Jihočeská univerzita v Českých Budějovicích
Pedagogická fakulta
Katedra fyziky

Statistické vyhodnocení experimentálních dat

Diplomová práce

Pavel Navrátil

Vedoucí práce:
RNDr. Pavel Kříž, Ph.D.

České Budějovice 2011

Anotace

NAVRÁTIL, PAVEL. Statistické vyhodnocení experimentálních dat. České Budějovice: Pedagogická fakulta Jihočeské univerzity, 2012.
Práce obsahuje teorii pravděpodobnosti a statistických souborů. Řešené a neřešené příklady z pravděpodobnosti, náhodné veličiny a rozdělení náhodné veličiny, náhodného vektoru, statistického souboru, regresní a korelační analýzy. Neřešené úholy jsou s výsledky.

Klíčová slova: pravděpodobnost, náhodná veličina, diskrétní rozdělení, spojité rozdělení, náhodný vektor, statistický soubor, regrese, korelace.

Abstract

NAVRÁTIL, PAVEL. Statistical interpretation of experimental data. České Budějovice: Pedagogical faculty, 2012.
This thesis contains theory of probability and statistical sets. Solved and unsolved problems of probability, random variable and distributions random variable, random vector, statistical sets, regression and correlation analysis. Unsolved problems contains solutions.

Keywords: probability, random variable, discrete distributions, continuous distributions, random vector, statistical set, regression, correlation.

Prohlášení

Prohlašuji že předloženou práci jsem vypracoval samostatně, pouze s použitím uvedené (citované) literatury.

Prohlašuji, že v souladu s § 47b zákona č. 111/1998 Sb. v platném znění souhlasím se zveřejněním své diplomové práce, a to v nezkrácené podobě fakultou elektronickou cestou ve veřejně přístupné části databáze STAG provozované Jihočeskou univerzitou v Českých Budějovicích na jejích internetových stránkách, a to se zachováním mého autorského práva k odevzdanému textu této kvalifikační práce. Souhlasím dále s tím, aby toutéž elektronickou cestou byly v souladu s uvedeným ustanovením zákona č. 111/1998 Sb. zveřejněny posudky školitele a oponentů práce i záznam o průběhu a výsledku obhajoby kvalifikační práce. Rovněž souhlasím s porovnáním textu mé kvalifikační práce s databází kvalifikačních prací Theses.cz provozovanou Národním registrem vysokoškolských kvalifikačních prací a systémem na odhalování plagiátů.

V Českých Budějovicích dne

....
Jméno a příjmení autora

Poděkování

Za cenné rady a připomínky při vypracování této diplomové práce děkuji RNDr. Pavlu Křížovi Ph.D.

Obsah

Seznam obrázků

Seznam tabulek

Úvod

Cílem mé diplomové práce bylo vytvořit studijní text k předmětu Statistické vyhodnocení experimentálních dat I. a II. Práce je určena studentům učitelství fyziky a měřící a výpočetní techniky, kteří jsou seznámeni se základy problematiky zpracování experimentálně určených dat a rozšiřují si tyto znalosti.

Práce je obsahově rozdělena na osm kapitol. Z počátku se zabývá zopakováním kombinatorických problémů a pravděpodobnosti. Poté tyto znalosti rozšiřuje o zpracování náhodné veličiny, náhodného vektoru a statistických souborů.

Práce je koncipována tak, aby student získal základní informace o teoretickém výkladu. Na konci každého teoretického výkladu jsou řešené příklady, které čtenáři pomohou v řešení úloh, které jsou pro kontrolu opatřeny výsledky.

Teoretické části této práce vycházejí z [1], [2] a [3]. Při sestavování neřešených úloh jsem použil zdroje [2], [4], [5] a [6].

Práce je vysázená systémem LaTEX, konkrétně v online verzi TEXonWeb. Obrázky a grafy jsou vytovřeny v programech GeoGebra 4 a Gimp.

Kapitola 1

Kombinatorika

Nejprve zopakujeme látku ze střední školy. Jedná se pouze o opakování, proto si uvedeme pouze definice a postup řešení si připomeneme na několika řešených úlohách a k procvičení dané látky poslouží několik neřešených úloh s výsledky.

Nejprve budeme zkoumat skupiny prvků v níž se jednotlivé prvky nemohou opakovat. Tyto skupiny prvků nazveme skupiny bez opakování. Pokračovat budeme skupinami v nichž se prvky mohou opakovat, tyto skupiny nazýváme skupiny s opakováním.

1.1 Základní kombinatorická pravidla

1.2 Faktoriál

Tento symbol zavádíme k ulehčení výpočtů.

Definice 1.1 (Faktoriál) Pro každé přirozené číslo n definujeme:

$$n! = 1 \cdot 2 \cdot 3 \cdot \ldots \cdot (n-1) \cdot n$$

Vlastnosti 1.1

1. $0! = 1$
2. *Rozklad faktoriálu* $n! = n \cdot (n-1)!$
3. *Rozklad faktoriálu* $(n+1)! = (n+1) \cdot n!$

■ **Příklad 1.1** Zjednodušte výraz:

$$\frac{(n+2)!}{n!} - 2\frac{(n+1)!}{(n-1)!} + \frac{n!}{(n-2)!}$$

Řešení

Nejprve prozkoumáme jednotlivé zlomky. Je potřeba se rozhodnout, který ze členů (čitatel, jmenovatel) je větší. Větší z členů vždy rozložíme tak, aby obsahoval stejný faktoriál jako menší člen. Poté už jen zkrátíme stejné členy a provedeme zbylé algebraické úpravy.

$$\begin{aligned}
\frac{(n+2)!}{n!} - 2\frac{(n+1)!}{(n-1)!} + \frac{n!}{(n-2)!} &= \frac{(n+2)\cdot(n+1)\cdot n!}{n!} - 2\frac{(n+1)\cdot n\cdot(n-1)!}{(n-1)!} + \\
&\quad + \frac{n\cdot(n-1)\cdot(n-2)!}{(n-2)!} = \\
&= (n+2)\cdot(n+1) - 2\cdot(n+1)\cdot n + n\cdot(n-1) = \\
&= n^2 + 3n + 2 - 2n^2 - 2n + n^2 - n = 2
\end{aligned}$$

Úlohy:

1. Zjednodušte a určete podmínky:

$$\frac{1}{n!} - \frac{n}{(n+1)!} =$$

2. Zjednodušte a určete podmínky:

$$\frac{n}{(n-3)!} - \frac{1}{(n-4)!} =$$

3. Zjednodušte a určete podmínky:

$$2\cdot\frac{n^2-16}{(n+4)!} + \frac{n^2+5}{(n+3)!} + \frac{3}{(n+2)!} =$$

4. Zjednodušte a určete podmínky:

$$\frac{2}{n!} - \frac{2n}{(n+1)!} - \frac{2n+4}{(n+2)!} =$$

5. Zjednodušte a určete podmínky:
$$\frac{(n-1)!}{3n!}+\frac{n!}{4\cdot(n+1)!}=$$
6. Zjednodušte a určete podmínky:
$$\frac{(n+5)!}{(n+3)!}-2\cdot\frac{(n+4)!}{(n+2)!}+\frac{(n+3)!}{(n+1)!}=$$
7. Zjednodušte a určete podmínky:
$$\frac{(n+1)!}{n!}+4\cdot\frac{(n+1)!}{(n-1)!}-\frac{9n!}{(n-1)!}=$$
8. Rozhodněte, které z čísel A a B je větší:
$$\begin{aligned} A &= 70!+73! \\ B &= 71!+72! \end{aligned}$$
9. Dokažte, že platí nerovnost:
$$\frac{1000!+1003!}{1001!+1002!}>1$$
10. Dokažte, že platí rovnost:
$$n!\cdot(n+1)+n\cdot(n-1)!-(n-1)!\cdot\left(n^2+n\right)=n!$$
11. Řešte rovnici s neznámou n:
$$5\cdot(n+1)!=(n+2)!$$
12. Řešte rovnici s neznámou n:
$$(n+1)!-16\cdot(n-1)!=n!$$
13. Řešte rovnici s neznámou n:
$$\frac{10-17n}{(n+1)!}+\frac{4}{(n-1)!}=0$$
14. Řešte nerovnici s neznámou n:
$$72n!<(n+2)!$$
15. Řešte nerovnici s neznámou n:
$$\frac{n!}{(n-2)!}+24\geq 10n$$

Řešení:

1. $\frac{1}{(n+1)!}$, podmínka: $n \geq 0$
2. $\frac{3}{(n-4)!}$, podmínka: $n \geq 4$
3. $\frac{1}{(n+1)!}$, podmínka: $n \geq -2$
4. 0, podmínka: $n \geq 0$
5. $\frac{7n+4}{12n^2+12n}$, podmínka: $n \geq 1$
6. 2, podmínka: $n \geq -1$
7. $(2n-1)^2$, podmínka: $n \geq 1$
8. $A > B$

11. $n = 3$
12. $n = 4$
13. $n = 2$
14. $n > 7$
15. $n \in \{2; 3\} \wedge n \geq 8$

