(|X-EX|\geq\varepsilon)\leq\frac{DX}{\varepsilon^2}\text{, за } \forall\varepsilon>0.$$

Доказателство:

$$P(|X-EX|\geq\varepsilon)=\int\limits_{x\in(EX-\varepsilon,EX+\varepsilon)}f(x)dx\leq\frac{1}{\varepsilon^2}\int\limits_{x\in(EX-\varepsilon,EX+\varepsilon)}(x-EX)^2f(x)dx\leq$$

$$\leq\frac{1}{\varepsilon^2}\int\limits_{-\infty}^{+\infty}(x-EX)^2f(x)dx=\frac{DX}{\varepsilon^2}$$

2. Закон за големите числа.

Т2: Ако $X_1,\ldots,X_n$ независими случайни величини имащи крайни дисперсии (т.е. $D(X_k)\leq C$ за $\forall k$), тогава

$$\lim_{n\to\infty}P(|\tfrac{1}{n}\sum_{k=1}^{n}X_k-\tfrac{1}{n}\sum_{k=1}^{n}E(X_k)|<\varepsilon)=1\text{, за } \forall\varepsilon>0.$$

Доказателство:

Ако за случайната величина $Z=\frac{1}{n}\sum_{k=1}^{n}X_k$ приложим неравенството на Чебишов, получаваме

$$P(|\tfrac{1}{n}\sum_{k=1}^{n}X_k-E\left(\tfrac{1}{n}\sum_{k=1}^{n}X_k\right)|\geq\varepsilon)\geq\frac{D\left(\frac{1}{n}\sum_{k=1}^{n}X_k\right)}{\varepsilon^2}=\frac{\frac{1}{n^2}\sum_{k=1}^{n}D(X_k)}{\varepsilon^2}\leq\frac{\frac{1}{n^2}nC}{\varepsilon^2}=\frac{C}{n\,\varepsilon^2}\underset{n\to\infty}{\longrightarrow}0$$

От където следва, че

$$P(|\tfrac{1}{n}\sum_{k=1}^{n}X_k-E\left(\tfrac{1}{n}\sum_{k=1}^{n}X_k\right)|<\varepsilon)\underset{n\to\infty}{\longrightarrow}1$$

Т3: Ако $X_1,\ldots,X_n$ случайни величини за които $\lim\limits_{n\to\infty}\left(D\left(\frac{1}{n}\sum_{k=1}^{n}X_k\right)\right)=0$, тогава

$$\lim_{n\to\infty}P(|\tfrac{1}{n}\sum_{k=1}^{n}X_k-\tfrac{1}{n}\sum_{k=1}^{n}E(X_k)|<\varepsilon)=1\text{, за } \forall\varepsilon>0.$$

Доказателство:(аналогично на горното)

Забележка 1: Ако е изпълнено

$$P(\lim_{n\to\infty}\left|\tfrac{1}{n}\sum_{k=1}^{n}X_k-\tfrac{1}{n}\sum_{k=1}^{n}E(X_k)\right|=0)=1$$

то горното се нарича усилен закон за големите числа.

3. Централна гранична теорема.

Т4: Ако $X_1, \ldots, X_n$ независими и еднакво разпределени случайни величини имащи крайни дисперсии (т.е. $D(X_k) \le C$ за $\forall k$), тогава

$$\lim_{n\to\infty} P\left(\alpha < \left| \frac{\frac{1}{n}\sum_{k=1}^{n} X_k - E\left(\frac{1}{n}\sum_{k=1}^{n} X_k\right)}{\sqrt{D\left(\frac{1}{n}\sum_{k=1}^{n} X_k\right)}} \right| < \beta \right) = \frac{1}{\sqrt{2\pi}} \int_{\alpha}^{\beta} e^{-x^2/2} dx$$

Доказателство:(няма да бъде дадено)

Т5: (теорема на Моавър-Лаплас)

Ако X е биномно разпределена случайна величини с параметри n и p /т.е. $X \in Bi(n, p)$/, тогава

$$\lim_{n\to\infty} P\left(\alpha < \left| \frac{m - np}{\sqrt{npq}} \right| < \beta \right) = \frac{1}{\sqrt{2\pi}} \int_{\alpha}^{\beta} e^{-x^2/2} dx$$

Доказателство:

Ще припомним , че случайната величина X е биномно разпределена случайна величини с параметри n и p, ако X е броя на "успехите" от n опита, където вероятността за успех при един опит е p и вероятността на събитието $X = k$ / $k = 0,1,\ldots,n$ / се пресмята по формулата на Бернули, т.е.

$$P(X = k) = C_n^k p^k (1-p)^{n-k} .$$

Ще въведем случайните величини $X_к$ е броя на "успехите" в $к$-тия пореден опит /очевидно $X_к \equiv \{0, 1\}$/, следователно

X_k	0	1
P(X_k)	1-p	p

Пресмятаме

$$E(X_k) = p$$

$$D(X_k) = E(X_k^2) - \left(E(X_k)\right)^2 = p - p^2 = p(1-p) .$$

Очевидно, че

$$X = X_1 + \ldots + X_n .$$

Пресмятаме

$$E(X) = E(\sum_{k=1}^{n} X_k) = \sum_{k=1}^{n} E(X_k) = np$$

$$D(X) = D(\sum_{k=1}^{n} X_k) = \sum_{k=1}^{n} D(X_k) = np(1-p) .$$

Заместваме горните три равенства в Т4 и получаваме Т5.

Неравенство на Чебишев. Закони за големите числа (задачи)

Зад.1 Да се оцени вероятността една сл. величина да се отклони от математическото си очакване на повече от три средно квадратични отклонения.

Решение:

Търсим $P(| X - EX | \ge 3\sigma)$.

Използвайки неравенството на Чебишев

$$P(|X - EX| \geq \varepsilon) \leq \frac{DX}{\varepsilon^2}, \, за \, \forall \varepsilon > 0,$$

получаваме

$$P(|X - EX| \geq 3\sigma) \leq \frac{DX}{(3\sigma)^2} = \frac{1}{9}.$$

Зад.2 Да се докаже, че

$$P(X \geq \varepsilon) \leq \frac{E(e^{ax})}{e^{a\varepsilon}}, \, за \, \forall \varepsilon > 0 \, и \, a > 0.$$

Решение:

$$P(X \geq \varepsilon) = \int\limits_{x>0} f(x)dx = \int\limits_{x>0} \frac{e^{ax}}{e^{ax}} f(x)dx \leq \frac{1}{e^{a\varepsilon}} \int\limits_{x>0} e^{ax} f(x)dx \leq \frac{1}{e^{a\varepsilon}} \int\limits_{-\infty}^{+\infty} e^{ax} f(x)dx = \frac{E(e^{ax})}{e^{a\varepsilon}}.$$

Забележка: В тази задача, функцията e^{ax}, може да се замести с произволна растяща функция, чийто минимална стойност в областта на интегриране е по-голяма от нула. На аналогични разсъждения се базира и доказателството на оригиналното неравенството на Чебишев.

Зад.3 Да се определи дали е в сила закона за големите числа за редицата от независими сл. величини $\{X_k\}_{k=1}^{+\infty}$, всяка от който има ред на разпределение

X_k	$\sqrt{2}$	0	$\sqrt{2}$
$P(X_k)$	$\frac{1}{4}$	$\frac{1}{2}$	$\frac{1}{2}$

Решение:

Пресмятаме

$$D(X_k) = E(X_k^2) - (EX_k)^2 = (2\frac{1}{4} + 0 + 2\frac{1}{4}) - (\sqrt{2}\frac{1}{4} + 0 + \sqrt{2}\frac{1}{4})^2 = 1.$$

Следователно горните сл.величини имат ограничени дисперсии. Като се използва този резултат лесно може да се докаже, че за такава редица от сл. величини е в сила (слабия) закон за големите числа.

Наистина използваме неравенството на Чебишев получаваме

(1) $$P\left(\left|\frac{1}{n}\sum_{k=1}^{n} X_k - \frac{1}{n}\sum_{k=1}^{n} EX_k\right| \geq \varepsilon\right) \leq \frac{D(\frac{1}{n}\sum_{k=1}^{n} X_k)}{\varepsilon^2} = \frac{\frac{1}{n^2}\sum_{k=1}^{n} D(X_k)}{\varepsilon^2} = \frac{\frac{1}{n^2} n}{\varepsilon^2} = \frac{1}{n\varepsilon^2}.$$

В горния израз пускаме $n \to +\infty$ и получаваме

$$\lim_{n \to +\infty} P\left(\left|\frac{1}{n}\sum_{k=1}^{n} X_k - \frac{1}{n}\sum_{k=1}^{n} EX_k\right| \geq \varepsilon\right) = 0, \, за \, \forall \varepsilon > 0.$$

Забележка: Горния израз е един от резултати наречени слаб закон за големите числа. Ако читателя е запознат с теория на апроксимациите, това е точно сходимост по вероятностна мярка.

Зад.4 Да се определи дали е в сила закона за големите числа за редица от независими сл. величини $\{X_k\}_{k=1}^{+\infty}$, ако:

а) $P(X_k = \pm 2^k) = \dfrac{1}{2}$;

б) $P(X_k = \pm 2^k) = 2^{-(2k+1)}$ и $P(X_k = 0) = 1 - 2^{-2k}$;

в) $P(X_k = \pm k) = \dfrac{1}{2k}$ и $P(X_k = 0) = 1 - \dfrac{1}{k}$;

Решение:

а) Пресмятаме

$$D(X_k) = E(X_k^2) - (EX_k)^2 = (2^{2k}\frac{1}{4} + 2^{2k}\frac{1}{4}) - (2^k\frac{1}{4} - 2^k\frac{1}{4})^2 = 4^k.$$

Пресмятаме

$$D(\frac{1}{n}\sum_{k=1}^{n} X_k) = \frac{1}{n^2}\sum_{k=1}^{n} DX_k = \frac{1}{n^2}\sum_{k=1}^{n} 4^k = \frac{1}{n^2}4\frac{4^n-1}{4-1} = \frac{4}{3}\frac{4^n-1}{n^2}.$$

Правим граничния преход

$$\lim_{n\to+\infty} D(\frac{1}{n}\sum_{k=1}^{n} X_k) = \frac{4}{3}\lim_{n\to+\infty}\frac{4^n-1}{n^2} = 0.$$

Като използваме горния резултат, от израза (1) получаваме (2), т.е слабия закон за големите числа е в сила.

б) Пресмятаме

$$D(X_k) = E(X_k^2) - (EX_k)^2 = (2^{2k}\frac{1}{2^{2k+1}} + 2^{2k}\frac{1}{2^{2k+1}}) - (2^k\frac{1}{2^{2k+1}} - 2^k\frac{1}{2^{2k+1}})^2 = 1.$$

Получихме, че сл.величини имат ограничени дисперсии. Както показахме в зад.3 от този факт директно следва, че за тази редица от сл. величини е в сила (слабия) закон за големите числа.

в) Пресмятаме

$$D(X_k) = E(X_k^2) - (EX_k)^2 = (k^2\frac{1}{2k} + 0 + k^2\frac{1}{2k}) - (k\frac{1}{2k} + 0 - k\frac{1}{2k})^2 = k.$$

Пресмятаме

$$D(\frac{1}{n}\sum_{k=1}^{n} X_k) = \frac{1}{n^2}\sum_{k=1}^{n} DX_k = \frac{1}{n^2}\sum_{k=1}^{n} k = \frac{1}{n^2}\frac{n(n+1)}{2}.$$

Правим граничния преход

$$\lim_{n\to+\infty} D(\frac{1}{n}\sum_{k=1}^{n} X_k) = \lim_{n\to+\infty}\frac{n(n+1)}{2n^2} = \frac{1}{2}.$$

Като използваме горния резултат, от израза (1) получаваме

$$\lim_{n\to+\infty} P(\left|\frac{1}{n}\sum_{k=1}^{n} X_k - \frac{1}{n}\sum_{k=1}^{n} EX_k\right| \geq \varepsilon) \leq \frac{1}{2\varepsilon^2}.$$

От този резултат не могат да се направят изводи дали е в сила слабия закон за големите числа!

Зад.5 Да се намери такова *s* за което да е в сила закона за големите числа за редица от независими сл. величини $\{X_k\}_{k=1}^{+\infty}$, ако $P(X_k = \pm k^s) = \dfrac{1}{2}$.

Решение:

Пресмятаме

$$D(X_k) = E(X_k^2) - (EX_k)^2 = (k^{2s}\frac{1}{2} + k^{2s}\frac{1}{2}) - (k^s\frac{1}{2} - k^s\frac{1}{2})^2 = k^{2s}.$$

Пресмятаме

$$\lim_{n\to+\infty} D(\frac{1}{n}\sum_{k=1}^{n} X_k) = \lim_{n\to+\infty}\frac{1}{n}\sum_{k=1}^{n} DX_k = \lim_{n\to+\infty}\frac{1}{n}\sum_{k=1}^{n} k^{2s} < \lim_{n\to+\infty}\frac{1}{n}\sum_{k=1}^{n} n^{2s} = \lim_{n\to+\infty} n^{2s-1}.$$

Ако в горния израз наложим условието *2s-1*<0,т.е $s < \frac{1}{2}$, получаваме

$$\lim_{n\to+\infty} D(\frac{1}{n}\sum_{k=1}^{n} X_k) = \lim_{n\to+\infty} n^{2s-1} = 0.$$

Както показахме в зад.4б от горния факт директно следва, че за тази редица от сл. величини е в сила (слабия) закон за големите числа.

Централна гранична теорема

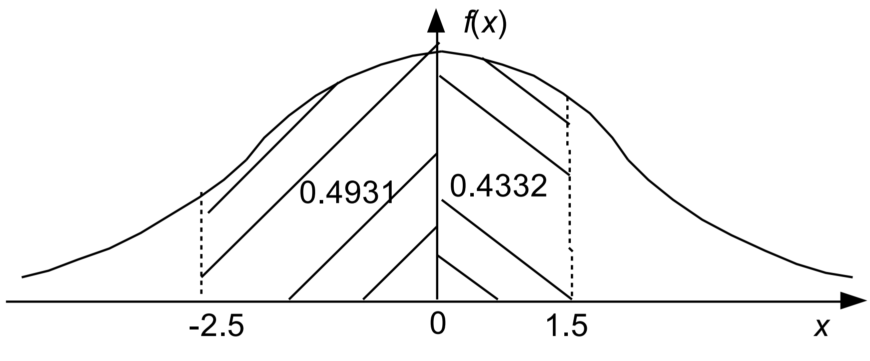
Зад.1 100 струга с еднаква мощност работят независимо един от друг. Захранването на всеки от тях е включено 80% от цялото време. Каква е вероятността в произволно взет момент от време да се окажат включени от 70 до 86 струга?

Решение:

Очевидно броя на включените струга "m" е биномно разпределена сл. величина с параметри $n = 100$ и $p = 0.8$, т.е. $Bi(n = 100, p = 0.8)$, тогава

$$P(70 \le m \le 86) = P\left(\frac{70 - np}{\sqrt{npq}} \le \frac{m - np}{\sqrt{npq}} \le \frac{86 - np}{\sqrt{npq}}\right) = P\left(\frac{70 - 80}{4} \le \frac{m - np}{\sqrt{npq}} \le \frac{86 - 80}{4}\right) =$$

$$= P(-2{,}5 \le \frac{m - np}{\sqrt{npq}} \le 1.5) = 0.4332 + 0.4931 = 0.927$$



Фигура 1

Зад.2 Вероятността на събитието *А* се определя по метода Монте-Карло. Колко минимално независими изпитания трябва да се направят, така че с вероятност 0.99 да се определи вероятността на събитието *А*, с грешка по-малка от 0.01 ?

Решение:

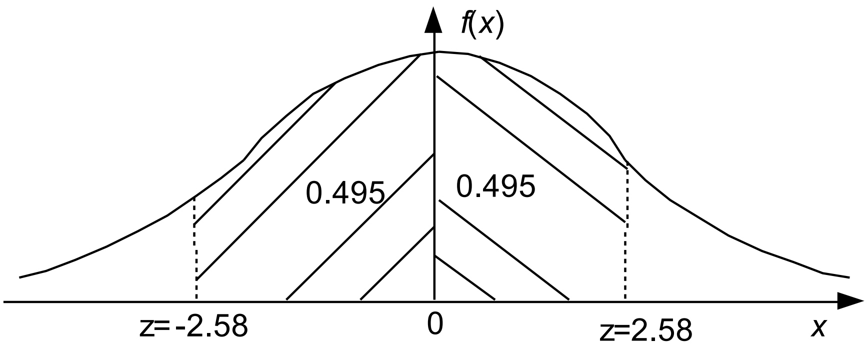
Очевидно броя на сбъдванията "m" на събитието *А*, е биномно разпределена сл. величина, за която е известно, че

$$P\left(\left|\frac{m}{n} - p\right| < 0.01\right) \ge 0.99$$

$$P(-0.01 \le \frac{m}{n} - p \le 0.01) \ge 0.99$$

$$P(-0.01\frac{n}{\sqrt{npq}} \le \frac{m-np}{\sqrt{npq}} - p \le 0.01\frac{n}{\sqrt{npq}}) \ge 0.99$$

$P(-z \le \dfrac{m-np}{\sqrt{npq}} - p \le z) \ge 0.99$, където $z = 0.01\dfrac{n}{\sqrt{npq}}$



Фигура 2

$$z = 0.01\frac{n}{\sqrt{npq}} \ge 2.58$$

$$\sqrt{n} \ge 258\sqrt{pq}$$

$n \ge 258^2 pq$, но $\max(pq) = \dfrac{1}{4}$

$$n \ge 258^2 \frac{1}{4} = 16641$$

Зад.3 Вероятността произволно избран детайл да се окаже дефектен е 0.1. Партидата не се приема ако се открият най-малко 10 дефектни изделия. Колко минимално детайла трябва да се проверят, така че с вероятност 0.6 да се твърди, че партидата няма да бъде приета ?

Решение:

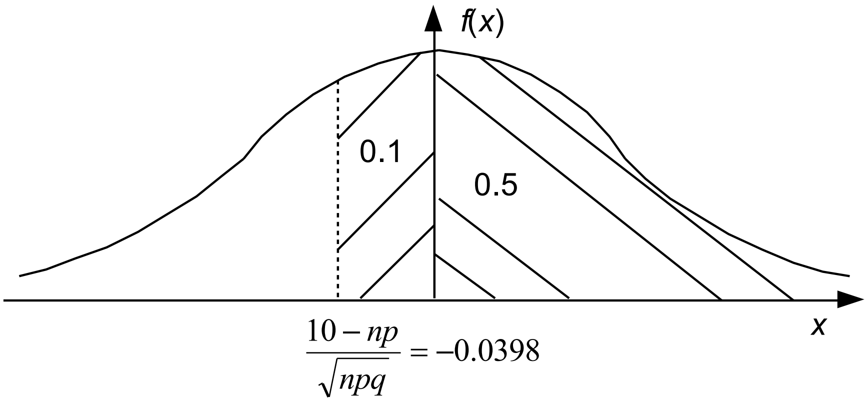
Очевидно броя "m"на проверените и оказали се дефектни изделия, е биномно разпределена сл. величина, за която е известно следното

$P(10 \le m \le n.) \ge 0.6$, $p = 0.1$ и $q = 1 - p = 0.9$

$$P\left(\frac{10-np}{\sqrt{npq}} \le \frac{m-np}{\sqrt{npq}} \le \frac{n-np}{\sqrt{npq}}\right) \ge 0.6$$

$P\left(\dfrac{10-np}{\sqrt{npq}} \le \dfrac{m-np}{\sqrt{npq}} \le 3\sqrt{n}\right) \ge 0.6$, понеже $3\sqrt{n} \ge 5$ то можем да считаме, че

$$P\left(\frac{10-np}{\sqrt{npq}} \le \frac{m-np}{\sqrt{npq}} \le +\infty\right) \ge 0.6$$



Фигура 3

$$\frac{10-np}{\sqrt{npq}}=-0.0398$$

$$100-n=-0.0398\sqrt{n}$$

$$n-0.12\sqrt{n}-100=0$$

$$\sqrt{n}=\frac{0.12\mp\sqrt{0.12^2+400}}{2}$$

$\sqrt{n}=10.1$ и $\sqrt{n}=9.9$

$$n\geq 10.1^2=102$$

Зад.4 Интегралът $J=\int_0^1 x^2dx$ се пресмята приблизително по метода Монте-Карло. Определете вероятността при 1000 изпитания, грешката от пресмятането му да не е по-голяма от 0.01 ?

Решение:

Ако приемем, че стойността на аргумента е равномерно разпределена сл. величина , т.е. $X\in U(0,1)$, то приближената стойност на интеграла се пресмята по формулата

$$J_н=\frac{1}{n}\sum_{\kappa=1}^{n}X_k^2$$

Тогава

$$E(J_н)=\frac{1}{n}\sum_{\kappa=1}^{n}E(X_k^2)=E(X^2)=J$$

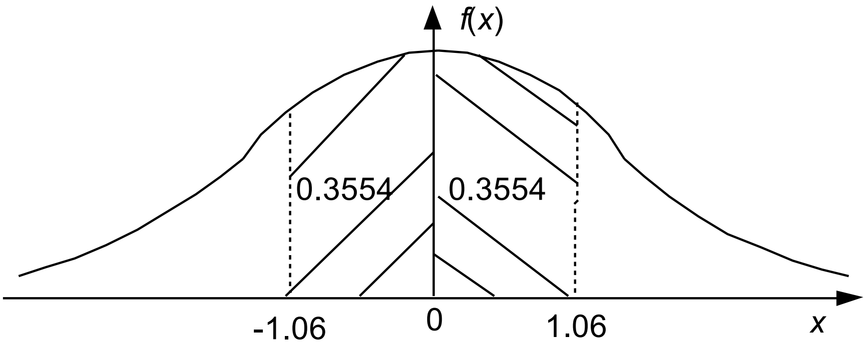
$$D(J_н)=\frac{1}{n^2}\sum_{\kappa=1}^{n}D(X_k^2)=\frac{1}{n}D(X^2)=\frac{1}{n}\Big(E(X^4)-(E(X^2))^2\Big)=$$

$$=\frac{1}{n}\left(\int_0^1 x^4dx-\left(\int_0^1 x^2dx\right)^2\right)=\frac{1}{n}\left(\frac{1}{5}-\left(\frac{1}{3}\right)^2\right)=\frac{4}{n\,45}=\frac{4}{45000}$$

Тогава $P\left(-\varepsilon\leq\frac{J_n-J}{\sqrt{\frac{4}{45000}}}\leq\varepsilon\right)\in N(0,1)$. Избираме $\varepsilon=\frac{0.01}{\sqrt{\frac{4}{45000}}}$ и получаваме

$$P\left(-\frac{0.01}{\sqrt{\frac{4}{45000}}} \le \frac{J_n - J}{\sqrt{\frac{4}{45000}}} \le \frac{0.01}{\sqrt{\frac{4}{45000}}}\right) \in N(0,1)$$

$$P\left(-1.06 \le \frac{J_n - J}{\sqrt{\frac{4}{45000}}} \le 1.06\right) = 2 \quad 0.3554 = 0.7108$$



Фигура 4

Глава 2 ПРИЛОЖНА СТАТИСТИКА
ТЕМА 1
Описателна статистика

В този въпрос ще бъдат намерени: извадковото средно - $\bar{x}$, извадковата дисперсия– s^2, медианата – *Me*, квартилите – Q_1, Q_3 и модата – *Mo*, на различни извадки от данни

Дефинициите на тези понятия ще бъдат дадени в процеса на излагане на материала.

Зад.1 (количествени /числови/ данни)

Тества се поведението на 20 души по време на шофиране, след употреба на разрешеното количество алкохол. Всеки от участниците в експеримента може да получи максимум 100 точки. Участниците са получили следните точки

74, 55, 82, 61, 68, 69, 61, 84, 52, 93

74, 79, 65, 88, 53, 57, 64, 67, 72, 77.

А) Да се определи извадковото средно - $\bar{x}$, извадковата дисперсия– s^2, медианата – *Me*, квартилите – Q_1, Q_3 и модата – *Mo*(за негрупираните данни).

Б) Данните да се групират в 5 групи и да се определи извадковото средно - $\bar{x}$, извадковата дисперсия– s^2, медианата – *Me*, квартилите – Q_1, Q_3 и модален интервал – *Mo* (за групираните данни).

Решение:

А) Негрупирана извадка

$\overline{X} = \frac{1}{n}\sum_{i=1}^{n} x_i$ - извадково средно (средната аритметична стойност)

$$\overline{X} = \frac{74+55+82+61+68+69+61+84+52+93+74+79+65+88+53+57+64+67+72+77}{20} = 70.25$$

$s^2 = \frac{1}{n-1}\sum_{i=1}^{n}(x_i - \overline{X})^2 = \frac{n}{n-1}\left(\overline{X^2} - \left(\overline{X}\right)^2\right)$ - извадкова дисперсия,

където $\overline{X^2} = \frac{1}{n}\sum_{i=1}^{n} x_i^2$

$$s^2 = \frac{(74-70.25)^2 + (55-70.25)^2 + (82-70.25)^2 + \ldots + (67-70.25)^2 + (72-70.25)^2 + (77-70.25)^2 +}{19} = 124.3$$

Съставяме вариационният ред (данните подредени във възходящ ред)

Номер	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20
Данна	52	55	57	61	61	63	64	65	67	68	69	72	74	74	77	79	82	84	88	93

Медианата – *Me* е още и 50% процентил, това е такава данна във вариационният ред, за която 50% от данните са преди нея и 50% от данните са след нея. Когато данните са четен брой, очевидно такава данна не съществува, за това се осредняват две съседни данни, като е в случая

$$Me = \frac{68+69}{2} = 68.5$$

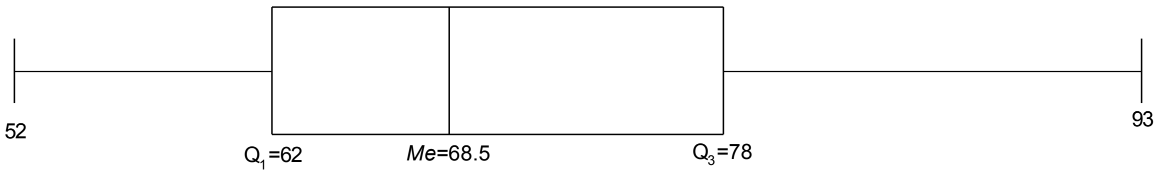
Първият квартил – Q_1 е още и 25% процентил, това е такава данна във вариационният ред, за която 25% от данните са преди нея и 75% от данните са след нея. Когато една четвърт от данните са четен брой, очевидно такава данна не съществува, за това се осредняват две съседни данни, като е в случая

$$Q_1 = \frac{61+63}{2} = 62$$

Третият квартил – Q_3 е още и 75% процентил, това е такава данна във вариационният ред, за която 75% от данните са преди нея и 25% от данните са след нея. Когато една четвърт от данните са четен брой, очевидно такава данна не съществува, за това се осредняват две съседни данни, като е в случая

$$Q_3 = \frac{77+79}{2} = 78$$

Много често вариационният ред се онагледява с така наречената "кутия с мустаци", фигура 1



Фигура 1

Модата – *Mo* е най-често повтарящата се данна, в случая имаме две моди

$Mo = 61$ и $Mo = 74$

Б) Групирана извадка

Групираме данните в 5 еднакво дълги интервала и получаваме

Интервал	50-60	60-70	70-80	80-90	90-100
m_i	3	8	5	3	1
F_i	3	3+8=11	11+5=16	16+3=19	19+1=20

където

m_i - се наричат честоти на попадение в *i*-тия интервал (това са броя на данните попадащи в този интервал);

F_i - се наричат кумулативни(натрупани) честоти на попадение в *i*-тия интервал.

Забележка: Когато една данна се оказва на границата на два интервала, обикновено тя се включва към по-горния интервал. Но има автори, които приемат че във всеки от двата съседни интервала има по 0.5 броя данна.

$\overline{X} = \frac{1}{n}\sum_{i=1}^{5}\overline{x}_i\ m_i$ - извадково средно, където $\overline{x}_i$ са средите на интервалите, а 5 е броя на интервалите на групиране

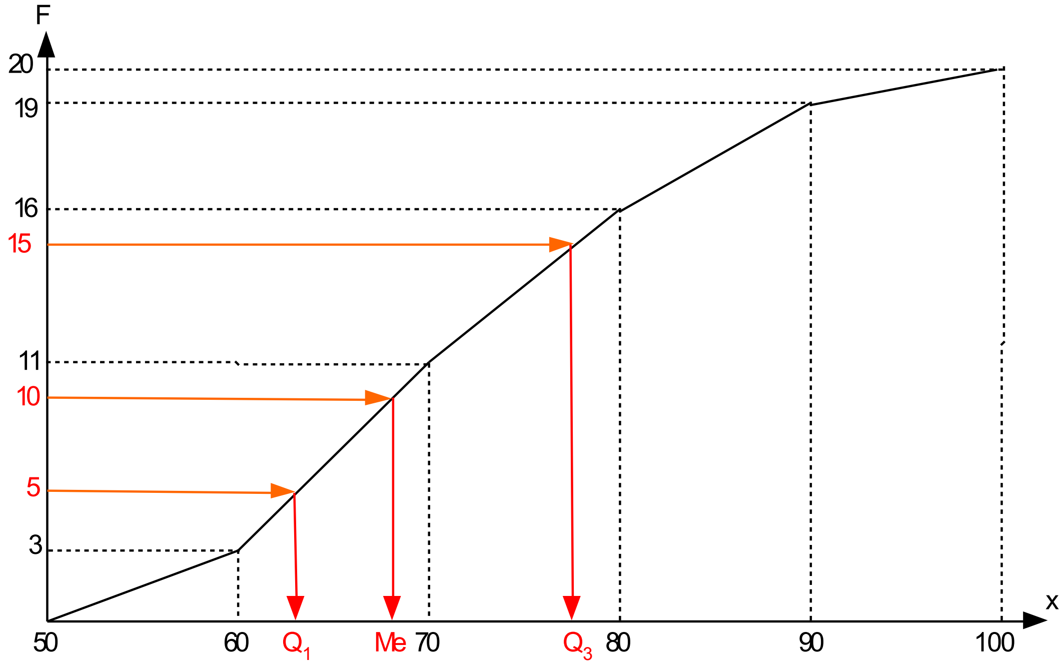
$$\overline{X} = \frac{55.3+65.8+75.5+85.3+95.1}{20} = 70.5$$

$s^2 = \frac{1}{n-1}\sum_{i=1}^{5}(x_i - \overline{X})^2 m_i = \frac{n}{n-1}\left(\overline{X^2} - \left(\overline{X}\right)^2\right)$ - извадкова дисперсия,

където $\overline{X^2} = \frac{1}{n}\sum_{i=1}^{5}x_i^2 m_i$

$$s^2 = \frac{(55-70.5)^2 3 + (65-70.5)^2 8 + (75-70.5)^2 5 + (85-70.5)^2 3 + (95-70.5)^2}{19} = 120.8$$

За да пресметнем медианата – *Me* и квартилите – Q_1, Q_3 ще начертаем графиката на кумулативните честоти, фигура 2



Фигура 2

Медианата – *Me* е тази стойност на *x,* за която $F(x)=\frac{1}{4}F\max=\frac{1}{4}20=5$. В случая, като използваме подобни триъгълници получаваме, че

$$Me=60+10\frac{7}{8}=68.75$$

Първият квартил – Q_1 е тази стойност на *x,* за която $F(x)=\frac{1}{2}F\max=\frac{1}{2}20=10$. В случая, като използваме подобни триъгълници получаваме, че

$$Q_1=60+10\frac{2}{8}=62.5$$

Третият квартил – Q_3 е тази стойност на *x,* за която $F(x)=\frac{3}{4}F\max=\frac{3}{4}20=15$. В случая, като използваме подобни триъгълници получаваме, че

$$Q_3=70+10\frac{4}{5}=78$$

Забележка: Забелязва се, че стойностите на *Me,* Q_1 и Q_3 за негрупирани данни и за групирани данни леко се различават. Това е съвсем естествено, понеже след групирането на едни данни очевидно се губи някаква информация. След групирането знаем колко данни попадат в даден интервал, но не и какви са точните им стойности. Вижда се че в случая тези разлики са съвсем минимални, което е косвен показател, че избраното групиране е добро. Няма строго фиксирани правила за групиране на данни, но най-често се стремим тези интервали да са с равни дължини и във всеки интервал да няма прекалено малко данни. За малка извадка от данни често не могат да се спазят всички емпирични правила за групиране, като е и в случая. За това малки извадки обикновено не се групират.

За групирани данни не говорим за модата – *Mo* , а използваме понятието модален интервал. Това е интервала с най-голямо m_i (най-голяма честота на попадение). В случая

Mo интервал е [60,70].

Зад.2 (качествени ординарни /имащи подредба/ данни)

На въпроса "Бихте ли купили това ново изделие на нашата фирма?", са получени следните отговори

Номер - i	Отговор	Честоти - m_i	Натрупани честоти - F_i
1	не	13	13
2	по-скоро не	30	43
3	по-скоро да	35	78
4	да	22	100

Забележка: Очевидно в този случай не могат да се намерят всички числови характеристики от задача първа, т.е. $\bar{x}$, s^2, *Me*, Q_1, Q_3 и *Mo*. За това се търсят само такива характеристики, които се явяват аналог на характеристиките от задача 1 и имат смисъл. Това са медианен, квартилен и модален класове (от данни).

Решение:

Медианен клас е интервалът на F_i съдържащ данната $\frac{1}{2}n = 50$, т.е.

Медианен клас = "по-скоро да"

Аналогично

Първи квартилен клас = "по-скоро не"

Трети квартилен клас = "по-скоро да"

Модален клас е интервалът с най-голямо m_i, т.е.

Модален клас = "по-скоро да".

Зад.3 (качествени номинални /нямащи подредба/ данни)

На въпроса "В кои от следните цветове бихте предпочели да си купили най-харесвания от вас модел?", са получени следните отговори

Номер - i	Отговор	Честоти - m_i
1	кафяв	65
2	син	75
3	розов	30

Забележка: В случая, можем да търсим само модален клас.

Решение:

Модален клас е интервалът с най-голямо m_i, т.е.

Модален клас = "син".

ТЕМА 2
Точкови оценки за параметрите на някои разпределения

Има различни методи за съставяне на точкови оценки, но ние ще разгледаме само - метода на неизместените оценки.

Нестрого(интуитивно) въвеждане на понятието неизместена оценка. Нека от една случайната величина, направим извадка от краен брой наблюдения (примерно *n* на брой). Ако разликата между средното аритметично на наблюденията и средното аритметично от всички възможни стойности на случайната величина клони към нула с нарастването на *n* , то казваме че средното аритметично на извадката е неизместена оценка за средно аритметичното на случайната величина.

Сега ще въведем и строгото правило, с което ще търсим неизместени оценки.

Т: Ако случайните величини $X_1, X_2, \cdots, X_n$ са независими и имат същото разпределение като случайната величина X , тогава:

А) $\overline{X} = \frac{1}{n}\sum_{i=1}^{n} X_i$ е неизместена оценка за EX ;

В) $s^2 = \frac{1}{n-1}\sum_{i=1}^{n}(X_i - \overline{X})^2$ е неизместена оценка за DX .

Понеже ще търсим неизместените оценки на параметрите на: нормалното, поасоновото и експоненциалното разпределения, то ще си припомним тези разпределения.

Поасоново разпределени - дискретната случайна величина X е Поасоново разпределена с параметър λ $/ X \in P_o(\lambda) /$, ако

$$P(X = k) = \frac{\lambda^k}{k!} e^{-\lambda} \text{ за } k = 0,1,\ldots,\infty .$$

Това разпределение има:

- математическо очакване $EX = \lambda$;
- дисперсия $DX = \lambda$.

Нормално разпределение - непрекъснатата случайна величина X е равномерно разпределена с параметри μ и σ $/ X \in N(\mu, \sigma) /$, ако плътността й е

$$f(x) = \frac{1}{\sigma\sqrt{2\pi}} e^{-\frac{1}{2}\left(\frac{x-\mu}{\sigma}\right)^2}, \text{ за } -\infty < x < +\infty .$$

Това разпределение има:

- математическо очакване $EX = \mu$;
- дисперсия $DX = \sigma^2$.

Експоненциално разпределени - дискретната случайна величина X е експоненциално разпределена с параметър λ $/ X \in Exp(\lambda) /$, ако

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

Това разпределение има:

- математическо очакване $EX = 1/\lambda$;
- дисперсия $DX = 1/\lambda^2$.

Зад.1

Отклоненията от стандартните размери на 100 детайла са нанесени в следната таблица

№	1	2	3	4	5	6	7	8	9	10
Интер-вал	-20;-15	-15;-10	-10;-5	-5;0	0;5	5;10	10;15	15;20	20;25	25;30
Среда	-17.5	-12.5	-7.5	-2.5	2.5	7.5	12.5	17.5	22.5	27,5
Брой	4	5	7	12	25	20	13	9	4	1

Ако предположим, че данните в таблицата са разпределени нормално с параметри μ и σ, да се намерят неизместените оценки на тези параметри.

Решение:

За групирана извадка имаме

$\overline{X} = \frac{1}{n}\sum_{i=1}^{k}\overline{x}_i\ m_i$ - извадково средно,

където:

mᵢ - се наричат честоти на попадение в *i*-тия интервал (това са броя на данните попадащи в този интервал);

Fᵢ - се наричат кумулативни(натрупани) честоти на попадение в *i*-тия интервал

k –брой интервали на групиране

Пресмятаме

$$\overline{X} = \frac{1}{n}\sum_{i=1}^{10}\overline{x}_i\ m_i = \frac{-15.5\ 4 - 12.5\ 5 - 7.5\ 7 - 2.5\ 12 + 2.5\ 25 + 7.5\ 20 + 12.5\ 13 + 17.5\ 9 + 22.5\ 5 + 27.5}{100} = 4.35$$

$$\overline{X^2} = \frac{1}{n}\sum_{i=1}^{10}\overline{x}_i^2\ m_i = \frac{15.5^2 4 + 12.5^2 5 + 7.5^2 7 + 2.5^2 12 + 2.5^2 25 + 7.5^2 20 + 12.5^2 13 + 17.5^2 9 + 22.5^2 5 + 27.5^2}{100} = 113.2$$

$$s^2 = \frac{n}{n-1}\left(\overline{X^2} - \left(\overline{X}\right)^2\right) = \frac{100}{99}(113.25 - 4.35^2) = 95.28$$

Тогава по метода на неизместените оценки имаме

$$\mu = EX = \overline{X} = 4.35$$

$$\sigma = \sqrt{DX} = s = \sqrt{95.28} = 9.76$$

Зад.2 Наблюдава са работата на 50 независими прибора по времето на гаранционния им срок.

Резултатите са нанесени в таблицата

Брой откази	Брой прибори дали толкова отказа
0	34
1	12
2	4

Ако предположим, че данните в таблицата са разпределени Поасоново с параметър λ, да се намери неизместената оценка на този параметър.

Решение:

Пресмятаме

$$\overline{X} = \frac{1}{n}\sum_{i=1}^{10}\overline{x}_i\ m_i = \frac{0.34 + 1.12 + 2.4}{50} = 0.4$$

Тогава по метода на неизместените оценки имаме

$$\lambda = EX = \overline{X} = 0.4$$

Зад.3 Времената за безотказна работа(в часове) на 6 прибора са

100, 170, 400, 250, 520, 680

Ако предположим, че данните в таблицата са разпределени експононенциално с параметър λ, да се намери неизместената оценка на този параметър.

Решение:

Пресмятаме

$$\overline{X} = \frac{1}{n}\sum_{i=1}^{n} x_i \quad = \frac{100+170+400+250+520+680}{6} = 353.3$$

Тогава по метода на неизместените оценки имаме

$$\frac{1}{\lambda} = EX = \overline{X} = 353.3$$

или

$$\lambda = \frac{1}{353.3}$$

ТЕМА 3
Определяне обема на извадката

Нека случайните величини $X_1, X_2, \cdots, X_n$ са независими и еднакво разпределени с $E(X_i) = \mu$ и $D(X_i) = \sigma^2$.

Образуваме случайните величини

$\overline{X} = \frac{1}{n}\sum_{i=1}^{n} X_i$ и $s^2 = \frac{1}{n-1}\sum_{i=1}^{n}(X_i - \overline{X})^2$.