1.3 Kombinační čísla

Definice 1.2 (Kombinační číslo) Pro všechna nezáporná celá čísla n a k, která splňují podmínku $k \leq n$ platí:

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

Vlastnosti 1.2 (Kombinační čísla)

1. *Pro všechna nezáporná n, k jestliže $k \leq n$ platí:*

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

Důkaz:

$$\begin{aligned}\binom{n}{n-k} &= \frac{n!}{(n-(n-k))! \cdot (n-k)!} = \frac{n!}{(n-n+k)! \cdot (n-k)!} = \frac{n!}{k! \cdot (n-k)!} = \\ &= \frac{n!}{(n-k)! \cdot k!} = \binom{n}{k}\end{aligned}$$

2. *Pro všechna přirozená n platí:*

$$\binom{n}{0} = \binom{n}{n} = 1 \tag{1.1}$$

Důkaz:

$$\begin{aligned}\binom{n}{0} &= \frac{n!}{(n-0)! \cdot 0!} = \frac{n!}{n! \cdot 0!} = \frac{1}{1 \cdot 1} = 1 \\ \binom{n}{n} &= \frac{n!}{(n-n)! \cdot n!} = \frac{n!}{0! \cdot n!} = \frac{1}{1 \cdot 1} = 1\end{aligned}$$

3. *Pro všechna přirozená n platí:*

$$\binom{n}{1} = n$$

Důkaz:

$$\binom{n}{1} = \frac{n!}{(n-1)! \cdot 1!} = \frac{n \cdot (n-1)!}{(n-1)! \cdot 1!} = \frac{n}{1 \cdot 1} = n$$

4. *Jestliže $n = k = 0$ platí:*

$$\binom{0}{0} = 1 \tag{1.3}$$

Důkaz:

$$\binom{0}{0} = \frac{0!}{(0-0)!\cdot 0!} = \frac{0!}{0!\cdot 0!} = \frac{1}{1\cdot 1} = 1$$

5. *Pro všechna nezáporná celá n, k jestliže $k \leq n$ platí:*

$$\binom{n}{k} + \binom{n}{k+1} = \binom{n+1}{k+1} \tag{1.4}$$

Důkaz:

$$\begin{aligned}
\binom{n}{k} + \binom{n}{k+1} &= \frac{n!}{(n-k)!\cdot k!} + \frac{n!}{(n-[k+1])!\cdot (k+1)!} = \\
&= \frac{n!}{(n-k)\cdot(n-k-1)!\cdot k!} + \frac{n!}{(n-[k+1])!\cdot(k+1)\cdot k!} = \\
&= \frac{n!}{(n-k)\cdot(n-[k+1])!\cdot k!} + \frac{n!}{(n-[k+1])!\cdot(k+1)\cdot k!} = \\
&= \frac{n!}{(n-[k+1])!\cdot k!}\cdot\left(\frac{1}{n-k} + \frac{1}{k+1}\right) = \\
&= \frac{n!}{(n-[k+1])!\cdot k!}\cdot\frac{k+1+n-k}{(n-k)\cdot(k+1)} = \\
&= \frac{n!}{(n-[k+1])!\cdot k!}\cdot\frac{n+1}{(n-k)\cdot(k+1)} = \\
&= \frac{n!\cdot(n+1)}{(n-k-1)!\cdot k!\cdot(n-k)\cdot(k+1)} = \\
&= \frac{(n+1)\cdot n!}{(n-k)\cdot(n-k-1)!\cdot(k+1)\cdot k!} = \\
&= \frac{(n+1)!}{(n-k)!\cdot(k+1)!} = \frac{(n+1)!}{([n+1]-[k+1])!\cdot(k+1)!} = \\
&= \binom{n+1}{k+1}
\end{aligned}$$

■ **Řešený příklad 1.1** Vyjádřete jedním kombinačním číslem součty:

a. $\binom{87}{24} + \binom{87}{64}$

b. $\binom{20}{20} + \binom{25}{24} + \binom{26}{24} + \binom{27}{24}$
Řešení:

Poznámka: Vyjádření výsledku pouze pomocí kombinačního čísla využíváme především v případě, že $n \geq k \geq 69$. Taková kombinační čísla běžná kalkulačka nespočítá. Pokud bychom i přesto chtěli získat celočíselný výsledek využijeme k výpočtu počítačový software např. tabulkový editor MS Office Excell, OpenOffice CALC, MatLab, Maple, Mathematica, Derive, ...

a. Nejprve použijeme vlastnost 1.2-1. a poté 1.2-5.

$$\binom{87}{24} + \binom{87}{64} = \binom{87}{24} + \binom{87}{87-64} = \binom{87}{24} + \binom{87}{23} = \binom{88}{24}$$

b. Nejprve použijeme vlastnost 1.2-2. a poté 1.2-5.

$$\begin{aligned}
\binom{20}{20} + \binom{25}{24} + \binom{26}{24} + \binom{27}{24} &= 1 + \binom{25}{24} + \binom{26}{24} + \binom{27}{24} = \\
&= \binom{25}{25} + \binom{25}{24} + \binom{26}{24} + \binom{27}{24} = \\
&= \binom{26}{25} + \binom{26}{24} + \binom{27}{24} = \\
&= \binom{27}{25} + \binom{27}{24} = \binom{28}{25}
\end{aligned}$$

Námět na samostatnou práci Najděte v Pascalově trojúhelníku výše zmíněné vlastnosti kombinačních čísel.

$$
\begin{array}{ccccccccccccccc}
 & & & & & & & \binom{1}{0} & & & & & & & \\
 & & & & & & \binom{1}{0} & & \binom{1}{1} & & & & & & \\
 & & & & & \binom{2}{0} & & \binom{2}{1} & & \binom{2}{2} & & & & & \\
 & & & & \binom{3}{0} & & \binom{3}{1} & & \binom{3}{2} & & \binom{3}{3} & & & & \\
 & & & \binom{4}{0} & & \binom{4}{1} & & \binom{4}{2} & & \binom{4}{3} & & \binom{4}{4} & & & \\
 & & \binom{5}{0} & & \binom{5}{1} & & \binom{5}{2} & & \binom{5}{3} & & \binom{5}{4} & & \binom{5}{5} & & \\
 & \binom{6}{0} & & \binom{6}{1} & & \binom{6}{2} & & \binom{6}{3} & & \binom{6}{4} & & \binom{6}{5} & & \binom{6}{6} & \\
\binom{7}{0} & & \binom{7}{1} & & \binom{7}{2} & & \binom{7}{3} & & \binom{7}{4} & & \binom{7}{5} & & \binom{7}{6} & & \binom{7}{7} \\
 & & & & & & & \vdots & & & & & & &
\end{array}
$$

Úlohy:

1. Řešte rovnici s neznámou x:
$$\binom{10}{4}x = \binom{12}{6}$$

2. Dokažte, že platí rovnost:
$$\left[\binom{11}{7}\right]^2 - \left[\binom{10}{7}\right] = \binom{10}{4} \cdot \left[\binom{10}{6} + \frac{1}{3} \cdot \binom{6}{5} \cdot \binom{10}{3}\right]$$

3. Dokažte, že platí rovnost:
$$\binom{15}{3} - \binom{10}{3} - \binom{5}{3} = \binom{10}{2} \cdot 5 + 10 \cdot \binom{5}{2}$$

4. Rozhodněte, které z čísel C a D je větší:
$$C = \binom{120}{60}$$
$$D = \binom{120}{61}$$

5. Dokažte, že platí rovnost:
$$n^3 = 6 \cdot \binom{n}{3} + 6 \cdot \binom{n}{2} + \binom{n}{1}$$

6. Dokažte, že platí rovnost:
$$\frac{3}{2} \cdot \binom{3n-1}{n} = \binom{3n}{n}$$

7. Zjednodušte a výsledek zapište jako kombinační číslo:
$$\binom{10}{1} + \binom{10}{0} + \binom{11}{9} =$$

8. Zjednodušte a výsledek zapište jako kombinační číslo:
$$\binom{12}{3} + \binom{4}{3} - \binom{12}{9} =$$

9. Zjednodušte a výsledek zapište jako kombinační číslo:
$$\binom{3}{3}+\binom{4}{3}+\binom{5}{3}+\binom{6}{3}=$$

10. Zjednodušte a výsledek zapište jako kombinační číslo:
$$\frac{n!}{(n-k)!\cdot k!}+\frac{n!}{(n-k+1)!\cdot(k-1)!}=$$

11. Řešte rovnici s neznámou x:
$$\binom{6}{3}=2\cdot\left[x+\binom{5}{1}\right]+\frac{1}{2}\cdot\binom{4}{3}$$

12. Řešte rovnici s neznámou x:
$$\left[x+\binom{1}{1}\right]\cdot\left[x-\binom{2}{2}\right]=\binom{3}{0}+\binom{3}{1}+\binom{3}{2}+\binom{3}{3}$$

13. Řešte rovnici s neznámou x:
$$\left[\binom{x}{2}\right]^2-2\cdot\binom{6}{5}\cdot\binom{x}{2}-\binom{7}{3}-\binom{5}{3}=0$$

14. Řešte rovnici s neznámou x:
$$\binom{8}{x}=\binom{8}{5}$$

15. Řešte rovnici s neznámou x:
$$\binom{10}{x}=45$$

16. Řešte rovnici s neznámou x:
$$\binom{x}{1}+\binom{x-3}{x-4}=2x-3$$

17. Řešte rovnici s neznámou x:
$$\binom{x-1}{x-3}-2\cdot\binom{x-2}{x-4}=0$$

18. Řešte nerovnici s neznámou x:
$$\binom{x+1}{x}+2x<50$$

Řešení:

1. $x = \frac{22}{5}$
4. $C > D$
7. $\binom{12}{2}$
8. $\binom{4}{3}$
9. $\binom{7}{4}$
10. $\binom{n+1}{k}$
11. $x = 4$
12. $x_1 = 3$, $x_2 = -3$
13. $x = 6$
14. $x_1 = 5$, $x_2 = 3$
15. $x_1 = 2$, $x_2 = 8$
16. $x \geq 4$
17. $x = 5$
18. $x < \frac{49}{3}$

1.4 Skupiny bez opakování

Skupiny bez opakování ještě můžeme rozdělit na dvě malé podskupiny a to podle toho, jestli nám záleží na pořadí prvků nebo ne.