Тогава за достатъчно голямо n /обикновено $n \geq 30$/ случайната величина $\frac{\overline{X}-\mu}{\sigma/\sqrt{n}}$ е стандартно нормално разпределена, т.е. $\frac{\overline{X}-\mu}{\sigma/\sqrt{n}} \in N(0,1)$

Забележка:

За да определим обема на извадката, че въведем понятието грешка $e = |\overline{X} - \mu|$. Ако грешката е известна /т.е. избрана съгласно желанието на клиента/ и ако знаем σ, то използвайки $\frac{\overline{X}-\mu}{\sigma/\sqrt{n}} \in N(0,1)$ можем да определим n.

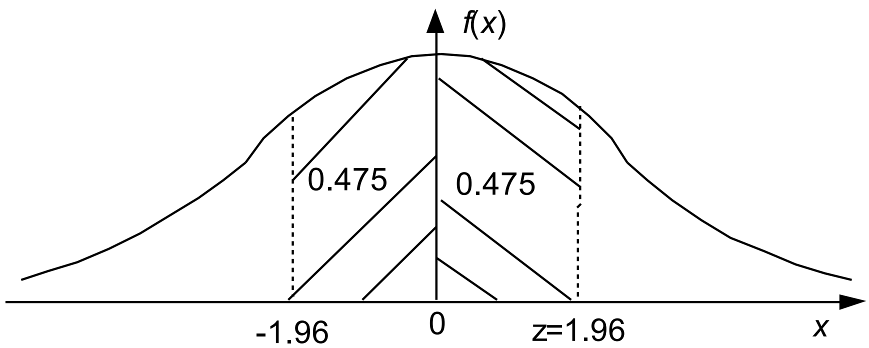
Зад.1 Собственик иска да определи със сигурност 95% при какъв минимален обем на извадката - n, максимално допустимата грешка между сумата похарчена от клиента и средната такава е 4 лв. Собственикът от собствен опит знае, че стандартното отклонение от похарчената от клиента сума е 9 лв.

Решение:

Знаем $|\overline{X} - \mu| = 4$ и $\sigma = 9$

Искаме

$P\left(\frac{|\overline{X}-\mu|}{\sigma/\sqrt{n}} < z\right) = 0.95$ или $P\left(-z < \frac{\overline{X}-\mu}{\sigma/\sqrt{n}} < z\right) = 0.95$ или $P\left(-z < \frac{4}{9/\sqrt{n}} < z\right) = 0.95$



Фигура 1

$\frac{4}{9/\sqrt{n}} = 1.96$ или $n = \left(\frac{9}{4}1.96\right)^2 \approx 19.45$

Следователно $n = 20$.

ТЕМА 4
Интервални оценки и проверка на параметрични хипотези

Въведение

Някой по-важни разпределения, свързани с интервални оценки и проверка на параметрични хипотези.

Означаваме $\overline{X} = \frac{1}{n_x}\sum_{i=1}^{n_X} X_i$, $s_x^2 = \frac{1}{n_x - 1}\sum_{i=1}^{n_X}(X_i - \overline{X})^2$.

Ако $X_i \in N(\mu, \sigma)$, тогава

а) за $n_x < 30$ и σ известно, то $\frac{\overline{X} - \mu}{\sigma / \sqrt{n_x}} \in N(0,1)$.

б) за $n_x < 30$ и σ неизвестно, то $\frac{\overline{X} - \mu}{s_x / \sqrt{n_x}} \in t(n-1)$.

в) за μ неизвестно, то $\frac{(n_x - 1){s_x}^2}{\sigma^2} \in \chi^2(n-1)$.

г) за μ известно, то $\frac{n_x \quad {s_1}^2}{\sigma^2} \in \chi^2(n)$, където $s_1^2 = \frac{1}{n_x}\sum_{i=1}^{n_x}(X_i - \mu)^2$.

Забележка: Ако $n_x \geq 30$, тогава е достатъчно X_i да са независими и еднакво разпределени с $E(X_i) = \mu$ и $D(X_i) = \sigma^2$ и тогава

$$\frac{\overline{X} - \mu}{s_x / \sqrt{n_x}} \in N(0,1).$$

Забележка: Припомняме, че:

а) $N(0,1)$ - стандартно нормално(Гаусово) разпределение;

б) $t(n)$ - *t*- разпределение(Стюдънтово разпределение) с *n* степени на свобода;

в) $\chi^2(n)$ - χ^2- разпределение с *n* степени на свобода.

Зад.1

Резултатите от четири независими измервания на разстоянието до даден обект са

2220, 2250, 2260 и 2270 m.

Средната грешка на прибора с който се мери това разстояние е 40 м.

Със сигурност:

а) 95% намерете интервала, в който варира измерваното разстояние;

б) 90% проверете основната хипотезата $H_0 : \mu = 2230$, срещу алтернативата $H_1 : \mu \neq 2230$.

Решение:

Имаме $\overline{X} = \frac{2220 + 2250 + 2260 + 2270}{4} = 2250$, $\sigma = 40$ и $n = 4$, тогава $\frac{\overline{X} - \mu}{\sigma / \sqrt{n}} \in N(0,1)$.

а) $P\left(-z < \frac{\overline{X} - \mu}{\sigma / \sqrt{n}} < z\right) = 0.95$

$$P\left(-z < \frac{2250-\mu}{40/\sqrt{4}} < z\right) = 0.95$$

$$P\left(-z < \frac{2250-\mu}{40/\sqrt{4}} < 0\right) + P\left(0 < \frac{2250-\mu}{40/\sqrt{4}} < z\right) = 0.95$$

$$2\,P\left(0 < \frac{2250-\mu}{40/\sqrt{4}} < z\right) = 0.95$$

$$P\left(0 < \frac{2250-\mu}{40/\sqrt{4}} < z\right) = 0.475$$

От таблицата за $N(0,1)$ намираме, че $z = 1.96$, тогава

$$-1.96 < \frac{2250-\mu}{40/\sqrt{4}} < 1.96$$

$$2250 - 1.96\frac{40}{\sqrt{4}} < \mu < 2250 + 1.96\frac{40}{\sqrt{4}}$$

$$2211 < \mu < 2289.$$

б) $P\left(-z < \frac{\overline{X}-\mu}{\sigma/\sqrt{n}} < z\right) = 0.9$

$$P\left(-z < \frac{2250-\mu}{40/\sqrt{4}} < z\right) = 0.9$$

$$P\left(-z < \frac{2250-\mu}{40/\sqrt{4}} < 0\right) + P\left(0 < \frac{2250-\mu}{40/\sqrt{4}} < z\right) = 0.9$$

$$2\,P\left(0 < \frac{2250-\mu}{40/\sqrt{4}} < z\right) = 0.9$$

$$P\left(0 < \frac{2250-\mu}{40/\sqrt{4}} < z\right) = 0.45$$

От таблицата за $N(0,1)$ намираме, че $z = 1.65$, тогава

$$-1.65 < \frac{2250-\mu}{40/\sqrt{4}} < 1.65$$

Заместваме μ със стойността му съгласно основната хипотеза, т.е. $\mu = 2230$ и получаваме

$$-1.65 < \frac{2250-2230}{40/\sqrt{4}} < 1.65$$

$-1.65 < 1 < 1.65$ - вярно

Понеже последното неравенство е вярно, то приемаме основната хипотеза, т.е. $\mu = 2230$.

Зад.2

Резултатите от 22 независими наблюдения на нормална случайна величина са нанесени в следната таблица

x_i - стойност	114	115	116	117	118
m_i – честота на наблюдение	2	5	8	4	3

Със сигурност:

а) 90% намерете доверителния интервал за μ;

б) 95% намерете доверителния интервал за σ;

в) 99% проверете основната хипотезата $H_0 : \mu = 113$, срещу алтернативата $H_1 : \mu < 113$;

г) 95% проверете основната хипотезата $H_0 : \sigma = 1.1$, срещу алтернативата $H_1 : \sigma > 1.1$.

Решение:

Имаме $\overline{X} = \dfrac{2.114 + 5.115 + 8.116 + 4.117 + 3.118}{22} \approx 116$,

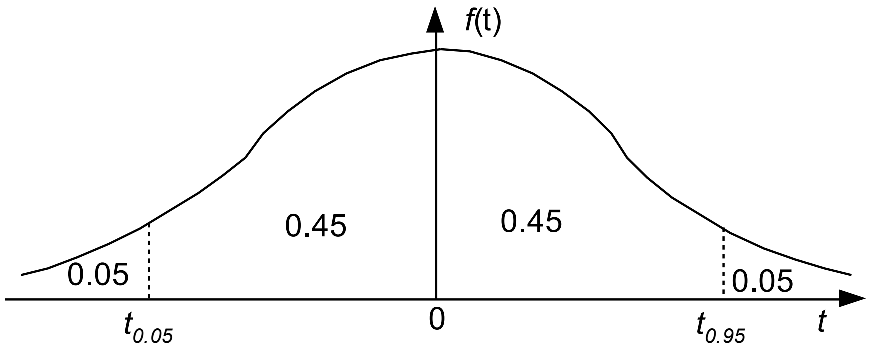
$$s^2 = \frac{2.(114-116)^2 + 5.(115-116)^2 + 8.(116-116)^2 + 4.(117-116)^2 + 3.(118-116)^2}{21} \approx 1.38$$

$s \approx 1.175$ и $n = 22 < 30$.

а) $\dfrac{\overline{X} - \mu}{s/\sqrt{n}} \in t(n-1)$

$$P\left(t_{0.05} < \frac{\overline{X} - \mu}{s/\sqrt{n}} < t_{0.95} \right) = 0.9$$

$$P\left(-t_{0.95} < \frac{\overline{X} - \mu}{s/\sqrt{n}} < t_{0.95} \right) = 0.9$$



Фигура 1

На фигура 1 долният индекс-α на стойността-t_α, показва площта под плътността $f(t)$ наляво от t_α. Стойността t_α се нарича α квантил на t разпределението.

От таблицата за $t(n)$ намираме, че $t_{0.95}(21) = 1.721$, тогава

$$-1.721 < \frac{116 - \mu}{1.175/\sqrt{22}} < 1.721$$

$$116 - 1.721\frac{1.175}{\sqrt{22}} < \mu < 116 + 1.721\frac{1.175}{\sqrt{22}}$$

$115.6 < \mu < 116.4.$

б) $\dfrac{(n-1)s^2}{\sigma^2} \in (n-1)$

$$P\left(\chi^2{}_{0.025} < \frac{(n-1)s^2}{\sigma^2} < \chi^2{}_{0.975} \right) = 0.95$$

От таблицата за $\chi^2(n)$ намираме, че $\chi^2{}_{0.025}(21) = 10.28$ и $\chi^2{}_{0.975}(21) = 35.48$, тогава

$$10.28 < \frac{(n-1)s^2}{\sigma^2} < 35.48$$

$$\frac{21 \quad 1.38}{35.48} < \sigma^2 < \frac{21 \quad 1.38}{10.28}$$

$$0.817 < \sigma^2 < 2.821$$

в) $\dfrac{\overline{X} - \mu}{s/\sqrt{n}} \in t(n-1)$

$P\left(t_{0.01} < \dfrac{\overline{X} - \mu}{s/\sqrt{n}} \right) = 0.99$ - едностранна хипотеза

Използваме, че $t_{0.01}(21) = -t_{0.99}(21)$ и от таблицата за $t(n)$ намираме, че $t_{0.01}(21) = -2.581$, тогава

$$-2.581 < \frac{116 - \mu}{1.38/\sqrt{21}}$$

Заместваме μ със стойността му съгласно основната хипотеза, т.е. $\mu = 113$ и получаваме

$-2.581 < \dfrac{116 - 113}{1.38/\sqrt{21}} = 0.54$ - вярно

Понеже последното неравенство е вярно, то приемаме основната хипотеза, т.е. $\mu = 113$

г) $\dfrac{(n-1)s^2}{\sigma^2} \in (n-1)$

$P\left(\dfrac{(n-1)s^2}{\sigma^2} < \chi^2{}_{0.95} \right) = 0.95$ - едностранна хипотеза

От таблицата за $\chi^2(n)$ намираме, че $\chi^2{}_{0.95}(21) = 32.67$, тогава

$$\frac{(n-1)s^2}{\sigma^2} < 32.67$$

Заместваме σ със стойността му съгласно основната хипотеза, т.е. $\sigma = 1.1$ и получаваме

$\dfrac{(n-1)s^2}{\sigma^2} = \dfrac{21 \quad 1.38}{1.1^2} = 23.95 < 32.67$ - вярно

Понеже последното неравенство е вярно, то приемаме основната хипотеза, т.е. $\sigma = 1.1$.

Зад.3

Социалното министерство поръчало с 95% вероятност да се провери в какви граници варира средно месечната сметка за ток на едно домакинство от двама пенсионери. За целта са наблюдавани 81 домакинства и е установено, че плащат 31.5±9 лева.

Решение:

Имаме $\overline{X} = 31.5$, $s = 9$ и $n = 81 > 30$, тогава $\dfrac{\overline{X} - \mu}{s/\sqrt{n}} \in N(0,1)$

$$P\left(-z < \frac{\overline{X} - \mu}{s/\sqrt{n}} < z \right) = 0.95$$

$$P\left(-z < \frac{31.5-\mu}{9/\sqrt{81}} < 0\right) + P\left(0 < \frac{31.5-\mu}{9/\sqrt{81}} < z\right) = 0.95$$

$$2\,P\left(0 < \frac{31.5-\mu}{9/\sqrt{81}} < z\right) = 0.95$$

$$P\left(0 < \frac{31.5-\mu}{9/\sqrt{81}} < z\right) = 0.475$$

От таблицата за $N(0,1)$ намираме, че $z = 1.96$, тогава

$$-1.96 < \frac{31.5-\mu}{9/\sqrt{81}} < 1.96$$

$$31.5 - 1.96\frac{9}{\sqrt{81}} < \mu < 31.5 + 1.96\frac{9}{\sqrt{81}}$$

$$29.54 < \mu < 33.46.$$

ТЕМА 5
Проверка на хипотези за разликата между две средни

Въведение

Някой по-важни разпределения, свързани с проверка на параметрични хипотези.

Означаваме $\overline{X} = \frac{1}{n_x}\sum_{i=1}^{n_x} X_i$, $\overline{Y} = \frac{1}{n_y}\sum_{j=1}^{n_y} Y_j$, $s_x^2 = \frac{1}{n_x - 1}\sum_{i=1}^{n_x}(X_i - \overline{X})^2$ и $s_y^2 = \frac{1}{n_y - 1}\sum_{j=1}^{n_y}(Y_j - \overline{Y})^2$.

Ако $X_i \in N(\mu_x, \sigma_x)$ и $Y_j \in N(\mu_y, \sigma_y)$, тогава:

а) за $n_x, n_y < 30$ и σ_x, σ_y известни, то $\dfrac{(\overline{X} - \overline{Y}) - (\mu_x - \mu_y)}{\sqrt{\dfrac{\sigma_x^2}{n_x} + \dfrac{\sigma_y^2}{n_y}}} \in N(0,1)$.

б) за $n_x, n_y < 30$ и σ_x, σ_y неизвестни и $\sigma_x = \sigma_y$, то

$$\frac{(\overline{X} - \overline{Y}) - (\mu_x - \mu_y)}{s_p\sqrt{\dfrac{1}{n_x} + \dfrac{1}{n_y}}} \in t(n_x + n_y - 2), \text{ където } s_p = \sqrt{\frac{(n_x - 1)s_x^2 + (n_y - 1)s_y^2}{n_x + n_y - 2}}.$$

в) за $n_x, n_y < 30$ и σ_x, σ_y неизвестни и $\sigma_x \neq \sigma_y$, то

$$\frac{(\overline{X} - \overline{Y}) - (\mu_x - \mu_y)}{\sqrt{\dfrac{s_x^2}{n_x} + \dfrac{s_y^2}{n_y}}} \in t(\nu), \text{ където } \nu \text{ е цялата част на } \frac{\left(s_x^2 / n_x + s_y^2 / n_y\right)^2}{\dfrac{(s_x^2 / n_x)^2}{n_x - 1} + \dfrac{(s_y^2 / n_y)^2}{n_y - 1}}.$$

г) $\dfrac{s_x^2}{s_y^2} \in F(n_x - 1, n_y - 1)$.

Забележка: Ако $n_x, n_y \geq 30$, тогава е достатъчно X_i и Y_j да са независими и еднакво разпределени с $E(X_i) = \mu_x$, $D(X_i) = {\sigma_x}^2$, $E(Y_i) = \mu_y$, $D(Y_i) = {\sigma_y}^2$ и тогава

$$\frac{(\overline{X} - \overline{Y}) - (\mu_x - \mu_y)}{\sqrt{\dfrac{s_x^2}{n_x} + \dfrac{s_y^2}{n_y}}} \in N(0,1).$$

Забележка: Припомняме, че:

а) $N(0,1)$ - стандартно нормално(Гаусово) разпределение;

б) $t(n)$ - *t*- разпределение(Стюдънтово разпределение) с *n* степени на свобода;

в) $\chi^2(n)$ - χ^2- разпределение с *n* степени на свобода.

г) $F(m,n)$- *F* разпределение с (m,n) степени на свобода.

Зад.1

В два еднотипни хотела собствениците са направили извадки с обем 35 и 45, за заплатите на служителите. Резултатите от извадката са $\overline{X} = 132$, $s_x = 12$, $\overline{Y} = 123$ и

$s_y = 17$. С вероятност 95% да се провери хипотезата, че заплатите в двата хотела са равни.

Решение:

Проверяваме, основна хипотеза $\mu_x - \mu_y = 0$, срещу алтернативата $\mu_x - \mu_y \neq 0$.

Понеже $n_x, n_y \geq 30$, тогава $\dfrac{(\overline{X} - \overline{Y}) - (\mu_x - \mu_y)}{\sqrt{\dfrac{s_x^2}{n_x} + \dfrac{s_y^2}{n_y}}} \in N(0,1)$

$$P\left(-z < \frac{(\overline{X} - \overline{Y}) - (\mu_x - \mu_y)}{\sqrt{\dfrac{s_x^2}{n_x} + \dfrac{s_y^2}{n_y}}} < z \right) = 0.95$$

$$2\,P\left(0 < \frac{(\overline{X} - \overline{Y}) - (\mu_x - \mu_y)}{\sqrt{\dfrac{s_x^2}{n_x} + \dfrac{s_y^2}{n_y}}} < z \right) = 0.95$$

$$P\left(0 < \frac{(\overline{X} - \overline{Y}) - (\mu_x - \mu_y)}{\sqrt{\dfrac{s_x^2}{n_x} + \dfrac{s_y^2}{n_y}}} < z \right) = 0.475$$

От таблицата за $N(0,1)$ намираме, че $z = 1.96$, тогава

$$-1.96 < \frac{(\overline{X} - \overline{Y}) - (\mu_x - \mu_y)}{\sqrt{\dfrac{s_x^2}{n_x} + \dfrac{s_y^2}{n_y}}} < 1.96$$

Заместваме основната хипотеза и получаваме

$$-1.96 < \frac{(132 - 123) - 0}{\sqrt{\dfrac{12^2}{35} + \dfrac{17^2}{45}}} < 1.96$$

$$-1.96 < 2.773 < 1.96$$

Понеже последното неравенство не е вярно, то отхвърляме основната хипотеза и приемаме алтернативата, т.е. приемам че $\mu_x - \mu_y \neq 0$.

Зад.2

С ниво на доверие $\alpha = 0.1$ да се провери основната хипотезата $H_0 : \mu_x - \mu_y = 0$, срещу алтернативата $H_1 : \mu_x - \mu_y \neq 0$. При следните извадкови данни $n_x = 8$, $\overline{X} = 100$, $s_x = 15$, $n_y = 10$, $\overline{Y} = 116$ и $s_y = 10$, получени от нормално разпределена генералната съвкупност.

Решение:

Понеже $n_x, n_y < 30$, първо проверяваме: хипотезата $\sigma_x = \sigma_y$, срещу алтернативата $\sigma_x \neq \sigma_y$. За целта използваме разпределението $\frac{s_x^2}{s_y^2} \in F(n_x - 1, n_y - 1)$.

$$P\left(F_{0.05} < \frac{s_x^2}{s_y^2} < F_{0.95} \right) = 0.9$$

$$F_{0.05}(7,9) < \frac{s_x^2}{s_y^2} < F_{0.95}(7,9)$$

Използваме свойствата на *F* разпределението и получаваме

$$\frac{1}{F_{0.95}(9,7)} < \frac{s_x^2}{s_y^2} < F_{0.95}(7,9)$$

От таблицата за *F* разпределението намираме, че $F_{0.95}(7,9) = 3.29$ и $F_{0.95}(9,7) = 3.68$, тогава

$$\frac{1}{3.68} < \frac{s_x^2}{s_y^2} < 3.29$$

Заместваме данните от извадката и получаваме

$$\frac{1}{3.29} < \frac{15^2}{10^2} < 3.68$$

$$0.3 < 2.25 < 3.68$$

Понеже последното неравенство е вярно, то приемаме основната хипотеза $\sigma_x = \sigma_y$.

Като отчетем полученото $\sigma_x = \sigma_y$, вече можем да пристъпим, към проверка на хипотеза $\mu_x - \mu_y = 0$, срещу алтернативата $\mu_x - \mu_y \neq 0$. Тогава

$$\frac{(\overline{X} - \overline{Y}) - (\mu_x - \mu_y)}{s_p \sqrt{\frac{1}{n_x} + \frac{1}{n_y}}} \in t(n_x + n_y - 2), \text{ където } s_p = \sqrt{\frac{(n_x - 1)s_x^2 + (n_y - 1)s_y^2}{n_x + n_y - 2}}.$$

$$P\left(t_{0.05} < \frac{(\overline{X} - \overline{Y}) - (\mu_x - \mu_y)}{s_p \sqrt{\frac{1}{n_x} + \frac{1}{n_y}}} < t_{0.95} \right) = 0.9$$

$$P\left(-t_{0.95} < \frac{(\overline{X} - \overline{Y}) - (\mu_x - \mu_y)}{s_p \sqrt{\frac{1}{n_x} + \frac{1}{n_y}}} < t_{0.95} \right) = 0.9$$

$$-t_{0.95}(16) < \frac{(\overline{X} - \overline{Y}) - (\mu_x - \mu_y)}{s_p \sqrt{\frac{1}{n_x} + \frac{1}{n_y}}} < t_{0.95}(16)$$

От таблицата за *t* разпределението намираме, че $t_{0.95}(16)=1.746$, тогава

$$-1.746<\frac{(\overline{X}-\overline{Y})-(\mu_x-\mu_y)}{s_p\sqrt{\dfrac{1}{n_x}+\dfrac{1}{n_y}}}<1.746$$

Заместваме данните от извадката и получаваме

$$-1.746<\frac{(100-116)-0}{s_p\sqrt{\dfrac{1}{8}+\dfrac{1}{10}}}<1.746,\ \text{за}\ s_p=\sqrt{\frac{7\ \ 225+9\ \ 100}{16}}=7.5$$

$$-1.746<-1.96<1.746$$

Понеже последното неравенство е вярно, то приемаме основната хипотеза $\mu_x-\mu_y=0$.

Зад.3

С ниво на доверие $\alpha=0.1$ да се провери основната хипотезата $H_0:\mu_x-\mu_y=0$, срещу алтернативата $H_1:\mu_x-\mu_y\neq 0$. При следните извадкови данни $n_x=13$, $\overline{X}=70$, $s_x^2=90$, $n_y=8$, $\overline{Y}=80$ и $s_y^2=23$, получени от нормално разпределена генералната съвкупност.

Решение:

Понеже $n_x,n_y<30$, първо проверяваме: хипотезата $\sigma_x=\sigma_y$, срещу алтернативата $\sigma_x\neq\sigma_y$. За целта използваме разпределението $\dfrac{s_x^2}{s_y^2}\in F(n_x-1,n_y-1)$.

$$P\left(F_{0.05}<\frac{s_x^2}{s_y^2}<F_{0.95}\right)=0.9$$

$$F_{0.05}(12,7)<\frac{s_x^2}{s_y^2}<F_{0.95}(12,7)$$

Използваме свойствата на *F* разпределението и получаваме

$$\frac{1}{F_{0.95}(7,12)}<\frac{s_x^2}{s_y^2}<F_{0.95}(7,9)$$

От таблицата за *F* разпределението намираме, че $F_{0.95}(7,12)=2.92$ и $F_{0.95}(12,7)=3.57$, тогава

$$\frac{1}{2.92}<\frac{s_x^2}{s_y^2}<3.57$$

Заместваме данните от извадката и получаваме

$$\frac{1}{2.92}<\frac{90}{23}<3.57$$

$$0.35<3.91<3.57$$

Понеже последното неравенство е вярно, то отхвърляме основната хипотеза $\sigma_x=\sigma_y$ и приемаме алтернативата, т.е. приемам че $\sigma_x\neq\sigma_y$.

Като отчетем полученото $\sigma_x \neq \sigma_y$, вече можем да пристъпим, към проверка на хипотеза $\mu_x - \mu_y = 0$, срещу алтернативата $\mu_x - \mu_y \neq 0$. Тогава

$\dfrac{(\overline{X}-\overline{Y})-(\mu_x-\mu_y)}{\sqrt{\dfrac{s_x^2}{n_x}+\dfrac{s_y^2}{n_y}}} \in t(\nu)$, където ν е цялата част на $\dfrac{\left(s_x^2/n_x + s_y^2/n_y\right)^2}{\dfrac{(s_x^2/n_x)^2}{n_x-1}+\dfrac{(s_y^2/n_y)^2}{n_y-1}}$.

$$P\left(t_{0.05} < \frac{(\overline{X}-\overline{Y})-(\mu_x-\mu_y)}{\sqrt{\frac{s_x^2}{n_x}+\frac{s_y^2}{n_y}}} < t_{0.95} \right) = 0.9$$

$$P\left(-t_{0.95} < \frac{(\overline{X}-\overline{Y})-(\mu_x-\mu_y)}{\sqrt{\frac{s_x^2}{n_x}+\frac{s_y^2}{n_y}}} < t_{0.95} \right) = 0.9$$

$-t_{0.95}(\nu') < \dfrac{(\overline{X}-\overline{Y})-(\mu_x-\mu_y)}{\sqrt{\dfrac{s_x^2}{n_x}+\dfrac{s_y^2}{n_y}}} < t_{0.95}(\nu')$, където ν' е цялата част на

$$\frac{\left(s_x^2/n_x + s_y^2/n_y\right)^2}{\frac{(s_x^2/n_x)^2}{n_x-1}+\frac{(s_y^2/n_y)^2}{n_y-1}} = \frac{(90/13+23/8)^2}{\frac{(90/13)^2}{12}+\frac{(23/8)^2}{7}} = 18.5$$

$$-t_{0.95}(18) < \frac{(\overline{X}-\overline{Y})-(\mu_x-\mu_y)}{\sqrt{\frac{s_x^2}{n_x}+\frac{s_y^2}{n_y}}} < t_{0.95}(18)$$

От таблицата за *t* разпределението намираме, че $t_{0.95}(18) = 1.734$, тогава

$$-1.734 < \frac{(\overline{X}-\overline{Y})-(\mu_x-\mu_y)}{\sqrt{\frac{s_x^2}{n_x}+\frac{s_y^2}{n_y}}} < 1.734$$

Заместваме данните от извадката и получаваме

$$-1.734 < \frac{(70-80)-0}{\sqrt{\frac{90}{13}+\frac{23}{8}}} < 1.734$$

$$-1.734 < -3.195 < 1.734$$

Понеже последното неравенство е вярно, то отхвърляме основната хипотеза $\mu_x - \mu_y = 0$ и приемаме алтернативата, т.е. приемам че $\mu_x - \mu_y \neq 0$.

ТЕМА 6
Критерии на знаците на Фишер

С този критерий се проверява, дали две извадки с еднакъв обем са еднородни (имат едно и също разпределение).

По-точно, при дадени извадки

$x_1, x_2, \ldots, x_n$

$y_1, y_2, \ldots, y_n$

с ниво на доверие α, се проверява основна хипотеза H_0 срещу алтернатива H_1, където:

H_0 : *извадките имат еднаква функция на разпределение*

H_1 : *извадките нямат еднаква функция на разпределение*.

При критерия на знаците се съставят разликите на

$$x_1 - y_1, x_2 - y_2, \ldots, x_n - y_n$$

и се намира числото $\upsilon_н^+$ - броя на разликите с положителни знаци.

Ако за някоя разлика имаме $x_i - y_i = 0$, то тя се маха от извадката и обема й се намалява с единица.

Ако двете извадки са еднородни, то $\upsilon_н^+$ е случайна величина, която е разпределена биномно с единична вероятност $p = 1/2$, т.е. $\upsilon_н^+ \in Bi(n, p = 1/2)$.

Следователно проверката на основна хипотеза H_0 срещу алтернатива H_1 е еквивалентна на проверката на основна хипотеза H_0 срещу алтернатива H_1, където

$H'_0 : \upsilon_n^+ \in Bi(n, p = 1/2)$

$H'_1 : \upsilon_n^+ \in Bi(n, p \neq 1/2)$

Трябва да се има в предвид, че при обем на извадката 10-15 наблюдения, критерият на знаците ще отхвърли основната хипотеза H_0, само когато p е близко до 1 или до 0. За това е добре критерият на знаците да се прави само при голям обем на извадката n, т.е. поне $n \geq 50$.

Ако $n \geq 50$ биномното разпределение може да се апроксимира (теоремата на Моавар-Лаплас) с нормално разпределение с параметри $\mu = n\,p = n/2$ и $\sigma = \sqrt{n\,p(1-p)} = \sqrt{n}/2$, т.е. $H_0 : \upsilon_n^+ \in N(\mu = n/2, \sigma = \sqrt{n}/2)$.

Следователно проверката на основна хипотеза H'_0 срещу алтернатива H'_1 е еквивалентна на проверката на основна хипотеза H''_0 срещу алтернатива H''_1, където

$H''_0 : \upsilon_н^+ \in N(\mu = n/2, \sigma = \sqrt{n}/2)$

$H''_1 : \upsilon_н^+ \notin N(\mu = n/2, \sigma = \sqrt{n}/2)$

Зад.1

Прави се извадка от 100 паралелни наблюдения, за продължителността на светене на лампите, произведени от два различни цеха. Установено е, че в 36 от случаите лампите на първия цех светят по-дълго от лампите на втория цех.

Със сигурност 95% проверете хипотезата, че лампите от двата цеха светят едно и също време.

Решение:

Имаме $\upsilon_n^+ = 36$, $n = 100$ и $\dfrac{\upsilon_n^+ - n/2}{\sqrt{n}/2} \in N(0,1)$

а) $P\left(-z < \dfrac{\upsilon_n^+ - n/2}{\sqrt{n}/2} < z\right) = 0.95$

$$P\left(-z < \frac{\upsilon_n^+ - n/2}{\sqrt{n}/2} < z\right) = 0.95$$

$$P\left(-z < \frac{\upsilon_n^+ - n/2}{\sqrt{n}/2} < 0\right) + P\left(0 < \frac{\upsilon_n^+ - n/2}{\sqrt{n}/2} < z\right) = 0.95$$

$$2\,P\left(0 < \frac{\upsilon_n^+ - n/2}{\sqrt{n}/2} < z\right) = 0.95$$

$$P\left(0 < \frac{\upsilon_n^+ - n/2}{\sqrt{n}/2} < z\right) = 0.475$$

От таблицата за $N(0,1)$ намираме, че $z = 1.96$, тогава

$$-1.96 < \frac{\upsilon_n^+ - n/2}{\sqrt{n}/2} < 1.96$$

$$-1.96 < \frac{36-50}{5} < 1.96$$

$-1.96 < 2.8 < 1.96$ - не е вярно

Понеже последното неравенство не е вярно, то отхвърляме основната хипотеза и приемаме алтернативната хипотеза, т.е. приемаме, че лампите от двата цеха не светят едно и също време.

ТЕМА 7
Тест на Пирсън

С този тест (при избрано ниво на доверие α) се проверява основна хипотеза H_0 срещу алтернатива H_1 , където:

H_0 : *Направената извадка има функцията на разпределени* $- F(x)$

H_1 : *Направената извадка няма функцията на разпределени* $- F(x)$.

Този тест се прилага за групирана извадка.

Тестът работи добре за обем на извадката $n \geq 40$ и честоти на попадение във всеки отделен интервал $m_i \geq 5$.

Теста използва разпределението

$$\sum_{i=1}^{k} \frac{(m_i - n\, p_i)^2}{n\, p_i} \in \chi^2(k - r - 1),$$

където

m_i - извадкови честоти;

$n\, p_i$ - теоретични честоти;

k -броя на интервалите на групиране на извадката;

r -броя на параметрите подлежащи на оценка на базата на данни от извадката.

Теста се прави така - пресмята се наблюдавания коефициент на Пирсън $\sum_{i=1}^{k} \frac{(m_i - n\, p_i)^2}{n\, p_i}$ и се проверява дали той е по-малък от теоретичния такъв $\chi^2_{1-\alpha}(k - r - 1)$. Ако това е изпълнено H_0 се приема, ако не е H_0 се отхвърля и се приема H_1.

Зад.1

С ниво на доверие $\alpha = 0.01$ да се провери еднакъв ли е добиват от районите

Район- i	1	2	3	4	5
Добив- m_i	110	130	70	90	100

Решение:

Издигаме хипотезите:

H_0 : Добивите от раййоните са еднакви и те са - n/5

H_1 : Добивите от райоите не са еднакви

където

$n = 110 + 130 + 70 + 90 + 100 = 500$ е обема на извадката.

$p_i = \frac{1}{5}$ са теоретичните вероятности за производство от всеки район.

Пресмятаме наблюдавания коефициент на Пирсън

$$\chi^2 = \sum_{i=1}^{5} \frac{(m_i - n\, p_i)^2}{n\, p_i} = \frac{(110 - 500\frac{1}{5})^2}{500\frac{1}{5}} + \frac{(130 - 500\frac{1}{5})^2}{500\frac{1}{5}} + \frac{(70 - 500\frac{1}{5})^2}{500\frac{1}{5}}$$

$$+ \frac{(90 - 500\frac{1}{5})^2}{500\frac{1}{5}} + \frac{(100 - 500\frac{1}{5})^2}{500\frac{1}{5}} = \frac{2000}{10} = 20$$

От таблицата за χ^2 намираме теоретичния коефициент на Пирсън $\chi^2_{1-\alpha}(k - r - 1) = \chi^2_{0.99}(4) = 13.28$, в случая $r = 0$ понеже при равномерно разпределение няма параметри, които трябва да се оценят на базата на данни от извадката.

Получихме $\chi^2 = 20 > \chi^2_{0.99}(4) = 13.28$, следователно отхвърляме основната хипотеза и приемаме алтернативната хипотеза, т.е. приемаме че добивът от районите не е еднакъв.

Зад.2

С ниво на доверие $\alpha = 0.01$ да се провери дали случайната величина *X* има Поасоново разпределение

x_i	0	1	2	3	4 или по-голямо
m_i	229	221	93	35	8

Решение:

Припомняме, че Поасоново разпределена(дискретна) случайна величина X с параметър λ / $X \in P_o(\lambda)$ / има ред на разпределение

$$P(X = k) = \frac{\lambda^k}{k!} e^{-\lambda} \text{ за } k = 0,1,\ldots,\infty .$$

Първо ще оценим параметъра λ с данни от извадката. Като се използва методът на неизместените оценки (Въпрос 2) се получава, че

$$\lambda = \overline{X} = \frac{1}{n}\sum_{i=1}^{5} \overline{x}_i \; m_i = \frac{0.229 + 1.221 + 2.93 + 3.35 + 4.8}{586} = \frac{544}{586} = 0.93,$$

където $n = 229 + 221 + 93 + 35 + 8 = 586$ е обема на извадката.

Издигаме хипотезите:

H_0 : Случайната величина има Поасоново разпределение

H_1 : Случайната величина няма Поасоново разпределение .

Сега трябва да пресметнем теоретичните вероятности $p_k = P(X = k)$, т.е.

$$p_0 = P(X = 0) = \frac{\lambda^0}{0!} e^{-\lambda} = 0.395$$

$$p_1 = P(X = 1) = \frac{\lambda}{1!} e^{-\lambda} = 0.366$$

$$p_2 = P(X = 2) = \frac{\lambda^2}{2!} e^{-\lambda} = 0.17$$

$$p_3 = P(X = 3) = \frac{\lambda^3}{3!} e^{-\lambda} = 0.055$$

$$p_4 = 1 - p_0 - p_1 - p_2 - p_3 = 0.016$$

Пресмятаме наблюдавания коефициент на Пирсън

$$\chi^2 = \sum_{i=0}^{4} \frac{(m_i - n\,p_i)^2}{n\,p_i} = \frac{(229 - 586\;0.395)^2}{586\;0.395} + \frac{(221 - 586\;\;0.306)^2}{586\;0.306} + \frac{(93 - 586\;0.17)^2}{586\;0.17}$$

$$+ \frac{(35 - 586\;0.055)^2}{586\;0.055} + \frac{(8 - 586\;0.016)^2}{586\;0.016} = 1.3$$

От таблицата за χ^2 намираме теоретичния коефициент на Пирсън $\chi^2_{1-\alpha}(k - r - 1) = \chi^2_{0.99}(3) = 11.34$, в случая $r = 1$ понеже при Пирсъновото разпределение има 1 параметър, които се оценява на базата на данни от извадката.

Получихме $\chi^2 = 1.3 < \chi^2_{0.99}(3) = 11.34$, следователно приемаме основната хипотеза, т.е. приемаме че данните са разпределени Пирсъново.

Зад.3

С ниво на доверие $\alpha = 0.01$ да се провери дали случайната величина *X* има нормално(Гаусово) разпределение

Интервал- i	2.9 – 3.9	3.9 – 4.9	4.9 – 5.9	5.9 - 6.9	6.9 – 7.9
m_i	5	15	23	19	6

Решение:

Първо ще оценим параметрите μ и σ с данни от извадката. Като се използва метода на неизместените оценки (Въпрос 2) се получава, че

$$\mu = \overline{X} = \frac{1}{n}\sum_{i=1}^{5}\bar{x}_i\ m_i = \frac{3.4\ 5 + 4.4\ 15 + 5.4\ 23 + 6.4\ 19 + 7.4\ 6}{5+15+23+19+6} = \frac{373.2}{68} = 5.5$$

$$s^2 = \frac{1}{n-1}\sum_{i=1}^{5}(\bar{x}_i - \overline{X})^2\ m_i =$$

$$= \frac{(3.4\text{-}5.5)^2 5 + (4.4\text{-}5.5)^2 15 + (5.4\text{-}5.5)^2 23 + (6.4\text{-}5.5)^2 19 + (7.4\text{-}5.5)^2 6}{67} = 1.16$$

$$s = \sqrt{1.16} \approx 1$$

Издигаме хипотезите:

H_0 : Случайната величина има Гаусово разпределение - $N(\mu = 5.5, \sigma = 1)$

H_1 : Случайната величина няма Гаусово разпределение - $N(\mu = 5.5, \sigma = 1)$.