Definice 1.3 (Variace) K-členná variace z n prvků je uspořádaná k-tice sestavená z těchto prvků tak, že každý se v ní vyskytuje nejvýše jednou.

Definice 1.4 (Počet variací) Počet všech variací k-té třídy z n různých prvků je dán vztahem.

$$V_k\,(n) = \frac{n!}{(n-k)!} = n \cdot (n-1) \cdot (n-2) \cdot \ldots \cdot (n-k+1) \tag{1.5}$$

Nejlépe si definici objasníme na příkladu.

■ **Řešený příklad 1.2** Na houpačku máme umístit tři různá závaží. Na výběr máme pět kusů závaží o různých hmotnostech označíme je m_1, m_2, m_3, m_4 a m_5. Určete:

a. počet všech možností jak můžeme na houpačku tato závaží rozložit,

b. počet všech možností jestliže chceme aby na houpačce bylo závaží m_1,

c. v kolika případech se nevyskytuje závaží m_1.

Řešení

a. Můžeme zvolit dvě různé strategie řešení, ukážeme si obě dvě.

1. Máme umístit tři závaží na houpačku. Na pomyslné první místo vybíráme z pěti závaží, na druhém místě se již nemůže opakovat závaží zvolené na první místo, protože jsme měli pouze jedno. To znamená výběr je již pouze ze čtyř závaží. Na třetím místě rozhodně nebude závaží z prvního ani druhého místa, proto vybíráme pouze ze tří závaží. Počet všech možných uspořádání je:

$$5 \cdot 4 \cdot 3 = 60$$

Poznámka: Tento způsob řešení je nevyhovující při určování počtu všech skupin, kde se vyskytuje více prvků.

2. Druhý způsob. Máme umístit tři závaží z nichž se ani jedno neopakuje a záleží na jejich pořadí. Na výběr máme pět závaží. To znamená, že tvoříme 3-člennou variaci bez opakování z 5-ti prvků.

$$V_3(5) = \frac{5!}{(5-3)!} = \frac{5!}{3!} = \frac{5 \cdot 4 \cdot 3 \cdot 2!}{2!} = 60$$

b. Chceme aby závaží m_1 bylo na houpačce. Toto závaží můžeme umístit na tři různé pozice, protože na pořadí závaží záleží je nutné nejprve vybrat počet všech možností jak umístit závaží m_1. Na toto závaží máme tři volná místa, jejich počet je dán:

$$V_1(3) = \frac{3!}{(3-1)!} = \frac{3!}{2!} = 3.$$

Na zbylá dvě místa můžeme vybírat ze čtyř závaží na jejichž pořadí záleží a počet těchto možností získáme takto

$$V_2(4) = \frac{4!}{(4-2)!} = \frac{4!}{2!} = 12.$$

Oba výběry na sobě závisí, proto celkový počet možných výběrů získáme takto:

$$V_1(3) \cdot V_2(4) = 3 \cdot 12 = 36.$$

Celkem máme 36 možností jak rozmístit závaží.

c. Stejně jako v případě a. máme dvě možnosti jak k danému případu přistupovat.

1. Určíme počet možností rozmístění závaží mezi nimiž není to co jsme označili m_1. Vybíráme ze čtyř závaží, na jejichž pořadí záleží, které můžeme umístit na tři místa. Počet těchto možností je dán takto:

$$V_3(4) = \frac{4!}{(4-3)!} = \frac{4!}{1!} = 24.$$

2. Druhá možnost je využití předchozích dvou výpočtů. Chceme počet všech výběrů v nichž se nevyskytuje závaží m_1. Víme, že počet všech výběrů bez omezení je $V_3(5)$, počet všech možností se závažím m_1 je $V_1(3) \cdot V_2(4)$. Z celkového počtu všech výběrů odstraníme ty co našemu předpokladu nevyhovují, tedy:

$$V_3(5) - V_1(3) \cdot V_2(4) = 60 - 36 = 24.$$

Počet možností jak vybrat závaží mezi nimiž není m_1 je 24.

■ **Řešený příklad 1.3** Určete počet prvků tak, aby počet pětičlenných variací z nich vytvořených byl 30-krát větší než počet tříčlenných variací.

Řešení:

$$\begin{aligned}
V_5(n) &= 30 \cdot V_3(n) \\
\frac{n!}{(n-5)!} &= 30 \cdot \frac{n!}{(n-3)!} \\
\frac{n \cdot (n-1) \cdot (n-2) \cdot (n-3) \cdot (n-4) \cdot (n-5)!}{(n-5)!} &= 30 \cdot \frac{n \cdot (n-1) \cdot (n-2) \cdot (n-3)!}{(n-3)} \\
n \cdot (n-1) \cdot (n-2) \cdot (n-3) \cdot (n-4) &= 30 \cdot n \cdot (n-1) \cdot (n-2) \\
(n-3) \cdot (n-4) &= 30 \\
n^2 - 7n + 12 - 30 &= 0 \\
n^2 - 7n - 18 &= 0 \\
n_{1,2} &= \frac{7 \pm \sqrt{(-7)^2 - 4 \cdot 1 \cdot (-18)}}{2} \\
n_{1,2} &= \frac{7 \pm \sqrt{121}}{2} \\
n_{1,2} &= \frac{7 \pm 11}{2} \\
n_1 &= 9 \\
n_2 &= -2 \notin \mathbb{N}
\end{aligned}$$

Výsledkem je pouze $n_1 = 9$, protože $n_2 = -2$ není přirozené číslo. Nelze vytvořit skupinu o minus dvou prvcích!

Úlohy:

1. Tři studenti VŠ jdou na oběd do menzy. V menze je následující výběr: nápoje - neslazený čaj, slazený čaj, sirup, nealkoholické pivo, pomerančový džus a minerální vodu. Polévky: česneková, bramborová, slepičí a hrachová. Hlavní jídlo: sekaná pečeně, svíčková omáčka, rajská omáčka, vepřový řízek, kuřecí řízek a pečené kuře. Příloha: vařené brambory, vařená rýže, bramborové hranolky, bramborové krokety a zeleninová obloha. Dezert: borůvkový, broskvový, jablečný a jahodový koláč.Ani jeden student nechce mít k obědu stejné jídlo jako ostatní.

   a. Rozhodněte kolik nejvýše možností mají na výběr.

b. Rozhodněte kolik možností studenti mají jestliže první z nich nejí bramborovou polévku, vepřový řízek a jablečný koláč.

2. Pedagogický sbor tvoří 7 žen a 4 muži. Určete kolika způsoby lze vybrat zástupce předmětové komise:

   a. předsedu, místopředsedu, zapisovatele, jednatele a pokladníka,

   b. daných pět zástupců tak, aby předseda byla žena a jednatel byl muž,

   c. zástupce tak, aby nejvýše dva byli muži,

   d. zástupce tak, aby alespoň jeden byla žena.

3. Určete počet všech možností jak se student VŠ může obléknout na zkoušku, jestliže má na výběr tři páry společenské obuvi, dva obleky, šest košil a osm kravat.

4. Kolik různých přirozených čtyřciferných čísel s různými ciframi lze sestavit z cifer 1, 2, 3, 4, 5? Kolik z nich je dělitelných 5? Kolik z nich je lichých?

5. Určete počet všech přirozených čísel větších než 2000, v jejichž zápisech se vyskytují cifry 1, 2, 4, 6, 8, a to každá nevjvýše jednou.

6. Z kolika prvků lze vytvořit 992 variací druhé třídy bez opakování?

7. Zvětší-li se počet prvků o 5, zvětší se počet variací druhé třídy bez opakování vytvořených z těchto prvků o 1170. Určete původní počet prvků.

8. Zmenší-li se počet prvků o 27, zmenší se počet variací druhé třídy bez opakování vytvořených z těchto prvků desetkrát. Určete původní počet prvků.

Řešení:

1. a. 2880, b. 1350
2. a. 55440, b. 14112, c. 8400, d. 5614
3. 288
4. 72
5. 216
6. 32
7. $n = 115$
8. 40

Definice 1.5 (Permutace - variace) Permutace z n prvků je každá n-členná variace z těchto prvků.

Definice 1.6 (Permutace - počet) Počet všech permutací z n prvků je dán takto:

$$P(n) = n! \tag{1.6}$$

Definice 1.7 (Permutace) Permutace z n prvků je uspořádaná n-tice sestavená z těchto prvků tak, že se v ní každý prvek vyskytuje právě jednou.

■ **Řešený příklad 1.4** Určete kolika způsoby se mohou studenti VŠ (Andrea, Eliška, Jiří, Karel a Marek) posadit na zadní sedadlo autobusu (5 míst) jestliže:

a. každý student si může sednou na místo jaké chce,

b. Eliška a Jiří chtějí sedět vedle sebe,

c. Andrea a Marek chtějí sedět vedle sebe a Karel chce sedět u okna.

Řešení:

a. Obsazujeme pět sedadel, tedy tvoříme pětičlennou skupinu a obsazujeme ji pěti studenty, tedy pěti prvky. Studenty rozlišujeme podle jména, tedy jedná se o uspořádanou pětici. Na základě těchto znalostí rozhodneme, že se jedná o permutace (bez opakování). Počet všech možností tedy vypadá takto:

$$P(5) = 5! = 5 \cdot 4 \cdot 3 \cdot 2 \cdot 1 = 120$$

Existuje 120 způsobů jak se mohou studenti posadit.

b. Dva studenti chtějí sedět vedle sebe a ostatním nezáleží na tom vedle koho budou sedět. Nejprve vybereme kolika způsoby se mohou daní dva studenti vedle sebe posadit. Vybíráme dvojici sedadel, tedy tvoříme dvoučlennou skupinu a obsazujeme ji dvěma studenty, tedy dvěma prvky. Studenty rozlišujeme podle jména, to znamená, že se jedná o uspořádanou dvojici. Na základě těchto znalostí rozhodneme, že se jedná o permutace (bez opakování). Počet možností vypadá takto:

$$P(2) = 2! = 2 \cdot 1 = 2$$

Poznámka: Možnost usazení Elišky a Jiřího můžeme určit také logickou zkušeností, kdy můžeme usadit buď dvojici Eliška a Jiří nebo Jiří a Eliška. Je tedy jasné, že máme dvě možnosti jak tyto dva studenty usadit.

Zatím jsme rozhodli pouze o způsobu rozsazení daných dvou studentů, ale nerozhodli jsme o jejich pozici na pětičlenném sedadle. Dvě nyní obsazená místa můžeme myšlenkově spojit a tím i studenty na nich sedící. Celkový počet sedadel se snížil z pěti na čtyři a počet studentů se také snížil z pěti na čtyři. Počet těchto možností je:

$$P\left(4\right) = 4! = 4 \cdot 3 \cdot 2 \cdot 1 = 24$$

Teprve po výběru usazení dvojice Eliška, Jiří vybereme jejich umístění na sedadlech, proto kombinatorické pravidlo součinu (druhý výběr nezávisí na prvním výběru):

$$P\left(2\right) \cdot P\left(4\right) = 2 \cdot 24 = 48$$

Existuje 48 způsobů jak se mohou studenti posadit tak, aby Eliška a Jiří seděli vedle sebe.

c. Chceme aby jeden student seděl na kraji sedačky. V zadání ovšem není řečeno na kterém. Můžeme ho tedy posadit na dva okraje $V\left(1;2\right)$. Můžeme říci, že jsme tohoto studenta již pevně usadili. Dále máme usadit dvojici studentů stejným způsobem jako v případě *b.*. Postup proto uvedeme ve zkrácené verzi. Existuje právě $P\left(2\right)$ možností jak usadit tyto dva studenty mezi sebou. Celkový počet míst i studentů se již snížil z původních pěti na čtyři (usazením studenta na okraj sedačky). Dále tento počet snížíme myšlenkovým spojením dvou míst a dvojicí studentů. Počet míst i studentů se tedy snížil na tři, proto je počet možností $P\left(3\right)$. Jednotlivé výběry na sobě závisí, proto znovu kombinatorické pravidlo součinu. Celkový počet možností:

$$V\left(1;2\right) \cdot P\left(2\right) \cdot P\left(3\right) = \frac{2!}{(2-1)! \cdot 1!} \cdot 2! \cdot 3! = 2 \cdot 2 \cdot 3 = 12$$

Existuje 12 způsobů jak se mohou studenti posadit tak, aby Andrea a Marek seděli vedle sebe a Karel seděl u okna.

Úlohy:

1. Kolika způsoby lze postavit 20 výrobků do řady?
2. Kolika způsoby lze postavit do řady vedle sebe do skříně 15 různých rezistorů?
3. Kolika způsoby lze postavit do řady na polici 10 různých žárovek a 5 LED diod tak, že nejprve budou žárovky a vedle nich diody.
4. Kolik různých devíticiferných čísel s různými ciframi lze sestavit z cifer 1 až 9?

Řešení:

1. $2,432 \cdot 10^{15}$
2. $15!$
3. 435456000
4. 362880

Definice 1.8 (Kombinace) K-členná kombinace z n prvků je neuspořádaná k-tice sestavená z těchto prvků tak, že se v ní každý prvek vyskytuje nejvýše jednou.

Definice 1.9 (Kombinace - počet) Počet $C_k\left(n\right)$ všech k-členných kombinací z n prvků je:

$$C_k\left(n\right) = \binom{n}{k} = \frac{n!}{(n-k)! \cdot k!} \tag{1.7}$$

Nejlépe si definici objasníme na příkladu.

■ **Řešený příklad 1.5** Do paralelního elektrického obvodu chceme zařadit čtyři žárovky. Na víběr máme 7 různých žárovek, dvě z nich nesvítí. Určete:

a. počet všech možností jak čtyři žárovky zapojit,

b. počet možností jak žárovky zapojit, jestliže chceme aby právě jedna žárovka nesvítila,

c. počet možností jak žárovky zapojit, jestliže chceme aby nejvýše jedna žárovka nesvítila,

d. počet možností jak žárovky zapojit, jestliže chceme aby alespoň jedna žárovka nesvítila.

Řešení:
Máme obsadit čtyři místa, to znamená, vybíráme čtveřice. Na výběr máme sedm žárovek, to znamená, máme sedm prvků. Jedná se o skupinu bez opakování v níž nezáleží na pořadí. Určujeme kombinace bez opakování.

a. Chceme zjistit počet všech možných zapojení. Hledáme čtveřice ze sedmi prvků, jejich počet je:

$$C_4\left(7\right) = \binom{7}{4} = \frac{7!}{(7-4)! \cdot 4!} = \frac{7!}{3! \cdot 4!} = 35.$$

Počet všech možných zapojení žárovek je 35.

b. Chceme zjistit počet všech zapojení v nichž jedna žárovka nesvítí. Nejprve vybereme žárovku co nesvítí. Dvě jsou rozbité a vybíráme právě jednu:

$$C_1\left(2\right) = \frac{2!}{1! \cdot 1!} = 2$$

Vybereme zbývájící tři žárovky pouze z nepoškozených.

$$C_3\,(5) = \frac{5!}{2! \cdot 3!} = 10$$

Celkem je tedy počet možností:

$$C_1\,(2) \cdot C_3\,(5) = 2 \cdot 10 = 20$$

c. Chceme zjistit počet všech zapojení v nichž svítí všechny žárovky nebo právě jedna nesvítí. Určíme počet zapojení v nichž svítí všechny žárovky a využijeme předchozí část při určování počtu možností zapojení v nichž právě jedna žárovka nesvítí. Pokud mají svítit všechny žárovky, znamená to, že vybíráme z pěti nepoškozených žárovek čtveřici.

$$C_4\,(5) = \frac{5!}{1! \cdot 4!} = 5$$

Celkem je pak možností:

$$C_4\,(5) + C_1\,(2) \cdot C_3\,(5) = 5 + 20 = 25$$

d. Chceme zjistit počet všech zapojení v nichž nesvítí právě jedna nebo dvě žárovky. Počet zapojení v nichž nesvítí jedna žárovka jsme už v předchozím určili. Musíme už jen určit počet zapojení v nichž nesvítí dvě žárovky.

$$C_2\,(2) \cdot C_2\,(5) = \frac{2!}{0! \cdot 2!} \cdot \frac{5!}{3! \cdot 2!} = 1 \cdot 10 = 10$$

Celkem je pak možností:

$$C_1\,(2) \cdot C_3\,(5) + C_2\,(2) \cdot C_2\,(5) = 20 + 10 = 30$$

Úlohy:

1. Ve třídě je 30 studentů. Kolika způsoby lze vybrat čtveřici na zkoušení?
2. Na běžecké trati závodí 8 sportovců. Do dalšího kola postupují první tři závodníci. Kolika způsoby můžeme vybrat postupující trojici?
3. Kolika způsoby lze 4 kondenzátory a 8 rezistorů zapojit do dvou složených obvodů střídavého proudu tak, aby v každém obvodu byly dva kondenzátory a 4 rezistory?
4. Ve skupině je 20 studentů, každý student má jiné jméno. Je mezi nimi je Jan a Jana. Kolika způsoby lze vybrat 8 studentů tak, aby mezi vybranými
   a. byl Jan,
   b. nebyl Jan,
   c. byl Jan a Jana,
   d. byl alespoň jeden z dvojice Jan a Jana,
   e. byl nejvýše jeden z dvojice Jan a Jana,
   f. nebyl Jan ani Jana.
5. V krabici je 10 multimetrů, z nichž jsou právě tři vadné. Kolika způsoby lze vybrat 5 multimetrů tak, aby
   a. žádný nebyl vadný,
   b. právě jeden byl vadný,
   c. nejvýše jeden byl vadný,
   d. právě dva byly vadné,
   e. nejvýše dva byly vadné,
   f. alespoň dva byly vadné.
6. Z kolika prvků lze vytvořit 990 kombinací durhé třídy bez opakování?
7. Zvětší-li se počet prvků o 4, zvětší se počet kombinací druhé třídy bez opakování vytvořených z těchto prvků o 30. Určete původní počet prvků.