Ще обърнем внимание на това, че нямаме наблюдения по-малки от 2.9 и по-големи от 7.9, но въпреки това съществува теоретична вероятности да се появят наблюдения извън този интервал. За това за да пресметнем теоретичните вероятности p_i ще разширим таблицата така

Интервал i	-∞ 2.9	2.9 – 3.9	3.9 – 4.9	4.9 – 5.9	5.9 - 6.9	6.9 – 7.9	7.9 ∞
m_i	0	5	15	23	19	6	0
$n\,p_i = 68p_i$	0.3	3.4	14.9	25.9	17.9	4.9	0.6

Като използване таблицата за стандартно нормално разпределение пресмятаме

$$p_0 = P(-\infty < X < 2.9) = P(-\infty < Z < \frac{2.9-5.5}{1}) = P(-\infty < Z < -2.6) =$$
$$= 0.5 - P(0 < Z < 2.6) = 0.5 - 0.4953 = 0.0047$$

$$p_1 = P(2.9 < X < 3.9) = P(\frac{2.9-5.5}{1} < Z < \frac{3.9-5.5}{1}) = P(-2.6 < Z < -1.6) =$$
$$= P(0 < Z < 2.6) - P(0 < Z < 1.6) = 0.4953 - 0.4452 = 0.0501$$

$$p_2 = P(3.9 < X < 4.9) = P(\frac{3.9-5.5}{1} < Z < \frac{4.9-5.5}{1}) = P(-1.6 < Z < -0.6) =$$
$$= P(0 < Z < 1.6) - P(0 < Z < 0.6) = 0.4452 - 0.2258 = 0.2194$$

$$p_3 = P(4.9 < X < 5.9) = P(\frac{4.9-5.5}{1} < Z < \frac{5.9-5.5}{1}) = P(-0.6 < Z < 0.4) =$$
$$= P(0 < Z < 0.6) + P(0 < Z < 0.4) = 0.2258 + 0.1554 = 0.3812$$

$$p_4 = P(5.9 < X < 6.9) = P(\frac{5.9-5.5}{1} < Z < \frac{6.9-5.5}{1}) = P(0.4 < Z < 1.4) =$$
$$= P(0 < Z < 1.4) - P(0 < Z < 0.4) = 0.4192 - 0.1554 = 0.2638$$

$$p_5 = P(6.9 < X < 7.9) = P(\frac{6.9-5.5}{1} < Z < \frac{7.9-5.5}{1}) = P(1.4 < Z < 2.4) =$$
$$= P(0 < Z < 2.4) - P(0 < Z < 1.4) = 0.4918 - 0.4192 = 0.0728$$

$$p_6 = P(7.9 < X < \infty) = P(\frac{7.9-5.5}{1} < Z < \infty) = P(2.4 < Z < \infty) =$$
$$= 0.5 - P(0 < Z < 2.4) = 0.5 - 0.4192 = 0.0082$$

Пресмятаме наблюдавания коефициент на Пирсън

$$\chi^2 = \sum_{i=0}^{6} \frac{(m_i - n\,p_i)^2}{n\,p_i} = \frac{0.3^2}{0.3} + \frac{(5-3.4)^2}{3.4} + \frac{(15-14.9)^2}{14.9}$$
$$+ \frac{(23-25.9)^2}{25.9} + \frac{(19-17.9)^2}{17.9} + \frac{(6-4.9)^2}{4.9} + \frac{0.6^2}{0.6} = 1.993$$

От таблицата за χ^2 намираме теоретичния коефициент на Пирсън $\chi^2_{1-\alpha}(k-r-1) = \chi^2_{0.99}(4) = 13.28$, в случая $r = 2$ понеже Гаусовото разпределение има 2 параметъра, които се оценка на базата на данни от извадката.

Получихме $\chi^2 = 1.993 < \chi^2_{0.99}(3) = 11.34$, следователно приемаме основната хипотеза, т.е. приемаме че данните са разпределени Гаусово.

ТЕМА 8
Корелация. Линейна регресия (метод на най-малките квадрати)

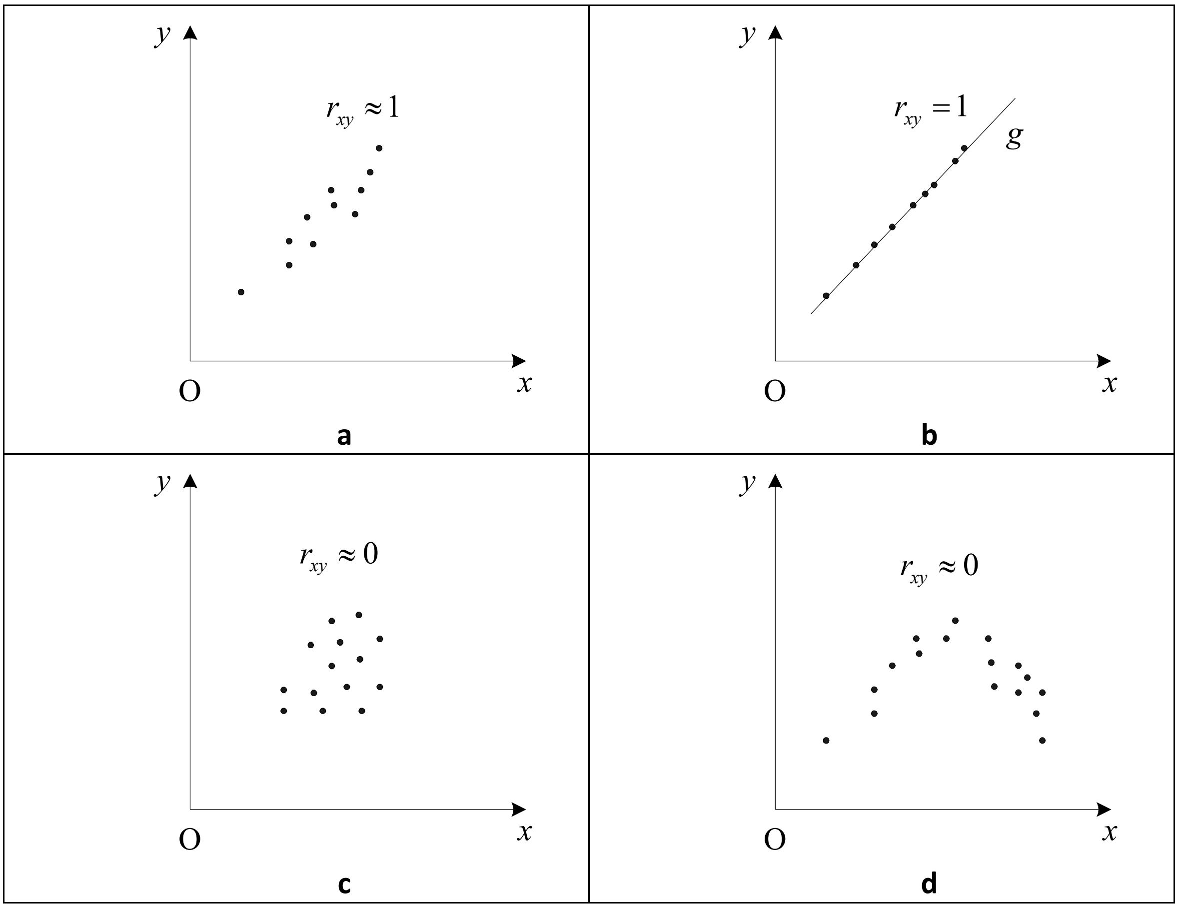
Зависимост между случайни величини.

Нека *X* и *Y* са случайни величини, за които предполагаме че са зависими (в някаква степен). За да проверим това, образуваме двойките данни (сдвоената извадка)

$x_1, x_2, \ldots, x_n$

$y_1, y_2, \ldots, y_n$.

След което нанасяме точките (x_i, y_i) на една координатна система, фиг. 1



Фигура 1

На фигура 1a се вижда, че когато *X* расте и *Y* расте, следователно между *X* и *Y* има зависимост, в този случай говорим за положителен тренд. В случая , когато *X* расте и *Y* намалява, говорим за отрицателен тренд.

На фигура 1b се вижда, че когато *X* заеме някаква стойност, то *Y* заема строго фиксирана стойност (лежи на правата *g*), следователно между *X* и *Y* има 100% зависимост.

На фигура 1c се вижда, че облакът от данни по-скоро е хаотичен и между *X* и *Y* няма зависимост.

На фигура 1d се вижда, че когато *X* расте и *Y* първо расте, но след определена стойност намалява. Следователно между *X* и *Y* има зависимост, но тази зависимост определено не е линейна (както беше на фигура 1a и 1b).

Коефициент на корелация r_{xy} между *X* и *Y*.

Забележка: В теория на вероятностите r_{xy} се въвежда така

$$r_{xy} = \frac{E\big((X-EX)(Y-EY)\big)}{\sqrt{E\big((X-EX)^2\big)}\sqrt{E\big((Y-EY)^2\big)}} = \frac{E\big((X-EX)(Y-EY)\big)}{\sqrt{DX}\sqrt{DY}}$$

и за r_{xy} се доказва твърдението

$|r_{xy}| = 1 \Leftrightarrow$ когато съществуват константи $a \neq 0$ и b, такива че $Y = aX + b$.

В статистиката коефициент на корелация- r_{xy} се пресмята така

$$r_{xy} = \frac{K_{xy}}{s_x\ s_y},$$

където

$K_{xy} = \frac{1}{n-1}\sum_{i=1}^{n}(x_i - \overline{X})(y_i - \overline{Y})$ - се нарича момент на корелация;

$\overline{X} = \frac{1}{n}\sum_{i=1}^{n} x_i$ и $\overline{Y} = \frac{1}{n}\sum_{i=1}^{n} y_i$ - извадкови средни;

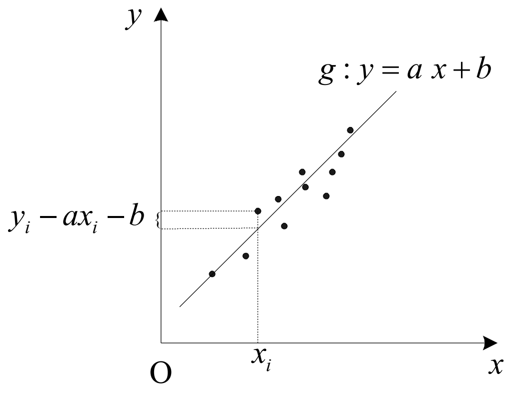
$s_x^2 = \frac{1}{n-1}\sum_{i=1}^{n}(x_i - \overline{X})^2$ и $s_y^2 = \frac{1}{n-1}\sum_{i=1}^{n}(y_i - \overline{Y})^2$ - извадкови дисперсии.

Изложеното преди малко твърдение може да се тълкува така – коефициента r_{xy} е "мярка" за съществуване на линейна зависимост между *X* и *Y* (фигура 1a и 1b), т.е. колкото $|r_{xy}|$ е по-близък до 1, толкова по-силна линейна зависимост има. Обикновено при $|r_{xy}| \approx 0.9$ говорим за съществуване на линейна зависимост между *X* и *Y* (фигура 1a), но стойността 0.9 не е строго фиксирана, тя се избира съгласно спецификата на всяка конкретна задача (може да е както по-голяма, така и по-малка).

Ще обърнем внимание на факта, че коефициентът r_{xy} не може да определи дали между *X* и *Y* има **нелинейна** зависимост (фигура 1d). В случая от фигура 1c няма зависимост, а в случая от фигура 1d има **нелинейна** зависимост, но и в двата случая $|r_{xy}| \approx 0$. Така, че r_{xy} се явява "мярка" за съществуване само на **линейна** зависимост между *X* и *Y*.

Линейна регресия (метод на най-малките квадрати).

Сега ще се върнем към фигура 1a, където имаме някаква линейна зависимост между X и Y (това може да се установи, като се провери че $|r_{xy}| \approx 0.9$). Сега ще търсим такава права $g: y = ax + b$, която е "най-близка" до всички данни от облака, фигура 2.



Фигура 2

За критерии на "близост", между правата g и облака от данни избираме сумата от квадратите на разликите между y_i и ординатата на точката от правата g с абциса x_i (фигура 2), т.е.

$$S(a,b)=\sum_{i=1}^{n}(y_i-ax_i-b)^2 \xrightarrow[a,b]{} \min .$$

Понеже горното е сума от квадрати, на която се търси най-малката стойност, то метода се нарича метод на най-малките квадрати.

Това е задача за намиране на минимум, на функцията на две променливи $S(a,b)$. От необходимото условие за екстремум получаваме

$$\left|\begin{array}{l}\dfrac{\partial S(a,b)}{\partial a}=0\\ \dfrac{\partial S(a,b)}{\partial b}=0\end{array}\right. \quad \text{или} \quad \left|\begin{array}{l}-2\sum_{i=1}^{n}(y_i-ax_i-b)x_i=0\\ -2\sum_{i=1}^{n}(y_i-ax_i-b)=0\end{array}\right. \quad \text{или} \quad \left|\begin{array}{l}a\sum_{i=1}^{n}x_i^2+b\sum_{i=1}^{n}x_i=\sum_{i=1}^{n}y_i\ x_i\\ a\sum_{i=1}^{n}x_i+b\sum_{i=1}^{n}1=\sum_{i=1}^{n}y_i\end{array}\right. \quad \text{или}$$

$$\left|\begin{array}{l}a\sum_{i=1}^{n}x_i^2+b\sum_{i=1}^{n}x_i=\sum_{i=1}^{n}y_i\ x_i\\ a\sum_{i=1}^{n}x_i+b\ n=\sum_{i=1}^{n}y_i\end{array}\right. .$$

Последната система се решава по метода на Крамер (с детерминанти) така

$$a=\frac{\begin{vmatrix}\sum_{i=1}^{n}y_i\ x_i & \sum_{i=1}^{n}x_i\\ \sum_{i=1}^{n}y_i & n\end{vmatrix}}{\begin{vmatrix}\sum_{i=1}^{n}x_i^2 & \sum_{i=1}^{n}x_i\\ \sum_{i=1}^{n}x_i & n\end{vmatrix}} \text{ и } b=\frac{\begin{vmatrix}\sum_{i=1}^{n}x_i^2 & \sum_{i=1}^{n}y_i\ x_i\\ \sum_{i=1}^{n}x_i & \sum_{i=1}^{n}y_i\end{vmatrix}}{\begin{vmatrix}\sum_{i=1}^{n}x_i^2 & \sum_{i=1}^{n}x_i\\ \sum_{i=1}^{n}x_i & n\end{vmatrix}} .$$

Зад.1

Информацията за 10 служители на една фирма е

Служител	Коефициент -X	Заплата - *Y*	X^2	Y^2	*X.Y*
1	1.8	105	3.24	11025	189
2	2.2	108	4.84	11664	237.6
3	1.7	95	2.89	9025	161.5
4	1.4	94	1.96	8836	131.6
5	2.7	106	7.29	11236	286.2
6	2.4	102	5.76	10404	244.8
7	1.3	90	1.69	8100	117
8	2.7	115	7.29	13225	310.5
9	2.1	99	4.41	9801	207.9
10	2.3	110	5.29	12100	253
$\sum$	20.6	1024	44.66	105416	2139.1

Да се пресметне коефициента на корелация r_{xy} и да се състави уравнението на линейната регресия $g: y = ax + b$.

Решение:

$$K_{xy} = \frac{1}{n-1}\sum_{i=1}^{n}(x_i - \overline{X})(y_i - \overline{Y}) = \frac{1}{n-1}\left(\sum_{i=1}^{n} x_i\ y_i - n\ \overline{X}\ \overline{Y}\right) =$$

$$= \frac{2139.1 - \dfrac{20.6\quad 1024}{10}}{9} = \frac{29.66}{9} = 3.295$$

$$s_x^2 = \frac{1}{n-1}\sum_{i=1}^{n}(x_i - \overline{X})^2 = \frac{1}{n-1}\left(\sum_{i=1}^{n} x_i^2 - n\ \overline{X}^2\right) =$$

$$= \frac{44.66 - \dfrac{20.6\quad 20.6}{10}}{9} = \frac{2.224}{9} = 0.2491$$

$$s_x^2 = \frac{1}{n-1}\sum_{i=1}^{n}(y_i - \overline{Y})^2 = \frac{1}{n-1}\left(\sum_{i=1}^{n} y_i^2 - n\ \overline{Y}^2\right) =$$

$$= \frac{105416 - \dfrac{1024\quad 1024}{10}}{9} = \frac{558.4}{9} = 62.04$$

$r_{xy} = \dfrac{K_{xy}}{s_x\ s_y} = \dfrac{3.295}{\sqrt{0.2491}\sqrt{62.04}} = 0.841$ - може да се приеме, че има някаква линейна зависимост

Коефициентите a и b смятаме по формулите

$$a = \frac{\begin{vmatrix} \sum_{i=1}^{n} y_i\ x_i & \sum_{i=1}^{n} x_i \\ \sum_{i=1}^{n} y_i & n \end{vmatrix}}{\begin{vmatrix} \sum_{i=1}^{n} x_i^2 & \sum_{i=1}^{n} x_i \\ \sum_{i=1}^{n} x_i & n \end{vmatrix}} = \frac{\begin{vmatrix} 2139.1 & 20.6 \\ 1024 & 10 \end{vmatrix}}{\begin{vmatrix} 44.66 & 20.6 \\ 20.6 & 10 \end{vmatrix}} = \frac{2139.1\quad 10 - 20.6\quad 1024}{44.66\quad 10 - 20.6^2} = \frac{296.6}{22.24} = 13.34$$

$$b = \frac{\begin{vmatrix} \sum_{i=1}^{n} x_i^2 & \sum_{i=1}^{n} y_i\ x_i \\ \sum_{i=1}^{n} x_i & \sum_{i=1}^{n} y_i \end{vmatrix}}{\begin{vmatrix} \sum_{i=1}^{n} x_i^2 & \sum_{i=1}^{n} x_i \\ \sum_{i=1}^{n} x_i & n \end{vmatrix}} == \frac{\begin{vmatrix} 44.66 & 2139.1 \\ 20.6 & 1024 \end{vmatrix}}{\begin{vmatrix} 44.66 & 20.6 \\ 20.6 & 10 \end{vmatrix}} = \frac{44.66\quad 1024 - 20.6\quad 2139.1}{44.66\quad 10 - 20.6^2} = \frac{1666.38}{22.24} = 74.93$$

т.е. $y = 13.34\ x + 74.93$

ТЕМА 9
Времеви редове

Първо се съставя модел на линейна регресия (Тема 8), след това този модел се коригира чрез добавяне на сезонна компонента и накрая той се използва за прогнозиране на бъдещо поведение.

Ще припомним(Тема 8), че на базата на извадката (x_i, y_i), за $i = 1, \ldots, n$ търсим връзка от вида $y = ax + b$, където коефициентите a и b се пресмятате от системата

$$\left| \begin{array}{l} a\sum_{i=1}^{n} x_i^2 + b\sum_{i=1}^{n} x_i = \sum_{i=1}^{n} y_i \ x_i \\ a\sum_{i=1}^{n} x_i + b\ n = \sum_{i=1}^{n} y_i \end{array} \right. .$$

Най-често системата се решава по метода на Крамер (с детерминанти), т.е.

$$a = \frac{\Delta_1}{\Delta} \text{ и } b = \frac{\Delta_2}{\Delta}, \text{ където}$$

$$\Delta_1 = \begin{vmatrix} \sum_{i=1}^{n} y_i \ x_i & \sum_{i=1}^{n} x_i \\ \sum_{i=1}^{n} y_i & n \end{vmatrix}, \ \Delta_2 = \begin{vmatrix} \sum_{i=1}^{n} x_i^2 & \sum_{i=1}^{n} y_i \ x_i \\ \sum_{i=1}^{n} x_i & \sum_{i=1}^{n} y_i \end{vmatrix} \text{ и } \Delta = \begin{vmatrix} \sum_{i=1}^{n} x_i^2 & \sum_{i=1}^{n} x_i \\ \sum_{i=1}^{n} x_i & n \end{vmatrix}.$$

Как се намира сезонната компонента ще бъде показано в самата задача.

Зад.1

Дадени са продажбите в хиляди левове на търговски обект по тримесечия за периода 2008-2010г.

Период		Продажби -Y_i	Тримесечие -t_i	t_i^2	$t_i.Y_i$
2008	1	6	1	1	6
	2	8	2	4	16
	3	32	3	9	96
	4	24	4	16	96
2009	1	10	5	25	50
	2	15	6	36	90
	3	32	7	49	224
	4	23	8	64	184
2010	1	10	9	81	90
	2	18	10	100	180
	3	30	11	121	330
	4	32	12	144	384
$\sum$		240	78	650	1746

Намерете линейния тренд на промяната на тримесечните продажби, сезонната компонента на продажбите и направете прогноза за тримесечните продажби за 2011г.