Řešení:

1. 27405
2. 56
3. 420
4. a. 50388, b. 75582, c. 18564, d. 82212, e. 107406, f. 43758
5. a. 21, b. 105, c. 126, d. 105, e. 231, f. 126
6. $n = 45$
7. $n = 6$

1.5 Skupiny s opakováním

Definice 1.10 (Variace s opakováním) K-členná variace z n prvků, je uspořádaná k-tice sestavená z těchto prvků tak, že se v ní každý prvek vyskytuje nejvýše k-krát. Na pořadí prvků záleží.

Definice 1.11 (Počet variací s opakováním) Počet všech variací k-té třídy s opakováním z n různých prvků je dán vztahem.

$$V_k^*(n) = n^k \tag{1.8}$$

Poznámka: Všimněte si, že již není dána podmínka $n \geq k$, právě proto, že se jednotlivé prvky se mohou opakovat.

Definice 1.12 (Permutace s opakováním) Permutace s opakováním je každá taková uspořádaná n-tice z n prvků, v níž se každý prvek vyskytuje alespoň jednou. Na pořadí prvků záleží.

Definice 1.13 (Počet permutací s opakováním) Počet všech permutací s opakováním z n prvků je dán takto:

$$P^*_{n_1,n_2,\ldots,n_k}(n) = \frac{(n_1 + n_2 + \ldots + n_k)!}{n_1! \cdot n_2! \cdot \ldots \cdot n_k!} = \frac{n!}{n_1! \cdot n_2! \cdot \ldots \cdot n_k!} \tag{1.9}$$

Kde čísla $n_1, n_2, \ldots, n_k$ označují počet opakování jednotlivých prvků.

Definice 1.14 (Kombinace s opakováním) je neuspořádaná k-tice z n prvků, kde se každý prvek vykytuje nejvýše k-krát a nezáleží na pořadí prvků.

Definice 1.15 (Počet kombinací s opakováním) Pro všechna přirozená n a k je počet kombinací s opakováním dán takto:

$$C_k^*(n) = \binom{n+k=1}{k} = \frac{(n+k-1)!}{(n-1)! \cdot k!} \tag{1.10}$$

■ **Řešený příklad 1.6** Kolik pěticiferných čísel lze sestavit z cifer 0, 1, 2, 3, 4, 5, 6, 7, jestliže se cifry mohou opakovat?

Řešení:
Počet všech pěticiferných čísel, které můžeme sestavit je dán takto:

$$V_5^*(8) = 8^5 = 32768$$

Takto vytvořená čísla obsahují i čísla, která začínají jednou i více nulami. Taková čísla ale nepovažujeme za pěticiferná. Musíme je proto z počtu všech čísel odstranit. Nebo zvolíme méně náročný postup při určení počtu všech pěticiferných čísel.
Na první pozici nechceme nulu, proto zvolíme první pozici tak, aby nulu nemohla nabýt. U ostatních pozic na cifře a jejím opakování nezáleží.

$$V_1(7) \cdot V_4^*(8) = \frac{7!}{6!} \cdot 8^4 = 28672$$

■ **Řešený příklad 1.7** Kolik osmiciferných čísel lze sestavit tak, aby se cifra 2 v zápisu opakovala třikrát, cifra 4 dvakrát a cifra 7 třikrát?
Řešení:
Máme určit osmiciferfné číslo, které má být sestaveno z konkrétního počtu opakujících se cifer jejichž počet je stejný jako stupeň skupiny, kterou máme určit. Jedná se o permutace s opakováním.

$$P_{2;3;3}^*(8) = \frac{(2+3+3)!}{2! \cdot 3! \cdot 3!} = \frac{40320}{2 \cdot 6 \cdot 6} = 560$$

■ **Řešený příklad 1.8** Kolika způsoby lze vybrat šest závaží, jestliže máme k dispozici pouze závaží o třech hmotnostech v dostatečném množství?
Řešení:
Na pořadní vybíraných závaží nezáleží a mohou se opakovat, proto se jedná o kombinaci s opakováním, kde vybíráme šestici ze tří druhů.

$$C_6^*(3) = \binom{3+6-1}{6} = \frac{8!}{2! \cdot 6!} = 28$$

Úlohy:

1. Kolik značek Morseovy abecedy lze sestavit z teček a čárek, vytváříme-li skupiny o jednom až čtyřech prvcích?
2. Ze sedmi prvků jsme vytvořili 2401 variací s opakováním stejné třídy. Kolik prvků obsahuje jedna variace?
3. Kolik různých hodů můžeme provést dvěma nebo třemi různobarevnými kostkami?
4. Kolik pěticiferných čísel lze sestavit z cifer 1, 2, 5, 7, 8, 9, jesliže se cifry mohou opakovat?
5. Kolika způsoby lze uspořádat deset knih v knihovně, kde 4 z nich jsou romány, 3 jsou poezie, 2 encyklopedie a jedna je učebnice.
6. Kolik různých permutací lze vytvořit použitím všech písmen slov: STATISTIKA, MATEMATIKA, FYZIKA.
7. Přístupový kód do trezoru je tvořen posloupností tří písmen a čtyř číslic. Kolik různých kódů je možné sestavit, jestliže k dispozici je 28 písmen a 10 číslic.
8. Kolik přirozených čísel lze sestavit z číslic:

   a. 2; 4; 6; 8; 9, sestavené číslo bude trojciferné,

   b. 9; 7; 5, sestavené číslo bude pěticiferné.

9. Kolika způsoby lze koupit v prodejně 5 sešitů, mají-li 3 druhy sešitů.
10. V laboratoři fyziky je 12 druhů diod. Kolika způsoby si můžeme vybrat šest diod, nebo čtyři různé diody?

Řešení:

1. 30
2. 4
3. 36; 216
4. 7776
5. 12600
6. 75600; 151200; 720
7. 219520000
8. 125; 243
9. 21
10. 12376; 495

Kapitola 2

Pravděpodobnost

Ještě než zavedeme samotnou definici pravděpodobnosti je potřeba zavést základní pojmy.

2.1 Jevová algebra

Nejjednodušším pojmem, který zavedeme je **Náhodný pokus** dále jen **pokus**. Tímto pojmem popisujeme děj nebo činnost, kterou provedeme v předem připravených podmínkách ať už tím myslíme prostor laboratoře nebo libovolný prostor mimo laboratoř. Pro relevantnost jakéhokoli pokusu je nutné, aby byl pokus znovu opakovatelný. Tedy vyžadujeme takové podmínky průběhu, které je možné znovu nastavit. Pokusy můžeme rozlišit podle aktivit pozorovatele takto:

- Ovlivňuje-li pozorovatel pokus tím, že sám připraví podmínky pokusu. Tím myslíme např. zapojení elektrického obvodu, přípravu mechanické, termodynamické, optické soustavy k měření daného fyzikálního (biologického, chemického) zákona nebo děje. Potom takový pokus nazýváme **experiment** = je řízený pozorovatelem.
- Snaží-li se pozorovatel co nejméně podmínky pokusu ovlivnit a pouze registrovat daný děj a zaznamenávat data, např. volný pád těles - listy stromů, kapky deště, atd. Potom takový pokus nazýváme pozorování = co největší eliminace zásahů pozorovatele.

Definice 2.1 (Jev) Jako jev označujeme daný výsledek pokusu. Tímto pojmem ovšem označujeme pouze výsledky u nichž má smysl uvažovat o tom zda nastal nebo nenastal a zároveň i jeho důsledky.

Dále budeme pracovat s pojmem **hromadný jev**, kterým rozumíme takové výsledky pokusů, které můžeme teoreticky nekonečně-krát opakovat nebo pozorovat. Popřípadě se jedná o výsledky měření, které získáme z předmětů stejného druhu.

Definice 2.2 (Jistý jev) Mějme pokus jehož podmínky nastavíme tak, že sledovaný jev vždy nastane. Potom takový jev nazveme jev jistý. Jistý jev značíme I.

Definice 2.3 (Nemožný jev) Mějme pokus jehož podmínky nastavíme tak, že sledovaný jev nikdy nenastane. Potom takový jev nazveme nemožný jev. Nemožný jev značíme $\emptyset$.

Definice 2.4 (Náhodný jev) Mějme pokus a jeho podmínky nastavíme tak, že jev může, ale nemusí nastat. Potom takový jev nazveme náhodný jev. Náhodný jev značíme velkými písmeny A, B, C.

Vzhledem k tomu, že většinou získáme více náhodných jevů, mluvíme potom o množině jevů Ω. Na takové množině poté můžeme zavést některé relace a operace.

Definice 2.5 (Dílčí jev) Mějme dva jevy A a B. Říkáme, že jev A je dílčí jev jevu B, nebo jev B obsahuje jev A. Značíme $A \subset B$. Tedy jev B nastane vždy, když nastane jev A.

Poznámka: Jev B nazýváme nadjev a jev A nazýváme dílčí jev.

Definice 2.6 (Ekvivalence jevů) Mějme dva jevy A a B. Jev A je ekvivalentní jevu B, jestliže jev A nastane právě tehdy a jen tehdy, když nastane jev B a platí $A \subset B$ a $B \subset A$. Značíme $A \Leftrightarrow B$ nebo $A = B$.

Definice 2.7 (Opačný jev) Mějme dva jevy A a $\bar{A}$. Řekneme, že jev $\bar{A}$ je opačný k jevu A, jestliže nastane jev $\bar{A}$ a zároveň nenastane jev A. Značíme $\bar{A} = \Omega - A$.

Definice 2.8 (Součet jevů) Mějve dva jevy A a B. Součtem jevů A a B nazveme takový jev, který nastane právě tehdy, když nastane alespoň jeden z jevů A nebo B. Značíme $A \cup B$ nebo $A + B$.

Definice 2.9 (Rozdíl jevů) Mějme dva jevy A a B. Rozdílem jevů A a B nazveme takový jev, který nastane právě tehdy, když nastane jev A a současně nenastane jev B. Značíme $A - B$.

Definice 2.10 (Součin jevů) Mějme dva jevy A a B. Součinem jevů A a B nazveme takový jev, který nastane právě tehdy, když nastanou jevy A a B současně. Značíme $A \cdot B$ nebo $A \cap B$.

Definice 2.11 (Disjunktní (neslučitelené) jevy) Mějme jevy A a B. Jevy A, B nazveme disjunktní, jestliže oba jevy nemohou nastat současně. Značíme $A \cdot B = 0$. Říkáme, že jevy A a B se vzájemně vylučují. Obecně pak řekneme, že jevy $A_1, A_2, \ldots, A_n$ jsou disjunktní (vzájemně neslučitelné), jesltiže pro každé dva jevy A_i, A_j platí $A_i \cdot A_j = 0$, kde $i \neq j$ a $i, j = 1, 2, \ldots, n$.

Definice 2.12 (Rozpad jevu) Mějme množinu jevů $A_1, A_2, \ldots, A_n$ a jev B. Řekneme, že jev B se rozpadá na dílčí jevy jevy $A_1, A_2, \ldots, A_n$, jestliže pro každé dva jevy A_i, A_j platí $A_i \cdot A_j = 0$, kde $i \neq j$ a $i, j = 1, 2, \ldots, n$.

Definice 2.13 (Elementární jev) Elementárním jevem nazveme takový jev, který se nerozpadá na dílčí jevy a značíme jej E.

Definice 2.14 (Úplná soustava disjunktních jevů) Úplná soustava disjunktních jevů je taková množina jevů, pro kterou platí, že aspoň jeden jev nastane. Platí
$A_1 + A_2 + \ldots + A_n = I$.

Pro nyní definované relace a operace s jevy uvedeme ještě základní vlastnosti:

Vlastnosti 2.1

Mějme libovolné jevy A, B, C.

1. *Relace část je reflexivní, antisymetrická a tranzitivní:*

   a. $A \subset A$,

   b. $A \subset B \Rightarrow B \supset A$,

   c. $A = B \wedge B = C \Rightarrow A \subset C$.

2. *Relace ekvivalence je reflexivní, symetrická a tranzitivní:*

   a. $A = A$,

   b. $A = B \Rightarrow B = A$,

c. $A = B \wedge B = C \Rightarrow A = C$.

3. *Operace součet je komutativní a asociativní:*

   a. $A + B = B + A$,

   b. $(A + B) + C = A + (B + C)$.

4. *Operace násobení je komutativní a asociativní:*

   a. $A \cdot B = B \cdot A$,

   b. $(A \cdot B) \cdot C = A \cdot (B \cdot C)$.

5. *Distributivní zákony:*

   a. $(A + B) \cdot C = A \cdot C + B \cdot C$,

   b. $A \cdot B + C = (A + C) \cdot (B + C)$.

6. *Idempotence*

   a. $A + A = A$,

   b. $A \cdot A = A$.

7. *Dvojí negace* $\overline{\overline{A}} = A$.

8. *De Morganovy zákony:*

   a. $\overline{A + B} = \bar{A} \cdot \bar{B}$,

   b. $\overline{A \cdot B} = \bar{A} + \bar{B}$.

9. *Absorpce:*

   a. $A + A \cdot B = A$,

   b. $A \cdot (A + B) = A$.

10. *Vlastnosti opačného, nemožného a jistého jevu:*

a. $A + \bar{A} = I$,

b. $A \cdot \bar{A} = 0$,

c. $A + 0 = A$,

d. $A \cdot 0 = 0$,

e. $A + I = I$,

f. $A \cdot I = A$.

Definice 2.15 (Borelovské jevové pole) Borelovské jevové pole β je množina jevů $A_1, A_2, \ldots, A_n, \ldots$, které splňují:

a) operace součet na množině β je uzavřená,

b) operace součin na množině β je uzavřená,

c) operace opačný jev na množině β je uzavřená,

d) navíc splňují předchozích pět prvních vlastností.

Dále si uvedeme základní vlastnosti borelovského jevového pole:

Vlastnosti 2.2

1. *Jev jistý patří do jevového pole* $I \in \beta$.
2. *Mějme jev* A*, který je prvkem jevového pole* β*. Platí, že jev* $\bar{A}$ *opačný k jevu* A *je také prvkem jevového pole* β*:*
$$A \in \beta \Rightarrow \bar{A} \in \beta.$$
Speciálně pro jistý jev:
$$I \in \beta \Rightarrow \emptyset \in \beta.$$
3. *Mějme množinu jevového pole* $A_1, A_2, \ldots, A_n, \ldots \in \beta$*. Potom musí platit, že:*
$$\begin{aligned} A_1 + A_2 + \ldots + A_n + \ldots &\in \beta \\ A_1 \cdot A_2 \cdot \ldots \cdot A_n \cdot \ldots &\in \beta. \end{aligned}$$

Pro lepší pochopení těchto vlastností si připomeneme co znamená pojem uzavřená operace vzhledem k množině

- pro sčítání tento pojem můžeme zjednodušeně vysvětlit takto: součet libovolných prvků množiny musí být prvkem dané množiny.
- pro násobení je vysvětlení analogické: součin libovolných prvků množiny musí být prvkem dané množiny.

2.2 Pravděpodobnost

Historický vývoj pravděpodobnosti můžeme vysledovat téměř u každé rozvíjející se společnosti, u které se můžeme setkat s hrami, v mnoha případech především s hazardní hrou. Mezi prvními významnými zástupci můžeme jmenovat především dvě jména Pierre de Fermat a Blaise Pascal, kteří se pravděpodobností jako matematickým problémem zabývali ve své korespondenci týkající se právě hazardních her a jiných kombinatorických problémů. O další rozvoj a matematický popis pravděpodobnosti se mimo jiných významně zasloužili Christian Huygens, Abraham de Moivre a Jacob Bernoulli. Za jednoho z nejvýznamnějších zástupců je dnes považován Pierre-Simon Laplace, který ve svém díle shrnul dosavadní poznatky jeho předchůdců a dále je rozpracoval. Důsledkem vývoje této disciplíny je dnes možné výstavbu pravděpodobnosti provést několika způsoby. Postupný vývoj si ilustrujeme na různých definicích pravděpodobnosti.

Ještě než přistoupíme k samotné klasické definici pravděpodobnosti shňme si dosavadní poznatky. Zatím jsme zavedli elementární jev, o kterém víme, že je dále nerozložitelný. Množinu všech elementárních jevů nazýváme úplná množina elementárních jevů. Libovolný jev A je částí úplné množiny elementárních jevů a je rozložitelný na $E_1, E_2, \ldots, E_m$ elementráních jevů. Dále jsme zavedli nemožný a jistý jev, o kterých víme, že jsou součástí úplné množiny jevů. Na základě těchto poznatků můžeme zavést klasickou definici pravděpodobnosti.