Решение:

Линейна регресия

Съставяме системата

$$\begin{cases} a\sum_{i=1}^{n} t_i^2 + b\sum_{i=1}^{n} t_i = \sum_{i=1}^{n} y_i \ t_i \\ a\sum_{i=1}^{n} t_i + b\ n = \sum_{i=1}^{n} y_i \end{cases} \qquad \begin{cases} a\ 650 + b\ 78 = 1746 \\ a\ 78 + b\ 12 = 240 \end{cases}$$

Коефициентите a и b смятаме по формулите на Крамер

$$a = \frac{\begin{vmatrix} 1746 & 78 \\ 240 & 12 \end{vmatrix}}{\begin{vmatrix} 650 & 78 \\ 78 & 12 \end{vmatrix}} = \frac{1746 \quad 12 - 240 \quad 78}{650 \quad 12 - 78^2} = \frac{2232}{1716} = 1.3$$

$$b = \frac{\begin{vmatrix} 650 & 1746 \\ 78 & 240 \end{vmatrix}}{\begin{vmatrix} 650 & 78 \\ 78 & 12 \end{vmatrix}} = \frac{650 \quad 240 - 1746 \quad 78}{650 \quad 12 - 78^2} = \frac{19812}{1716} = 11.54 \text{ т.е.}$$

$y = 1.3\ x + 11.54$ - линейна регресия (тренд)

Сезонна компонента

Година / Тримесечие	2008	2009	2010	$\overline{Y}_i$	I_i	$Y_{i,2011}$ Тренд	$\hat{Y}_{i,2011}$ Коригирана
1	6	10	10	26/3=8.76	0.452	28.44	12.84
2	8	15	18	41/3=13.67	0.712	29.74	21.17
3	32	32	30	94/3=31.33	1.632	31.04	50.65
4	24	23	32	69/3=23	1.198	32.34	38.74
$\sum$				76.67			

За удобство означаваме y_{ij} продажбите за i-тото тримесечие на j-тата година, където $i = 1,2,3,4$ и $j = 1,2,3$. Тогава

$$\overline{Y}_i = \frac{1}{n}\sum_{i=1}^{n} y_{ij} = \frac{1}{3}\sum_{i=1}^{3} y_{ij}$$

$$\overline{Y} = \frac{1}{m}\sum_{i=1}^{m} \overline{Y}_i = \frac{1}{4}\sum_{i=1}^{4} \overline{Y}_i = \frac{76.67}{4} = 19.17$$

$$I_i = \frac{\overline{Y}_i}{\overline{Y}}$$

Прогноза за тримесечните продажби:

а) тренд

$$\overline{Y}_{1,2011} = 1.3 \quad 13 + 11.54 = 28.44$$

$$\overline{Y}_{2,2011} = 1.3 \quad 14 + 11.54 = 29.74$$

$$\overline{Y}_{3,2011} = 1.3 \quad 15 + 11.54 = 31.04$$

$$\overline{Y}_{4,2011} = 1.3 \quad 16 + 11.54 = 32.34$$

б) коригиране със сезонна компонента

$$\hat{Y}_{1,2011} = \overline{Y}_{1,2011}\, I_1 = 28.44 \quad 0.452 = 12.83$$

$$\hat{Y}_{2,2011} = \overline{Y}_{2,2011}\, I_2 = 29.74 \quad 0.712 = 21.17$$

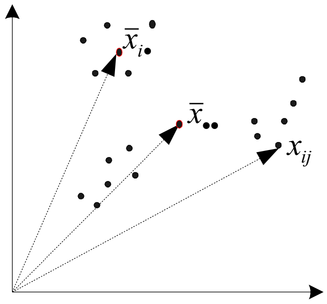
$$\hat{Y}_{3,2011} = \overline{Y}_{3,2011} \, I_3 = 31.04 \quad 1.632 = 50.65$$

$$\hat{Y}_{4,2011} = \overline{Y}_{4,2011} \, I_4 = 32.34 \quad 1.198 = 38.74$$

ТЕМА 10
Дисперсионен анализ

Въведение

Да разгледаме извадката $\{x_{ij}\}_{i=1}^{m}{}_{j=1}^{n}$ с двумерни индекси, показана на фигура 1.



Фигура 1

От фигурата се вижда, че данните са групирани в три подмножества. В повечето случай тези подмножества не са така ясно изразени и тогава възниква въпроса доколко можем да твърдим, че такова групиране има. На този въпрос може да се отговори с помощта на дисперсионния анализ.

Обикновено групирането (фигура 1) се дължи на това, че факторът X заема различни стойности по време на изпитанията и новото което той заеме определя до коя група ще се доближи съответното наблюдение.

Да означим с

$\bar{x}_{\bullet\bullet} = \dfrac{1}{n\,m}\sum_{i=1}^{m}\sum_{j=1}^{n} x_{ij}$ - извадковото средно на всички данни

$s^2 = \dfrac{1}{(n-1)\,(m-1)}\sum_{i=1}^{n}\sum_{j=1}^{m}(x_{ij} - \bar{x}_{\bullet\bullet})^2$ - извадкова дисперсия на всички данни

$\bar{x}_{i\bullet} = \dfrac{1}{n}\sum_{j=1}^{n} x_{ij}$ - извадковото средно на *i*-тата група от данни.

При дисперсионния анализ цялата извадкова диспепсия се разбива на две дисперсии(едномерен дисперсионен анализ) и ако се предположи, че генералната съвкупност е разпределена нормално, то тогава частно на тези две дисперсии има F-разпределение (въз основа на данни от извадката се проверява дали това наистина е така).

Тук ще разгледаме само еднофакторен и двуфакторен дисперсионен анализ.

Еднофакторен дисперсионен анализ

Дадена е таблицата

		Нива на фактора		
		1	…	*n*
Наблюдения	1	x_{11}	…	x_{1n}
	⋮			
	m	x_{m1}	…	$x_{m\,n}$

В горната таблица, x_{ij} е отклика при i-тото наблюдение за j-тото ниво на фактора.

Сега ще намерим едно представяне на сумата

$$\sum_{i=1}^{m}\sum_{j=1}^{n}(x_{ij}-\bar{x}_{\bullet\bullet})^2 = \sum_{i=1}^{m}\sum_{j=1}^{n}(x_{ij}-\bar{x}_{\bullet j}+\bar{x}_{\bullet j}-\bar{x}_{\bullet\bullet})^2 =$$

$$= \sum_{i=1}^{m}\sum_{j=1}^{n}(x_{ij}-\bar{x}_{\bullet j})^2 + 2\sum_{i=1}^{m}\sum_{j=1}^{n}(x_{ij}-\bar{x}_{\bullet j})(\bar{x}_{\bullet j}-\bar{x}_{\bullet\bullet}) + \sum_{i=1}^{m}\sum_{j=1}^{n}(\bar{x}_{\bullet j}-\bar{x}_{\bullet\bullet})^2 =$$

$$= \sum_{i=1}^{m}\sum_{j=1}^{n}(x_{ij}-\bar{x}_{\bullet j})^2 + 2\sum_{i=1}^{m}\sum_{j=1}^{n}(x_{ij}-\bar{x}_{\bullet j})(\bar{x}_{\bullet j}-\bar{x}_{\bullet\bullet}) + m\sum_{j=1}^{n}(\bar{x}_{\bullet j}-\bar{x}_{\bullet\bullet})^2 =$$

$= SSE + 2Dif + SStr$,

където

$SSR = \sum_{i=1}^{m}\sum_{j=1}^{n}(x_{ij}-\bar{x}_{\bullet j})^2$ - остатъчна сума

$SStr = m\sum_{j=1}^{n}(\bar{x}_{\bullet j}-\bar{x}_{\bullet\bullet})^2$ - междугрупова сума

$Dif = \sum_{i=1}^{m}\sum_{j=1}^{n}(x_{ij}-\bar{x}_{\bullet j})(\bar{x}_{\bullet j}-\bar{x}_{\bullet\bullet}) = 0$

За да е коректен този резултат ще покажем, че наистина $Dif = 0$, т.е.

$$\mathrm{D}if = \sum_{i=1}^{m}\sum_{j=1}^{n}(x_{ij}-\bar{x}_{\bullet j})(\bar{x}_{j\bullet}-\bar{x}_{\bullet\bullet}) =$$

$$= \sum_{i=1}^{m}\sum_{j=1}^{n}x_{ij}\ \bar{x}_{\bullet j} - \sum_{i=1}^{m}\sum_{j=1}^{n}x_{ij}\ \bar{x}_{\bullet\bullet} - \sum_{i=1}^{m}\sum_{j=1}^{n}(\bar{x}_{\bullet j})^2 + \sum_{i=1}^{m}\sum_{j=1}^{n}\bar{x}_{\bullet j}\ \bar{x}_{\bullet\bullet} =$$

$$= \sum_{j=1}^{n}\bar{x}_{\bullet j}\sum_{i=1}^{m}x_{ij} - \bar{x}_{\bullet\bullet}\sum_{i=1}^{m}\sum_{j=1}^{n}x_{ij} - \sum_{j=1}^{n}(\bar{x}_{\bullet j})^2\sum_{i=1}^{m}1 + \bar{x}_{\bullet\bullet}\sum_{j=1}^{n}\bar{x}_{\bullet j}\sum_{i=1}^{m}1 =$$

$$= \sum_{j=1}^{n}\bar{x}_{\bullet j}\ m\ \bar{x}_{\bullet j} - \bar{x}_{\bullet\bullet}\ n\ m\ \bar{x}_{\bullet\bullet} - \sum_{j=1}^{n}(\bar{x}_{\bullet j})^2 m + \bar{x}_{\bullet\bullet}\sum_{j=1}^{n}\bar{x}_{i\bullet}\ m =$$

$$= m\sum_{j=1}^{n}(\bar{x}_{\bullet j})^2 - n\ m\ (\bar{x}_{\bullet\bullet})^2 - m\sum_{j=1}^{n}(\bar{x}_{\bullet j})^2 + \bar{x}_{\bullet\bullet}\ n\bar{x}_{\bullet\bullet}\ m = 0\,.$$

Окончателно получихме

$$\sum_{i=1}^{m}\sum_{j=1}^{n}(x_{ij}-\bar{x}_{\bullet\bullet})^2 = \sum_{i=1}^{m}\sum_{j=1}^{n}(x_{ij}-\bar{x}_{\bullet j})^2 + m\sum_{j=1}^{n}(\bar{x}_{\bullet j}-\bar{x}_{\bullet\bullet})^2 = SSR + SStr\,.$$

За да се провери дали наистина групите от фигура 1 съществуват (дали различните нива на фактора водят до създаване на такива групи) се проверява дали

$$\frac{SStr/(n-1)}{SSR/\big(n(m-1)\big)} \in F_{1-\alpha}\big(n-1, n(m-1)\big),$$

т.е. проверява се дали частното на дисперсиите $SStr/(n-1)$ и $SSR/\big(n(m-1)\big)$ има *F*-разпределение с $(n-1)$ и $n(m-1)$ степени на свобода.

Приема се.че частното на дисперсиите има *F*-разпределение, ако пресметната стойност на *F*-разпределението (въз основа на данни от извадката) е по-малка от теоретичната му стойност.

Зад.1

Един завод има три еднакви цеха. Данните за произведената от тях продукция са нанесени в таблицата.

		Цех		
		1	2	*3=n*
Наблюдения	1	30	35	39
	2	32	39	38
	3	34	38	42
	4=*m*	28	36	41
	$\bar{x}_{\bullet j}$	31	37	40

С ниво на доверие $\alpha = 0.05$, да се провери дали продукциите произведени от тези три цеха са равни.

Решение

$\bar{x}_{\bullet 1} = \frac{30+32+34+28}{4} = 31$, $\bar{x}_{\bullet 2} = \frac{35+39+38+36}{4} = 37$ и $\bar{x}_{\bullet 3} = \frac{39+38+42+41}{4} = 40$.

$$\bar{x}_{\bullet\bullet} = \frac{31+37+40}{3} = 36$$

$$SStr = m\sum_{j=1}^{n}(\bar{x}_{\bullet j} - \bar{x}_{\bullet\bullet})^2 = 4\left((31-36)^2 + (37-36)^2 + (40-36)^2\right) = 168$$

За да пресметнем $\sum_{i=1}^{m}\sum_{j=1}^{n}(x_{ij} - \bar{x}_{\bullet\bullet})^2$ съставяме таблицата с клетки $(x_{ij} - \bar{x}_{\bullet\bullet})$.

		Цех		
		1	2	*3=n*
Наблюдения	1	-6	-1	3
	2	-4	3	2
	3	-2	2	6
	4=*m*	-8	0	5

Тогава

$$\sum_{i=1}^{m}\sum_{j=1}^{n}(x_{ij} - \bar{x}_{\bullet\bullet})^2 = 6^2 + 1^2 + 3^2 + 4^2 + 3^2 + 2^2 + 2^2 + 2^2 + 6^2 + 8^2 + 0^2 + 5^2 = 208$$

$$SSR = \sum_{i=1}^{m}\sum_{j=1}^{n}(x_{ij} - \bar{x}_{\bullet\bullet})^2 - SStr = 208 - 168 = 40$$

Проверяваме дали

$$F_{набл} = \frac{SStr/(n-1)}{SSR/\left(n(m-1)\right)} = \frac{162/2}{40/9} = 18.5 < F_{1-\alpha}\left(n-1, n(m-1)\right) = F_{0.95}(2,9) = 4.26.$$

Понеже горното неравенство не е вярно, то отхвърляме основната хипотеза и приемаме, че продукциите от трите цеха не са равни.

Двуфакторен дисперсионен анализ

Дадена е таблицата

		Нива на фактора 1		
		1	…	*n*
Нива на фактора 2	1	x_{11}	…	x_{1n}
	⋮			
	m	x_{m1}	…	x_{mn}

В горната таблица, x_{ij} е отклика при i-тото ниво на фактора 2 и j-тото ниво на фактора 1. За разлика от еднофакторния експеримент тук имаме само едно наблюдение при дадено ниво на някой фактор.

Аналогично на еднофакторния експеримент тук се доказва, че е в сила

$$\sum_{i=1}^{m}\sum_{j=1}^{n}(x_{ij}-\bar{x}_{\bullet\bullet})^2 = \sum_{i=1}^{m}\sum_{j=1}^{n}(x_{ij}-\bar{x}_{i\bullet}-\bar{x}_{\bullet j}+\bar{x}_{\bullet\bullet})^2 + n\sum_{i=1}^{m}(\bar{x}_{i\bullet}-\bar{x}_{\bullet\bullet})^2 + m\sum_{j=1}^{n}(\bar{x}_{\bullet j}-\bar{x}_{\bullet\bullet})^2 =$$

$$= SSR + SStr1 + +SStr2$$

където

$SSR = \sum_{i=1}^{m}\sum_{j=1}^{n}(x_{ij}-\bar{x}_{i\bullet}-\bar{x}_{\bullet j}+\bar{x}_{\bullet\bullet})^2$ - остатъчна сума

$SStr1 = n\sum_{i=1}^{m}(\bar{x}_{i\bullet}-\bar{x}_{\bullet\bullet})^2$ - междугрупова сума за фактор 1

$SStr2 = m\sum_{j=1}^{n}(\bar{x}_{\bullet j}-\bar{x}_{\bullet\bullet})^2$ - междугрупова сума за фактор 2

Аналогично на еднофакторния експеримент тук се проверява дали

$$\frac{SStr1/(m-1)}{SSR/\left((n-1)(m-1)\right)} \in F_{1-\alpha}\left(m-1,(n-1)(m-1)\right)$$

$$\frac{SStr2/(n-1)}{SSR/\left((m-1)(n-1)\right)} \in F_{1-\alpha}\left(n-1,(m-1)(n-1)\right).$$

Зад.2

24 кандидат-студенти, които са от три различни региона на България, се явяват на 8 различни теста. Резултатите са нанесени в таблицата

		Регион			
		1	2	*3=n*	$\bar{x}_{i\bullet}$
Тест	1	87	91	80	258/3=86
	2	74	83	60	217/3=72.3
	3	63	71	58	192/3=64
	4	62	76	73	211/3=70.3
	5	83	76	85	244/3=81.3
	6	73	84	71	228/3=76
	7	85	90	77	252/84=84
	8=*m*	70	69	63	202/3=67.3
	$\bar{x}_{\bullet j}$	597/8=74.6	640/8=80	567/8=70.9	$\bar{x}_{\bullet\bullet}=75.2$

С ниво на доверие $\alpha = 0.01$, да се провери: дали тестовете са с еднаква трудност и дали студентите от трите региона са еднакво подготвени.

Решение

$$SStr(тестове) = n\sum_{i=1}^{m}(\bar{x}_{i\bullet}-\bar{x}_{\bullet\bullet})^2 == 3\left((86-75.2)^2+(72.3-75.2)^2+(64-75.2)^2+(70.3-75.2)^2+\right.$$

$$\left.+(86-75.2)^2+(72.3-75.2)^2+(64-75.2)^2+(70.3-75.2)^2\right)=1356.6$$

$$SStr(райони) = m\sum_{j=1}^{n}(\bar{x}_{\bullet j}-\bar{x}_{\bullet\bullet})^2 = 8\left((74.6-75.2)^2+(80-75.2)^2+(70.9-75.2)^2\right)=335.1$$

За да пресметнем $\sum_{i=1}^{m}\sum_{j=1}^{n}(x_{ij}-\bar{x}_{\bullet\bullet})^2$ съставяме таблицата с клетки $(x_{ij}-\bar{x}_{\bullet\bullet})^2$.

		Регион		
		1	2	*3=n*
Тест	1	11.8^2	15.8^2	4.8^2
	2	1.2^2	7.8^2	15.2^2
	3	12.2^2	4.2^2	17.2^2
	4	13.2^2	0.8^2	2.2^2
	5	7.8^2	0.8^2	9.8^2
	6	2.2^2	8.8^2	4.2^2
	7	9.8^2	14.8^2	1.8^2
	8=*m*	5.2^2	6.2^2	12.2^2
	$\sum$	652.52	664.32	820.52

$$\sum_{i=1}^{m}\sum_{j=1}^{n}(x_{ij}-\bar{x}_{\bullet\bullet})^2 = 652.52+664.32+820.52=2137.3$$

$$SSR=\sum_{i=1}^{m}\sum_{j=1}^{n}(x_{ij}-\bar{x}_{\bullet\bullet})^2 - SStr(тестове)-SStr(райони)=2137.3-1356.6-335.1=445.6$$

Проверяваме дали

$$F_{набл}=\frac{SStr(тестове)/(m-1)}{SSR/\left((n-1)(m-1)\right)}=\frac{1356.6/7}{445.6/14}=6.8>F_{1-\alpha}\left(m-1,(n-1)(m-1)\right)=F_{0.99}\left(7,14\right)=4.28$$

Понеже горното неравенство не е вярно, то отхвърляме основната хипотеза и приемаме, че тестовете не са с еднаква трудност.

Проверяваме дали

$$F_{набл}=\frac{SStr(райони)/(n-1)}{SSR/\left((n-1)(m-1)\right)}=\frac{335.1/2}{445.6/14}=5.26<F_{1-\alpha}\left(n-1,(n-1)(m-1)\right)=F_{0.99}\left(2,14\right)=6.36$$

Понеже горното неравенство е вярно, то приемаме основната хипотеза, т.е приемаме че студентите от трите региона са еднакво подготвени.

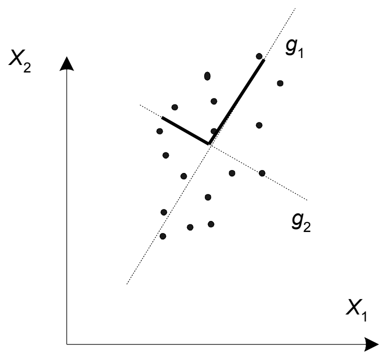
ТЕМА 11
Факторен анализ (метод на главните компоненти)

Въведение.

Да разгледаме случайните величини $X_1,\ldots,X_n$. Ако за всяка от тях направим *m* наблюдения, то получаваме матрицата от данни

$$X = \begin{pmatrix} x_{11} & \ldots & x_{1n} \\ \vdots & & \vdots \\ x_{m1} & \ldots & x_{mn} \end{pmatrix}.$$

Можем да считаме, че това са координатите на *m* точки в *n* мерно пространство. Нека облакът от *m* точки има форма на *n* мерен елипсоид. Ако *n* =2 елипсоида става елипса, фигура 1.



Фигура 1

От фигурата се вижда, че съществуват две прави g_1 и g_2, по които данните имат най-силно разсейване. Големината на това разсейване се характеризира с дължините на двете полуоси на елипсата (удебелени отсечки на фигура 1). Дължините на тези полуоси са точно сингулярните стойности на матрицата *X*, а прави g_1 и g_2 се наричат главни оси.

Забележка: В техниката дължините на полуосите(сингулярните стойности) се интерпретират, като най-големите усилвания на *n* мерен изходен сигнал, при входен сигнал с норма единица.

Сингулярно разлагане и главни компоненти.