Definice 2.16 (Klasická (Laplace) definice pravděpodobnosti) Mějme množinu N elementárních jevů $E_1, E_2, \ldots, E_n$. Tuto množinu označujeme jako úplnou množinu elemetárních jevů Ω, v níž jsou všechny jevy stejně možné. Mějme jev A, který je možné rozložit na m elementárních jevů $E_1, E_2, \ldots, E_m$, kde $m \leq n$ a je částí množiny Ω. Potom pravděpodobnost, že nastane právě jev A je reálné číslo:

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

V početní praxi většinou počet n všech elementárních jevů úplné množiny Ω označujeme jako počet všech možných výsledků daného pokusu. Tím myslíme jak výsledky, které vyhovují jevu A, tak i ty, které nevyhovují jevu A. Počet m všech možných elementárních jevů, na které se rozkládá jev A potom označujeme jako počet všech příznivých výsledků daného pokusu. Tím získáme známější klasickou definici pravděpodobnosti ve tvaru:

$$P(A) = \frac{\text{počet všech příznivých výsledků pokusu}}{\text{počet všech možných výsledků pokusu}}. \tag{2.1}$$

Z této definice přímo vyplývají vlastnosti pravděpodobnosti náhodného jevu A:

Vlastnosti 2.3

1. *Pravděpodobnost jevu A je vždy kladné číslo $P(A) \geq 0$.*
2. *Pravděpodobnost jistého jevu je $P(I) = 1$.*
3. *Jsou-li jevy A, B disjunktní (vzájemně neslučitelné), potom pro pravděpodobnost součtu dvou jevů platí:*
$$P(A+B) = P(A) + P(B). \tag{2.2}$$
Nejsou-li jevy A, B disjunktní (vzájemně neslučitelné), potom pro pravděpodobnost součtu dvou jevů platí:
$$P(A+B) = P(A) + P(B) - P(A \cdot B). \tag{2.3}$$
4. *Pro pravděpodobnost jevu opačného k jevu A platí:*
$$P\left(\bar{A}\right) = 1 - P(A). \tag{2.4}$$
5. *Pravděpodobnost nemožného jevu $\emptyset$ je:*
$$P(\emptyset) = 0. \tag{2.5}$$
6. *Pro hodnotu pravděpodobnosti libovolného jevu platí:*
$$0 \leq P(A) \leq 1. \tag{2.6}$$

Poznámka:

a. První a šestou vlastnost často využíváme k ověření správnosti výsledku daného výpočtu pravděpodobnosti.

b. U jevů, které nejsou disjunktní obsahuje pravděpodobnost jevu A i případy, kdy nastane jev B, stejně tak pro pravděpodobnost jevu B. Proto je důležité z celkové pravděpodobnosti součtu těchto jevů odstranit pravděpodobnost, kdy oba jevy nastávají současně, protože se nyní ve výpočtu vyskytuje dvakrát

c. Pro klasickou definici pravděpodobnosti je velice důležitý předpoklad, že všechny elementární jevy jsou stejně pravděpodobné. Neplatí-li tento předpoklad, nelze klasickou definici pravděpodobnosti použít. Nastíněný problém si ukážeme na klasickém příkladu.

■ **Řešený příklad 2.1** Máme tři klasické kostky. Hru házení kostkami vyhrajeme, jestliže padne součet vyšší než deset. Určete pravděpodobnost, že:

a) padne součet jedenáct,

b) padne součet dvanáct.

Řešení: Určíme příznivé výsledky daného počtu:

a) chceme, aby součet hodnot kostek byl jedenáct, vyhovují tedy trojice:

$$(6;4;1) \qquad (6;3;2) \qquad (5;5;1) \qquad (5;4;2) \qquad (5;3;3) \qquad (4;4;3)\,.$$

b) chceme, aby součet hodnot kostek byl dvanáct, vyhovují tedy trojice:

$$(6;5;1) \qquad (6;4;2) \qquad (6;3;3) \qquad (5;5;2) \qquad (5;4;2) \qquad (4;4;4)\,.$$

Zdá se, že počet příznivých výsledků je u obou součtů stejný a tedy pravděpodobnost obou jevů je:

$$P\,(11) = P\,(12)\,.$$

Problém nastane, uvědomíme-li si, že počet příznivých výsledků pro všechny trojice nemusí být stejný. Prozkoumejme tedy důsledněji počet všech příznivých výsledků pro každou vyhovující trojici hodnot. V tomto případě si tedy uvědomujeme, že na pořadí hodnot jednotlivých kostek záleží. Proto pravděpodobnosti obou hodů vypadají takto:

$$\begin{aligned}
P(11) &= \frac{P(3)+P(3)+P^*_{2;1}(3)+P(3)+P^*_{1;2}(3)+P^*_{2;1}(3)}{V^*_3(6)} = \\
&= \frac{3!+3!+\frac{(2+1)!}{2!\cdot 1!}+3!+\frac{(1+2)!}{1!\cdot 2!}+\frac{(2+1)!}{2!\cdot 1!}}{6^3} = \\
&= \frac{6+6+3+6+3+3}{216} = \frac{27}{216} \\
P(12) &= \frac{P(3)+P(3)+P^*_{1;2}(3)+P^*_{2;1}(3)+P(3)+P^*_3(3)}{V^*_3(6)} = \\
&= \frac{3!+3!+\frac{(1+2)!}{1!\cdot 2!}+\frac{(2+1)!}{2!\cdot 1!}+3!+\frac{3!}{3!}}{6^3} = \\
&= \frac{6+6+3+3+6+1}{216} = \frac{25}{216}
\end{aligned}$$

Dodržíme-li všechna pravidla pro počet pravděpodobnosti ukáže se, že pravděpodobnost součtu jedenáct je větší než pravděpodobnost součtu 12. Tímto příkladem jsme se přesvědčili jak je důležité důkladně kontrolovat předpoklad, že všechny výsledky jevu jsou stejně možné.

■ **Řešený příklad 2.2** Máme balíček 64 karet, kde jsou 4 žolíky. Určete s jakou pravděpodobností bude mezi osmi taženými kartami nejvýše jeden nebo alespoň 3 žolíky.
Řešení:

Nejprve určíme jaká je pravděpodobnost na nejvýše jeden žolík. Tento jev musíme rozdělit na dva nezávislé jevy. Jev A nevytáhneme ani jeden žolík, jev B vytáhneme právě jeden žolík. Také určíme celkový počet možných tahů karet.

$$\begin{aligned}
A &= C_8(60) = \binom{60}{8} \\
B &= C_1(4)\cdot C_7(60) = \binom{4}{1}\cdot\binom{60}{7} \\
N &= C_8(64) = \binom{64}{8}
\end{aligned}$$

Pravděpodobnost, že mezi taženými kartami bude nejvýše jeden žolík pak je:

$$P(K) = \frac{C_8(60)+C_1(4)\cdot C_7(60)}{C_8(64)} = 0,9271$$

Jaká je pravděpodobnost, že mezi taženými kartami budou alespoň tři žolíky. Znovu jev rozdělíme na dva, kdy jsou taženy právě tři žolíky a kdy jsou taženy právě čtyři žolíky.

$$\begin{aligned} A &= C_3(4) \cdot C_5(60) = \binom{3}{4} \cdot \binom{60}{5} \\ B &= C_4(4) \cdot C_4(60) = \binom{4}{4} \cdot \binom{60}{4} \end{aligned}$$

Pravděpodobnost, že mezi taženými kartami budou alespoň tři žolíky pak je:

$$P(L) = \frac{C_3(4) \cdot C_5(60) + C_4(4) \cdot C_4(60)}{C_8(64)} = 0,0050$$

Klasická definice pravděpodobnosti začne být příliš komplikovaná při velkém počtu výsledků pokusů a úplně selhává v případě, že množina všech možných výsledků je nespočetná. To byl důvod pro další zkoumání, které dospělo k dalším definicím pravděpodobnosti. Zavedeme tedy další definici, kterou nazýváme geometrická. Ta je založena na principu porovnávání míry geometrických útvarů. Mírou geometrického útvaru myslíme objem, povrch nebo délka geometrického objektu.

Definice 2.17 (Geometrická definice pravděpodobnosti) Mějme libovolný jev A, který je částí jevového pole τ. Jev A vymodelujeme nějakým geometrickým útvarem α jehož míra je $|\alpha|$. Jistý jev v daném jevovém poli vymodelujeme útvarem τ jehož míra je $|\tau|$. Dále předpokládáme, že $\alpha \subset \tau$ a, že každý výsledek je v modelu jistého jevu stejně možný. Potom platí, že pravděpodobnost jevu A je:

$$P(A) = \frac{|\alpha|}{|\tau|}. \tag{2.7}$$

Další matematický vývoj pravděpodobnosti dospěl ke statistické definici pravděpodobnosti, která se velice dobře hodí k popisu pravděpodobnosti jevu, který se často opakuje. Připomeňme, že tento opakující se jev nazýváme hromadný.