Всяка матрицата $X_{m\times n}$ има така наречено сингулярно разлагане. За удобство ще считаме, че $m < n$. Тогава сингулярното разлагане на *X* е

$$X = U\,\Sigma\, V^T$$

където:

$U_{m\times m} = \begin{pmatrix} u_{11} & \ldots & u_{1m} \\ \vdots & & \vdots \\ u_{m1} & \ldots & u_{mm} \end{pmatrix}$ - квадратна матрица, чийто стълбове $U_i = \begin{pmatrix} u_{1i} \\ \vdots \\ u_{mi} \end{pmatrix}$ са собствените вектори на $X\,X^T$;

$\Sigma_{m\times n} = \begin{pmatrix} \sigma_1 & \ldots & 0 & 0 & \ldots & 0 \\ \vdots & & \vdots & \vdots & & \vdots \\ 0 & \ldots & \sigma_m & 0 & \ldots & 0 \end{pmatrix}$- матрица, чийто диагонални елементи $\sigma_1,\ldots,\sigma_m$ са различни от нула, а всички останало са нулеви. Диагоналните елементи се наричат сингулярни стойности и са квадратите на собствените стойности на матрицата $X X^T$. При това ранга на А е равен на броя на ненулевите сингулярни стойности;

$V_{n\times n} == \begin{pmatrix} v_{11} & \dots & v_{1n} \\ \vdots & & \vdots \\ v_{n1} & \dots & v_{nn} \end{pmatrix}$ - квадратна матрица, чийто стълбове $V_i = \begin{pmatrix} v_{1i} \\ \vdots \\ v_{ni} \end{pmatrix}$ са собствените вектори на $X^T X$.

Забележка: В случая когато за $m > n$, сингулярните стойности на X са квадратите на собствените стойности на матрицата $X^T X$, а матрицата Σ има *n* на брой диагонални елементи $\sigma_1, \dots, \sigma_n$.

Забележка: Понеже в статистиката обикновено се работи с матрици X имащи голяма размерност, то намирането на сингулярното разлагане на X на ръка, не е удобно. За това ние ще намираме сингулярното разлагане на X с помощта на командата “svd( )” от продукта MATLAB.

За случая $m < n$ ще намерим още едно представяне на сингулярното разлагане на *X*

$$X = U\,\Sigma\,V^T = \begin{pmatrix} u_{11} & \dots & u_{1m} \\ \vdots & & \vdots \\ u_{m1} & \dots & u_{mm} \end{pmatrix} \begin{pmatrix} \sigma_1 & \dots & 0 & 0 & \dots & 0 \\ \vdots & & \vdots & \vdots & & \vdots \\ 0 & \dots & \sigma_m & 0 & \dots & 0 \end{pmatrix} \begin{pmatrix} v_{11} & \dots & v_{1n} \\ \vdots & & \vdots \\ v_{n1} & \dots & v_{nn} \end{pmatrix}^T =$$

$$= \begin{pmatrix} \sigma_1 u_{11} & \dots & \sigma_m u_{1m} & 0 & \cdots & 0 \\ \vdots & & \vdots & \vdots & & \vdots \\ \sigma_1 u_{m1} & \dots & \sigma_m u_{mm} & 0 & \dots & 0 \end{pmatrix} \begin{pmatrix} v_{11} & \dots & v_{n1} \\ \vdots & & \vdots \\ v_{1n} & \dots & v_{nn} \end{pmatrix} =$$

$$= \begin{pmatrix} \sigma_1 u_{11} v_{11} + \sigma_2 u_{12} v_{12} + \dots + \sigma_m u_{1m} v_{1m} & \dots & \sigma_1 u_{11} v_{n1} + \sigma_2 u_{12} v_{n2} + \dots + \sigma_m u_{1m} v_{nm} \\ \vdots & & \vdots \\ \sigma_1 u_{m1} v_{11} + \sigma_2 u_{m2} v_{12} + \dots + \sigma_m u_{mm} v_{1m} & \dots & \sigma_1 u_{m1} v_{n1} + \sigma_2 u_{m12} v_{n2} + \dots + \sigma_m u_{mm} v_{nm} \end{pmatrix} =$$

$$= \sigma_1 \begin{pmatrix} u_{11} \\ \vdots \\ u_{m1} \end{pmatrix} \begin{pmatrix} v_{11} & \dots & v_{n1} \end{pmatrix} + \dots + \sigma_m \begin{pmatrix} u_{1m} \\ \vdots \\ u_{mm} \end{pmatrix} \begin{pmatrix} v_{1m} & \dots & v_{nm} \end{pmatrix}$$

Записваме горното в матричен вид и получаваме:

$$X = U\,\Sigma\,V^T = \sum_{i=1}^{m} \sigma_i\, U_i\, V_i^T$$, където U_i и V_i са стълбовете на U и V.

Елементите $\sigma_i\, U_i\, V_i^T$ на горната сума се наричат главни компоненти.

Понеже $\sigma_1 > \sigma_2 > \dots$ и понеже сингулярните стойности много бързо намаляват, то матрицата *X* може да се апроксимира само с първите няколко главни компонента. Обикновено се използват само тези главни компоненти, за които сумата на сингулярните им стойности е 90% от сумата на всички сингулярни стойност, т.е.

$$X = U\,\Sigma\,V^T \approx \sum_{i=1}^{k} \sigma_i\, U_i\, V_i^T$$, където $\frac{\sigma_1 + \sigma_1 + \dots + \sigma_k}{\sigma_1 + \sigma_1 + \dots + \sigma_m} = 90\%$.

От гледна точка на факторният анализ, броя *k* на главните компоненти в горното разлагане, може да се интерпретира, като броя на линейно независимите фактори, които влияят върху стойността на *X*.

Зад.1

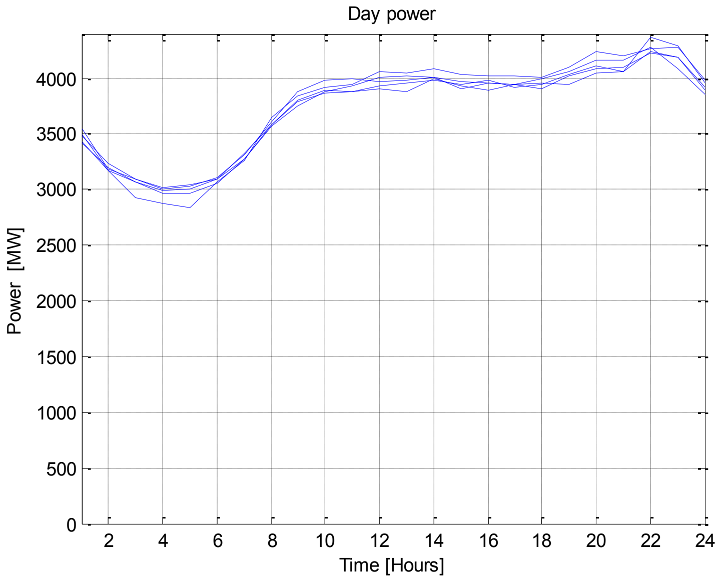
В долната таблица е даден денонощният график на потреблението на ел. енергия(MW) в България за пет летни работни дни

	Дни				
Часове	1	2	3	4	5
1	3430	3474	3419	3539	3498
2	3175	3231	3192	3190	3175
3	2933	3092	3064	3099	3067
4	2880	3001	2993	3016	2964
5	2838	3033	3010	3043	2960
6	3074	3111	3095	3100	3052
7	3262	3308	3277	3325	3272
8	3644	3606	3579	3584	3569
9	3846	3885	3792	3802	3752
10	3922	3981	3868	3890	3882
11	3940	3992	3876	3873	3936
12	4064	3972	3904	3924	4009
13	4050	3978	3884	3954	4024
14	4080	4007	3999	3983	4005
15	4038	3968	3924	3941	3909
16	4024	3960	3889	3979	3954
17	4020	3941	3949	3915	3938
18	4007	3902	3951	3940	3995
19	4101	4020	3943	4030	4062
20	4240	4085	4043	4116	4162
21	4197	4101	4065	4058	4160
22	4262	4232	4367	4241	4279
23	4278	4188	4292	4188	4082
24	3985	3891	3951	3915	3848

Да се разложи на главни компоненти товаровия график и да се интерпретира статистически (от гледна точка на факторния анализ) това разлагане.

Решение:

В случая матрицата X е с размерност $24^{\times}5$, като във всеки от петте стълба съдържа денонощния график на ел. товара за един ден. За да онагледим тези данни начертаваме всеки от тези стълбове на фигура 2



Фигура 2

След сингулярно разлагане на *X* за сингулярните стойности получаваме $\sigma_1 = 41213$, $\sigma_2 = 370$, $\sigma_3 = 259$, $\sigma_4 = 176$ и $\sigma_5 = 127$

Тогава $\dfrac{\sigma_1}{\sum_{i=1}^{5} \sigma_i} = 97.79\ \%$. Следователно матрицата *X* може да се приближи добре само с първата компонента, т.е

$$X_{24\times 5} \approx \sigma_1\, U_1\, V_1^T$$

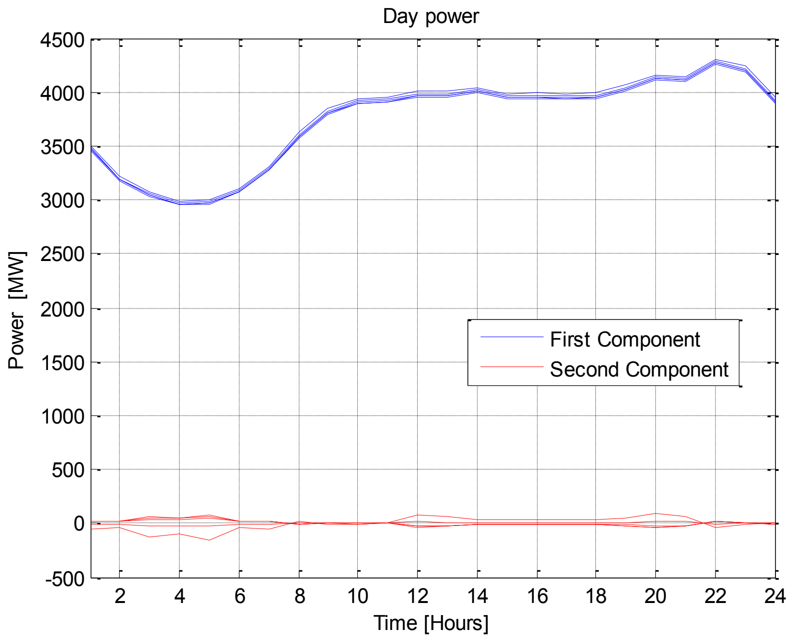
където

$\sigma_1 = 41213$

$U_1^T = (-0.1884\ \ -0.1732\ \ -0.1655\ \ -0.1612\ \ -0.1615\ \ -0.1675\ \ -0.1784\ \ -0.1951\ \ -0.2070$
$-0.2121\ \ -0.2129\ \ -0.2157\ \ -0.2158\ \ -0.2178\ \ -0.2146\ \ -0.2149\ \ -0.2145\ \ -0.2148\ \ -0.2187$
$-0.2240\ \ -0.2233\ \ -0.2320\ \ -0.2282\ \ -0.2126)$

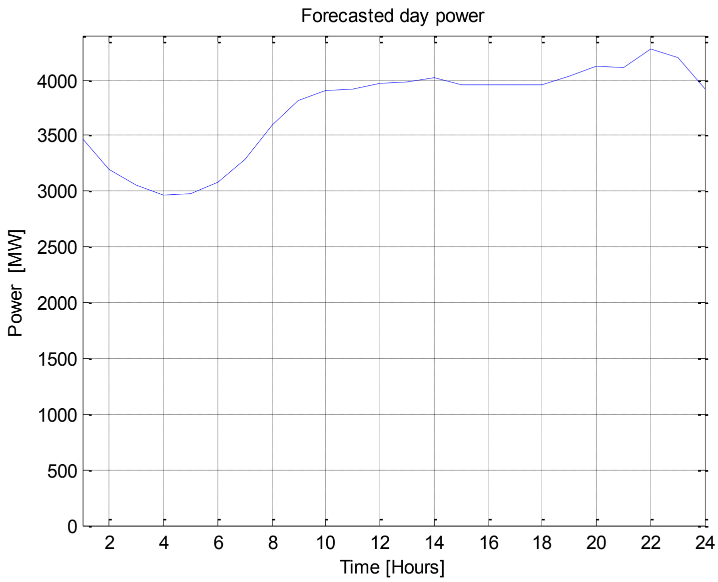
$V_1^T = (-0.4506\ \ -0.4480\ \ -0.4450\ \ -0.4463\ \ -0.4462)$.

За да покажем колко добре първата компонента приближава матрицата *X* начертаваме стълбовете на първата и втората компонента на фигура 3



Фигура 3

Статистическият извод от фигура 3, е че една компонента е напълно достатъчна, за да се приближи достатъчно добре матрицата *X*. Следователно съществува един главен фактор, които определя поведението на данните *X*. Това очевидно е стандартния работен режим на средно статистическия човек. Останалите компоненти могат да се смятат като случайни отклонения от този режим. В такъв случай ако усредним петте главни компоненти от фигура 3, то получената крива може да се използва като прогнозна стойност за денонощният товаров график за следващия работен ден, фигура 4.



Фигура 4

Приложение. Код на MATLAB, с които се извършват пресмятанията

```
clear all
x =[3430,3474,3419,3539,3498
3175,3231,3192,3190,3175
2933,3092,3064,3099,3067
2880,3001,2993,3016,2964
```

```
2838,3033,3010,3043,2960
3074,3111,3095,3100,3052
3262,3308,3277,3325,3272
3644,3606,3579,3584,3569
3846,3885,3792,3802,3752
3922,3981,3868,3890,3882
3940,3992,3876,3873,3936
4064,3972,3904,3924,4009
4050,3978,3884,3954,4024
4080,4007,3999,3983,4005
4038,3968,3924,3941,3909
4024,3960,3889,3979,3954
4020,3941,3949,3915,3938
4007,3902,3951,3940,3995
4101,4020,3943,4030,4062
4240,4085,4043,4116,4162
4197,4101,4065,4058,4160
4262,4232,4367,4241,4279
4278,4188,4292,4188,4082
3985,3891,3951,3915,3848]
%
figure(1)
for i=1:5
    plot([1:24],x(:,i));hold on
end
xlim([1,24])
ylim([0,4400])
title('Day power')
xlabel('Time [Hours]')
ylabel('Power  [MW]')
grid
%
[n,m]=size(x)
%
[u,d,v] = svd(x);
d
u(:,1)
v(:,1)
err=d(1,1)/(d(1,1)+d(2,2)+d(3,3)+d(4,4)+d(5,5))
%
k=min(n,m);
for i=1:k
    komp(:,:,i)=d(i,i)*u(:,i)*v(:,i)';
end
%
figure(2)
for i=1:k
    aa=plot([1:24],komp(:,i,1));hold on
end
%
for i=1:k
    bb=plot([1:24],komp(:,i,2),'r');hold on
end
xlim([1,24])
%ylim([0,4400])
legend([aa,bb],'First Component','Second Component',0)
```

```
title('Day power')
xlabel('Time [Hours]')
ylabel('Power  [MW]')
grid
%
middel=(komp(:,1,1)+komp(:,2,1)+komp(:,3,1)+komp(:,4,1)+…
komp(:,5,1))/5
%
figure(3)
for i=1:k
    plot([1:24],middel)
end
xlim([1,24])
ylim([0,4400])
title('Forecasted day power')
xlabel('Time [Hours]')
ylabel('Power  [MW]')
grid
```

Приложение 1. Някои по-важни разпределения

Биномно разпределение - дискретната случайна величина X е биномно разпределена с параметри n и p / $X \in B_i(n,p)$ / ако

$$P(X=k) = C_n^k p^k (1-p)^{n-k} \text{ за } k = 0,1,\dots,n.$$

Това разпределение има:

- математическо очакване $EX = np$;
- дисперсия $DX = np(1-p)$.

Поасоново разпределени - дискретната случайна величина X е Поасоново разпределена с параметър λ / $X \in P_o(\lambda)$ /, ако

$$P(X=k) = \frac{\lambda^k}{k!} e^{-\lambda} \text{ за } k = 0,1,\dots,\infty.$$

Това разпределение има:

- математическо очакване $EX = \lambda$;
- дисперсия $DX = \lambda$.

Нормално разпределение - непрекъснатата случайна величина X е равномерно разпределена с параметри μ и σ / $X \in N(\mu,\sigma)$ /, ако плътността й е

$$f(x) = \frac{1}{\sigma\sqrt{2\pi}} e^{-\frac{1}{2}\left(\frac{x-\mu}{\sigma}\right)^2}, \text{ за } -\infty < x < +\infty.$$

Това разпределение има:

- математическо очакване $EX = \mu$;
- дисперсия $DX = \sigma^2$.

Експоненциално разпределение - дискретната случайна величина X е експоненциално разпределена с параметър λ / $X \in Exp(\lambda)$ /, ако

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

Това разпределение има:

- математическо очакване $EX = 1/\lambda$;
- дисперсия $DX = 1/\lambda^2$.

Приложение 2. Разпределения получени чрез аритметични действия със случайни величини

1. Ако $X_1, \ldots, X_n \in N(0,1)$, тогава $\sum_{i=1}^{n} X_i^2 \in \chi^2(n)$ - се нарича χ^2 разпределение с *n* степени на свобода.

Това разпределение има:

- математическо очакване $E(\chi^2) = n$;
- дисперсия $D(\chi^2) = 2n$.

2. Ако $X \in N(0,1)$ и $Y \in \chi^2(n)$, тогава $\frac{X}{\sqrt{Y/n}} \in t(n)$ - се нарича *t* разпределение с *n* степени на свобода.

Това разпределение има:

- математическо очакване $E(t) = 0$;
- дисперсия $D(t) = \frac{n}{n-2}$.

3. Ако $X \in \chi^2(n)$ и $Y \in \chi^2(m)$, тогава $\frac{X/n}{Y/m} \in F(n,m)$ - се нарича *F* разпределение с *(n,m)* степени на свобода.

Това разпределение има:

- математическо очакване $E(F) = \frac{m}{m-2}$;
- дисперсия $D(F) = \frac{2m^2(m+n-2)}{n(n-2)^2(m-4)}$.

Приложение 3. Някой разпределения използвани в статистиката

Нека $\overline{X} = \dfrac{1}{n_x}\sum_{i=1}^{n_x} X_i$, $\overline{Y} = \dfrac{1}{n_y}\sum_{j=1}^{n_y} Y_j$, $s_x^2 = \dfrac{1}{n_x - 1}\sum_{i=1}^{n_x}(X_i - \overline{X})^2$ и $s_y^2 = \dfrac{1}{n_y - 1}\sum_{j=1}^{n_y}(Y_j - \overline{Y})^2$.

1. Ако $X_i \in N(\mu, \sigma)$, тогава

а) за $n_x < 30$ и σ известно, то $\dfrac{\overline{X} - \mu}{\sigma / \sqrt{n_x}} \in N(0,1)$.

б) за $n_x < 30$ и σ неизвестно, то $\dfrac{\overline{X} - \mu}{s_x / \sqrt{n_x}} \in t(n-1)$.

в) за μ неизвестно, то $\dfrac{(n_x - 1){s_x}^2}{\sigma^2} \in \chi^2(n-1)$.

г) за μ известно, то $\dfrac{n_x \quad {s_1}^2}{\sigma^2} \in \chi^2(n)$, където $s_1^2 = \dfrac{1}{n_x}\sum_{i=1}^{n_x}(X_i - \mu)^2$.

Забележка: Ако $n_x \geq 30$, тогава е достатъчно X_i да са независими и еднакво разпределени с $E(X_i) = \mu$ и $D(X_i) = \sigma^2$ и тогава

$$\frac{\overline{X} - \mu}{s_x / \sqrt{n_x}} \in N(0,1).$$

2. Ако $X_i \in N(\mu_x, \sigma_x)$, $Y_j \in N(\mu_y, \sigma_y)$, тогава:

а) за $n_x, n_y < 30$ и σ_x, σ_y известни, то $\dfrac{(\overline{X} - \overline{Y}) - (\mu_x - \mu_y)}{\sqrt{\sigma_x^2 / n_x + \sigma_y^2 / n_y}} \in N(0,1)$.

б) за $n_x, n_y < 30$ и σ_x, σ_y неизвестни и $\sigma_x = \sigma_y$, то

$\dfrac{(\overline{X} - \overline{Y}) - (\mu_x - \mu_y)}{s_p \sqrt{1/n_x + 1/n_y}} \in t(n_x + n_y - 2)$, където $s_p = \sqrt{\dfrac{(n_x - 1)s_x^2 + (n_y - 1)s_y^2}{n_x + n_y - 2}}$.

в) за $n_x, n_y < 30$ и σ_x, σ_y неизвестни и $\sigma_x \neq \sigma_y$, то

$\dfrac{(\overline{X} - \overline{Y}) - (\mu_x - \mu_y)}{\sqrt{s_x^2 / n_x + s_y^2 / n_y}} \in t(\nu)$, където ν е цялата част на $\dfrac{\left(s_x^2 / n_x + s_y^2 / n_y\right)^2}{\dfrac{(s_x^2 / n_x)^2}{n_x - 1} + \dfrac{(s_y^2 / n_y)^2}{n_y - 1}}$.

г) $\dfrac{s_x^2}{s_y^2} \in F(n_x - 1, n_y - 1)$.

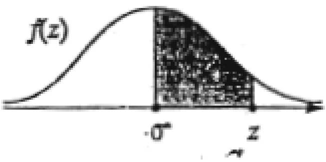
Забележка: Ако $n_x, n_y \geq 30$, тогава е достатъчно X_i и Y_j да са независими и еднакво разпределени с $E(X_i) = \mu_x$, $D(X_i) = {\sigma_x}^2$, $E(Y_i) = \mu_y$ и $D(Y_i) = {\sigma_y}^2$ и тогава

$$\frac{(\overline{X} - \overline{Y}) - (\mu_x - \mu_y)}{\sqrt{s_x^2 / n_x + s_y^2 / n_y}} \in N(0,1)$$

3. $\sum_{i=1}^{k} \dfrac{(m_i - n\,p_i)^2}{n\,p_i} \in \chi^2(k - r - 1)$, където m_i - извадкови честоти, $n\,p_i$ - теоретични честоти, k -броя на интервалите на групиране на извадката и r -броя на параметрите подлежащи на оценка на базата на данни от извадката.

Приложение 4. Таблица на стандартно нормално разпределение

Таблица А1. Стойности на $P(0 \le Z \le z)$ за сл. в. $Z \sim N(0,1)$



В таблицата са дадени стойностите на $P(0 \le Z \le z)$, т.е. защрихованото лице под графиката на стандартната нормална плътност $f(z)$. Например $P(0 \le Z \le 2,43) = 0,4925$.

z	0,00	0,01	0,02	0,03	0,04	0,05	0,06	0,07	0,08	0,09
0,0	0,0000	,0040	,0080	,0120	,0160	,0199	,0239	,0279	,0319	,0359
0,1	,0398	,0438	,0478	,0517	,0557	,0596	,0636	,0675	,0714	,0754
0,2	,0793	,0832	,0871	,0910	,0948	,0987	,1026	,1064	,1103	,1141
0,3	,1179	,1217	,1255	,1293	,1331	,1368	,1406	,1443	,1480	,1517
0,4	,1554	,1591	,1628	,1664	,1700	,1736	,1752	,1808	,1844	,1879
0,5	,1915	,1950	,1985	,2019	,2054	,2088	,2123	,2157	,2190	,2224
0,6	,2258	,2291	,2324	,2357	,2389	,2422	,2454	,2486	,2518	,2549
0,7	,2580	,2612	,2642	,2673	,2704	,2734	,2764	,2794	,2823	,2852
0,8	,2881	,2910	,2939	,2967	,2996	,3023	,3051	,3078	,3106	,3133
0,9	,3159	,3186	,3212	,3238	,3264	,3289	,3315	,3340	,3365	,3389
1,0	,3413	,3438	,3461	,3485	,3508	,3531	,3554	,3577	,3599	,3612
1,1	,3643	,3665	,3686	,3708	,3729	,3749	,3770	,3790	,3810	,3830
1,2	,3849	,3869	,3888	,3907	,3925	,3944	,3962	,3980	,3997	,4015
1,3	,4032	,4049	,4066	,4082	,4099	,4115	,4131	,4147	,4162	,4177
1,4	,4192	,4207	,4222	,4236	,4251	,4265	,4279	,4292	,4306	,4319
1,5	,4332	,4345	,4357	,4370	,4382	,4394	,4406	,4418	,4429	,4441
1,6	,4452	,4463	,4474	,4484	,4495	,4505	,4515	,4525	,4535	,4545
1,7	,4554	,4564	,4573	,4582	,4591	,4599	,4608	,4616	,4625	,4633
1,8	,4641	,4649	,4656	,4664	,4671	,4678	,4686	,4693	,4699	,4706
1,9	,4713	,4719	,4726	,4732	,4738	,4744	,4750	,4756	,4762	,4767
2,0	,4772	,4778	,4783	,4788	,4793	,4798	,4803	,4808	,4812	,4817
2,1	,4821	,4826	,4830	,4834	,4838	,4842	,4846	,4850	,4854	,4857
2,2	,4861	,4864	,4868	,4871	,4875	,4878	,4881	,4884	,4887	,4890
2,3	,4893	,4896	,4898	,4901	,4904	,4906	,4909	,4911	,4913	,4916
2,4	,4918	,4920	,4922	,4925	,4927	,4929	,4931	,4932	,4934	,4936
2,5	,4938	,4940	,4941	,4943	,4945	,4946	,4948	,4949	,4951	,4952
2,6	,4953	,4955	,4956	,4957	,4959	,4960	,4961	,4962	,4963	,4964
2,7	,4965	,4966	,4867	,4968	,4969	,4970	,4971	,4972	,4973	,4974
2,8	,4974	,4975	,4976	,4977	,4977	,4978	,4979	,4979	,4980	,4981
2,9	,4981	,4982	,4982	,4983	,4984	,4984	,4985	,4985	,4986	,4986
3,0	,4987	,4987	,4987	,4988	,4988	,4989	,4989	,4989	,4990	,4990
3,1	,4990	,4991	,4991	,4991	,4992	,4992	,4992	,4993	,4993	,4993
3,2	,4993	,4993	,4994	,4994	,4994	,4994	,4994	,4995	,4995	,4995
3,3	,4995	,4995	,4995	,4996	,4996	,4996	,4996	,4996	,4996	,4997
3,4	,4997	,4997	,4997	,4997	,4997	,4997	,4997	,4997	,4997	,4998
3,5	,4998	,4998	,4998	,4998	,4998	,4998	,4998	,4998	,4998	,4998
3,6	,4998	,4998	,4999	,4999	,4999	,4999	,4999	,4999	,4999	,4999
3,7	,4999	,4999	,4999	,4999	,4999	,4999	,4999	,4999	,4999	,4999
3,8	,4999	,4999	,4999	,4999	,4999	,4999	,4999	,4999	,4999	,4999
3,9	,5000	,5000	,5000	,5000	,5000	,5000	,5000	,5000	,5000	,5000

Приложение 5. Таблица на *t* разпределение

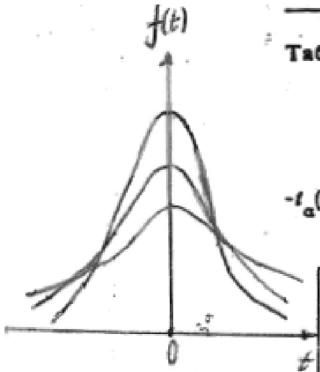


Таблица А2. Квантили $t_{1-\alpha}(\nu)$ (при едностранна) и $t_{1-\alpha/2}(\nu)$ (при двустранна критична област) на *t*-разпределението.

$$P\left\{ T \le t_{\alpha}(\nu) \right\} = \alpha; \quad P\left\{ t_{\alpha/2}(\nu) \le T \le t_{1-\alpha/2}(\nu) \right\} = 1 - \alpha$$

$-t_{\alpha}(\nu) = t_{1-\alpha}(\nu)$; За $\alpha = 0{,}05$ и $\nu = 8$: $t_{1-\alpha}(\nu) = t_{0,95}(8) = 1{,}859$.

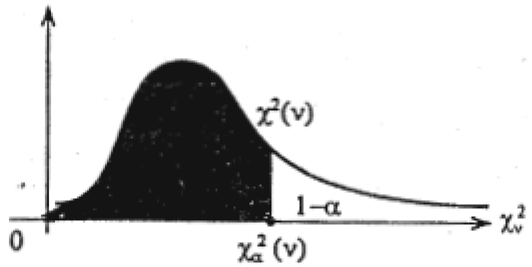
$\ge \le$ Двустранна критична област

	α	0,001	0,002	0,01	0,02	0,05	0,1	0,2
	$t_{1-\alpha/2}(\nu)$	$t_{0,9995}$	$t_{0,999}$	$t_{0,995}$	$t_{0,99}$	$t_{0,975}$	$t_{0,95}$	$t_{0,90}$
ν степени на свобода ν	1	636,619	318,309	63,657	31,820	12,706	6,314	3,077
	2	31,599	22,327	9,925	6,965	4,303	2,920	1,886
	3	12,924	10,214	5,841	4,541	3,182	2,353	1,638
	4	8,610	7,173	4,604	3,747	2,776	2,132	1,533
	5	6,867	5,893	4,032	3,365	2,571	2,015	1,476
	6	5,959	5,208	3,707	3,143	2,447	1,943	1,430
	7	5,408	4,785	3,499	2,998	2,365	1,895	1,415
	8	5,041	4,501	3,355	2,896	2,306	$\boxed{1,859}$	1,397
	9	4,781	4,297	3,250	2,821	2,262	1,833	1,383
	10	4,587	4,144	3,169	2,763	2,228	1,812	1,372
	11	4,437	4,025	3,106	2,718	2,201	1,796	1,363
	12	4,318	3,930	3,054	2,681	2,178	1,782	1,356
	13	4,221	3,852	3,012	2,650	2,160	1,771	1,350
	14	4,140	3,787	2,977	2,624	2,149	1,761	1,345
	15	4,073	3,733	2,947	2,602	2,131	1,753	1,341
	16	4,015	3,686	2,928	2,583	2,120	1,745	1,337
	17	3,965	3,646	2,898	2,567	2,110	1,740	1,333
	18	3,922	3,610	2,878	2,552	2,101	1,734	1,330
	19	3,883	3,579	2,861	2,540	2,093	1,729	1,327
	20	3,849	3,552	2,845	2,528	2,086	1,725	1,325
	21	3,819	3,527	2,831	2,518	2,080	1,721	1,323
	22	3,792	3,505	2,819	2,508	2,074	1,717	1,321
	23	3,768	3,485	2,807	2,500	2,069	1,714	1,319
	24	3,745	3,467	2,797	2,492	2,064	1,711	1,318
	25	3,725	3,450	2,787	2,485	2,059	1,708	1,316
	26	3,707	3,435	2,779	2,477	2,055	1,706	1,315
	27	3,690	3,421	2,770	2,473	2,052	1,703	1,314
	28	3,674	3,408	2,763	2,467	2,048	1,701	1,312
	29	3,659	3,396	2,756	2,462	2,045	1,699	1,311
	30	3,646	3,385	2,750	2,457	2,042	1,697	1,310
	40	3,551	3,307	2,704	2,423	2,021	1,684	1,303
	50	3,496	3,261	2,678	2,403	2,009	1,676	1,299
	60	3,460	3,232	2,660	2,390	2,001	1,670	1,296
	100	3,402	3,174	2,632	2,368	1,987	1,662	1,291
	500	3,310	3,107	2,586	2,334	1,965	1,648	1,283
	$t_{1-\alpha}(\nu)$	$t_{0,9995}$	$t_{0,999}$	$t_{0,995}$	$t_{0,99}$	$t_{0,975}$	$t_{0,95}$	$t_{0,9}$
	α	0,0005	0,001	0,005	0,01	0,025	0,05	0,1

$\ne$ Едностранна критична област

Приложение 6. Таблица на χ^2 разпределение

Таблица А3. Квантили $\chi^2_\alpha(\nu)$ на χ^2-разпределението с ν ст.св.



$$P\left\{\chi^2_\nu \leq \chi^2_\alpha(\nu)\right\} = \alpha;$$
$$P\left\{\chi^2_\nu > \chi^2_\alpha(\nu)\right\} = 1-\alpha.$$

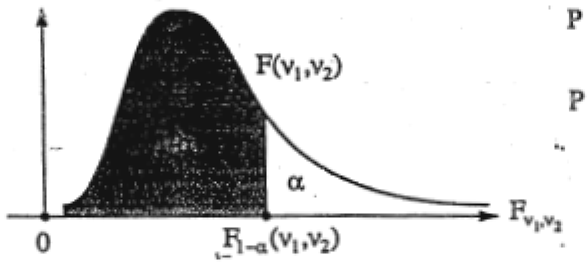
За $\alpha = 0{,}025$ и $\nu = 8$ имаме: $\chi^2_\alpha(\nu) = \chi^2_{0,025}(8) = 2{,}18$;

За $\alpha = 0{,}975$ и $\nu = 5$ имаме: $\chi^2_\alpha(\nu) = \chi^2_{0,975}(5) = 12{,}83$.

α	0,01	0,025	0,05	0,1	0,9	0,95	0,975	0,99
$\chi^2_\alpha(\nu)$ / ν степени на свобода	$\chi^2_{0,01}$	$\chi^2_{0,025}$	$\chi^2_{0,05}$	$\chi^2_{0,1}$	$\chi^2_{0,9}$	$\chi^2_{0,95}$	$\chi^2_{0,975}$	$\chi^2_{0,99}$
1	0,0002	0,0009	0,0039	0,0158	2,70	3,84	5,02	6,63
2	0,0201	0,0506	0,103	0,211	4,60	5,99	7,38	9,21
3	0,115	0,2158	0,352	0,584	6,25	7,81	9,35	11,34
4	0,297	0,4844	0,711	1,063	7,78	8,49	11,14	13,28
5	0,554	0,8312	0,145	1,610	9,24	11,07	$\boxed{12,83}$	15,09
6	0,87	1,24	1,64	2,20	10,64	12,59	14,45	16,81
7	1,24	1,69	2,17	2,83	12,02	14,07	16,01	18,47
8	1,65	$\boxed{2,18}$	2,73	3,49	13,36	15,51	17,53	20,09
9	2,09	2,70	3,33	4,17	14,68	16,92	19,02	21,67
10	2,56	3,25	3,94	4,87	15,99	18,31	20,48	23,21
11	3,05	3,82	4,57	5,58	17,27	19,67	21,92	24,27
12	3,57	4,40	5,23	6,30	18,55	21,03	23,34	26,22
13	4,11	5,01	5,89	7,04	19,81	22,36	24,74	27,69
14	4,66	5,63	6,57	7,79	21,06	23,68	26,12	29,14
15	5,23	6,26	7,26	8,55	22,31	25,00	27,49	30,58
16	5,81	6,91	7,96	9,31	23,54	26,30	28,84	32,00
17	6,41	7,56	8,67	10,09	24,77	27,59	30,19	33,41
18	7,01	8,23	9,39	10,86	25,99	28,87	31,53	34,80
19	7,63	8,91	10,12	11,65	27,20	30,14	32,85	36,19
20	8,26	9,59	10,85	12,44	28,41	31,41	34,17	37,57
21	8,90	10,28	11,59	13,24	29,61	32,67	35,48	38,93
22	9,54	10,98	12,34	14,04	30,81	33,92	36,78	40,29
23	10,20	11,69	13,09	14,85	32,01	35,17	38,08	41,64
24	10,86	12,40	13,84	15,65	33,20	36,41	39,36	42,97
25	11,52	13,12	14,61	16,47	34,38	37,62	40,65	44,31
26	12,20	13,84	15,38	17,29	35,56	38,88	41,92	45,64
27	12,88	14,57	16,15	18,11	36,74	40,11	43,19	46,96
28	13,56	15,31	16,93	18,94	37,92	41,34	44,46	48,28
29	14,26	16,05	17,71	19,77	39,09	42,56	45,72	49,59
30	14,95	16,79	18,49	20,60	40,26	43,77	46,98	50,89

Приложение 7. Таблица на *F* разпределение

Таблица А4. Квантили $F_{1-\alpha}(\nu_1,\nu_2)$ на F-разпределението.



$$P\left\{F_{\nu_1,\nu_2} \le F_{1-\alpha}(\nu_1,\nu_2)\right\} = 1-\alpha;$$

$$P\left\{F_{\nu_1,\nu_2} > F_{1-\alpha}(\nu_1,\nu_2)\right\} = \alpha;$$

$$F_{\alpha}(\nu_1,\nu_2) = \frac{1}{F_{1-\alpha}(\nu_2,\nu_1)}.$$

За $\alpha = 0{,}05$; $\nu_1 = 8$ и $\nu_2 = 7$: $F_{1-\alpha}(\nu_1,\nu_2) = F_{0,95}(8,7) = 3{,}73$.

Таблица за $\alpha = 0{,}05$.

Ст. св. ν_2 на		Степени на свобода ν_1 на числителя										
		1	2	3	4	5	6	7	8	9	10	12
	1	161	200	216	225	230	234	237	239	241	242	244
з	2	18,5	19,0	19,1	19,2	19,3	19,3	19,4	19,4	19,4	19,4	19,4
н	3	10,1	9,55	9,28	9,12	9,01	8,94	8,88	8,84	8,81	8,78	8,74
а	4	7,71	6,94	6,59	6,39	6,26	6,16	6,09	6,04	6,00	5,96	5,91
м	5	6,61	5,79	5,41	5,19	5,05	4,95	4,88	4,82	4,78	4,74	4,68
е	6	5,99	5,14	4,76	4,53	4,39	4,28	4,21	4,15	4,10	4,06	4,00
н	7	5,59	4,74	4,35	4,12	3,97	3,87	3,79	3,73	3,68	3,63	3,57
а	8	5,32	4,46	4,07	3,84	3,69	3,58	3,50	3,44	3,39	3,34	3,28
т	9	5,12	4,26	3,83	3,63	3,48	3,37	3,29	3,23	3,18	3,13	3,07
е	10	4,96	4,10	3,71	3,48	3,33	3,22	3,14	3,07	3,02	2,97	2,91
л	11	4,84	3,98	3,59	3,36	3,20	3,09	3,01	2,95	2,90	2,86	2,79
я	12	4,45	3,88	3,49	3,26	3,11	3,00	2,92	2,85	2,80	2,70	2,69

За $\alpha = 0{,}01$; $\nu_1 = 6$ и $\nu_2 = 9$: $F_{1-\alpha}(\nu_1,\nu_2) = F_{0,99}(6,9) = 5{,}80$.

Таблица за $\alpha = 0{,}01$.

Ст. св. ν_2 на		Степени на свобода ν_1 на числителя										
		1	2	3	4	5	6	7	8	9	10	12
	1	4052	4999	5403	5625	5764	5859	5928	5982	6022	6056	6106
з	2	98,5	99,0	99,2	99,3	99,3	99,3	99,4	99,4	99,4	99,4	99,4
н	3	34,1	30,8	29,5	28,7	28,2	27,9	17,7	27,5	27,3	27,2	27,0
а	4	21,2	18,0	16,7	16,0	15,5	15,2	15,0	14,8	14,7	14,6	14,4
м	5	16,2	13,3	12,1	11,4	11,0	10,7	10,4	10,3	10,1	10,0	9,89
е	6	13,4	10,9	9,78	9,15	8,75	8,47	8,26	8,10	7,98	7,87	7,72
н	7	12,2	9,55	8,45	7,85	7,46	7,19	7,00	6,84	6,71	6,62	6,47
а	8	11,3	8,65	7,59	7,01	6,93	6,37	6,19	6,08	5,91	5,82	5,67
т	9	10,6	8,02	6,99	6,42	6,06	5,80	5,62	5,47	5,35	5,38	5,11
е	10	10,0	7,58	6,55	5,99	5,64	5,39	5,21	5,06	4,93	4,85	4,71
л	11	9,65	7,20	6,22	5,67	5,32	5,07	4,88	4,74	4,63	4,54	4,40
я	12	9,33	6,93	5,95	5,41	5,06	4,82	4,65	4,50	4,39	4,30	4,18

Литература:

1. Н. Обрешков, “Теория на вероятностите”, Наука и изкуство, София, 1963
2. Б. Гнеденко, “Курс теории вероятностей”, Наука, Москва, 1969
3. А. А. Свешникова (редактор), “Сборник задач по теории вероятностей, математической статистике и теории случайных функции”, Наука, Москва, 1963
4. Н. Колев, “Приложна статистика –I ”, ТУ-София, 1993
5. Боян Димитров, Елена Каращранова, “Статистика учебник за студенти от нематематически специалности”, ВПИ "Неофит Рилски" – Благоевград, 1993
6. Христо Карапенев, “Теория на вероятностите и математическа статистика - модул от математика 4 ”, ТУ-София, 2000
7. Б. Гилев, “Висша математика“,Б. Гилев, “Висша математика – I и II част“, София, Партнер БГ, 2023, ISBN 978-619-7066-30-2, ISBN 978-619-7066-31-9, online: https://math.vtu.bg/zapiski.htm