Definice 2.18 (Statistická definice pravděpodobnosti) Mějme hromadný jev A s počtem realizací n_A a počet všech realizací pokusu označíme n. Poměr $\frac{n_A}{n}$ nazýváme relativní četnost jevu A v n pokusech. Potom pro pravděpodobnost jevu A platí:

$$P(A) = \lim_{n \to \infty} \frac{n_A}{n}. \tag{2.8}$$

Postupný další vývoj a popis pravděpodobnosti dospěl do dnes používané moderní podoby, o kterou se významně zasloužil matematik A. N. Kolmogorov. Jeho definice popisuje pravděpodobnost jako objektivní vlastnost náhodného jevu, která nezávisí na tom, zda ji umíme nebo neumíme změřit.

Definice 2.19 (Axiomatická definice pravděpodobnosti) Mějme borelovské jevové pole β s jevy $A, A_1, A_2, \ldots, An, \ldots$. Pravděpodobnost jevu A je reálné číslo $P(A)$, které splňuje axiomy:

1. Pravděpodobnost pro všechny jevy $A \in \beta$

$$P(A) \geq 0 \tag{2.9}$$

2. Pravděpobodnost jistého jevu:

$$P(I) = 1 \tag{2.10}$$

3. Pravděpodobnost součtu skupiny navzájem neslučitelných jevů:

$$P(A_1 + A_2 + A_3 + \ldots + A_n + \ldots) = P(A_1) + P(A_2) + P(A_3) + \ldots + P(A_n) + \ldots \tag{2.11}$$

Z axiomatcké definice pravděpodobnosti, pak přímo vyplývají její vlastnosti:

Vlastnosti 2.4

1. *Pravděpodobnost výskytu nemožného jevu*

$$P(\emptyset) = 0 \tag{2.12}$$

2. *Pravděpodobnost výskytu opačného jevu*

$$P(\bar{A}) = 1 - P(A) \tag{2.13}$$

3. *Je-li jev A důsledkem jevu B, potom platí:*

   a.

$$0 \leq P(A) \leq P(B) \tag{2.14}$$

   b.

$$P(B - A) = P(B) - P(A) \tag{2.15}$$

4. *Pravděpodobnost součtu jevů, které nejsou vzájemně disjunktní (vzájmně se nevylučují)*

$$P(A + B) = P(A) + P(B) - P(A \cdot B) \tag{2.16}$$

5. *Hodnota pravděpodobnosti každého jevu A je vždy $P(A) \in \langle 0; 1 \rangle$*

2.2.1 Podmíněná pravděpodobnost

Jak nadpis napovídá, jedná se o případy kdy do daného systému podmínek, že nastane jev A, ještě přidáme další podmínku. Přidáváme podmínku, že před tím než nastane jev A musí nastat jev B. Potom takto nově zavedenou pravděpodobnost nazýváme podmíněná pravděpodobnost jevu A.

Definice 2.20 (Podmíněná pravděpodobnost) Mějme dva jevy A a B, jejichž pravděpodobnost výskytu je $P(A)$ a $P(B) \neq 0$. Pravděpodobnost, že nastane jev A za podmínky výskytu jevu B pak budeme značit $P(A/B)$ a definujeme čtvrtým axiomem:

$$P(A/B) = \frac{P(A \cdot B)}{P(B)} \tag{2.17}$$

V případě, že jev B je nemožný, podmíněná pravděpodobnost $P(A/B)$ není definována.

■ **Řešený příklad 2.3** Házíme třemi různě barevnými kostkami: bílá, červená a modrá. Určete jaká je pravděpodobnost, že na bílé kostce padne číslo 3, jestliže součet hodu všech tří kostek je 10.

Řešení:

Nejprve označme jednotlivé jevy:

a) Výsledek že na bílé kostce padne číslo 3 označíme jako jev A.

b) Součet hodu všech tří kostek je 10 označíme jako jev B.

Pro podmíněnou pravděpodobnost, že nastane jev A za podmínky B je potřeba určit:

1. K výpočtu podmíněné pravděpodobnosti musíme zjistit pravděpodobnost jevu B. K určení počtu výsledků, které označíme jako jev B použijeme tabulku (2.3) s jejich výpisem.

   Potom pravděpodobnost výskytu jevu B je dána takto:

   $$P(B) = \frac{27}{V_3^*(6)} = \frac{27}{216}$$

2. Pravděpodobnost $P(A \cdot B)$. Určíme tedy součin jevů $A \cdot B$. Využijeme všechny výsledky, které odpovídají jevu B a z nich vybereme pouze ty, které začínají číslicí 3 z tabulky (2.3). Potom pravděpodobnost součinu jevu A a B je:

   $$P(A \cdot B) = \frac{6}{V_3^*(6)} = \frac{6}{216}$$

1;3;6	2;2;6	3;1;6	4;1;5	5;1;4	6;1;3
1;4;5	2;3;5	3;2;5	4;2;4	5;2;3	6;2;2
1;5;4	2;4;4	3;3;4	4;3;3	5;3;2	6;3;1
1;6;3	2;5;3	3;4;3	4;4;2	5;4;1	
	2;6;2	3;5;2	4;5;1		
		3;6;1			

Tabulka 2.1: Výsledky jevu B

Nyní můžeme určit podmíněnou pravděpodobnost jevu A:

$$P(A/B) = \frac{P(A \cdot B)}{P(B)} = \frac{\frac{6}{216}}{\frac{27}{216}} = \frac{6}{27} = \frac{2}{9}$$

Pravděpodobnost, že na bílé kostce padne číslo tři, když je součet hodu 10 je přibližně 22,2%.

Dále pomocí podmíněné pravděpodobnosti můžeme ukázat, jak určit pravděpodobnost součinu dvou jevů. Samozřejmě bude záležet na tom jestli se jedná o jevy disjunktní nebo ne. Nejprve disjunktní jevy, pro které platí:

$$P(A \cdot B) = P(A) \cdot P(B). \tag{2.18}$$

Pro jevy, které nejsou disjunktní je určení pravděpodobnosti součin jevů komplikovanější. Můžeme využít již definovaného čtvrtého axiomu a potom pro pravděpodobnost součinu dvou závislých jevů platí:

$$P(A \cdot B) = P(A/B) \cdot P(B) = P(B/A) \cdot P(B). \tag{2.19}$$

S podmíněnou pravděpodobností ještě navíc úzce souvisí i nezávislot jevů. Pozor neplést si nezávislost a neslučitelnost jevů, jedná se o dva různé pojmy. Platí-li pro podmíněnou pravděpodobnost $P(A/B) = P(A)$. Což znamená, že výskyt jevu B neovlivňuje pravděpodobnost jevu A. Potom říkáme, že jev A nezávisí na jevu B. Zároveň s tímto tvrzením platí i obrácené tvrzení, že jev B je nezávislý na jevu A. Poslední tvrzení můžeme lehce ukázat pomocí pravděpodobnosti součinu dvou jevů:

$$\begin{aligned} P(A \cdot B) &= P(A/B) \cdot P(B) = P(A) \cdot P(B) \\ \text{Zároveň:} & \\ P(A \cdot B) &= P(B/A) \cdot P(A) = P(B) \cdot P(A) \end{aligned}$$

Proto nezávislot dvou jevů definujeme takto:

Definice 2.21 (Nezávislé jevy) Mějme dva jevy A a B a jejich pravděpodobnost $P(A)$ a $P(B)$. Tyto jevy jsou nezávislé, jestliže platí:

$$P(A/B) = P(A) \qquad \text{a} \qquad P(B/A) = P(B). \tag{2.20}$$

Také jevy A a B z nichž jeden má pravděpodobnost nulovou jsou nezávislé.

Nezávsilost jevů lze samozřejmě rozšířit i na libovolný počet n jevů:

Definice 2.22 (n nezávislých jevů) Mějme množinu jevů $A_1, A_2, A_3, \ldots, A_n$ a jejich pravděpodobnosti $P(A_1), P(A_2), P(A_3), \ldots, P(A_n)$. Jevy $A_1, A_2, A_3, \ldots, A_n$ jsou vzájemně nezávislé, jestliže platí:

$$P(A_1 \cdot A_2 \cdot A_3 \cdot \ldots \cdot A_n) = P(A_1) \cdot P(A_2) \cdot P(A_3) \cdot \ldots \cdot P(A_n) \tag{2.21}$$

Navíc je možné získat skupinu n jevů, pro které bude platit, že právě jen každé dva různé jevy ze skupiny $A_1, A_2, A_3, \ldots, A_n$ jsou nezávislé. V takovém případě tuto skupinu nazveme: skupina po dvou nezávislých jevů. Z předchozího také vyplívá, že je-li skupina jevů vzájemně nezávislá, jsou také všechny jevy po dvou nezávislé. Opačné tvrzení neplatí!

Vraťme se ještě na chvíli k poznámce, že je důležité nezaměňovat jevy nezávislé a neslučitelné. Ukážeme na dvou neslučitelných jevech. Pro pravděpodbnost součinu dvou neslučitelných jevů, jejichž pravděpodobnost je nenulová $P(A) \neq 0$ a $P(B) \neq 0$ platí:

$$P(A \cdot B) = 0.$$

Pokud by jevy A a B byly zároveň nezávislé, muselo by platit i tvrzení:

$$P(A \cdot B) = P(A) \cdot P(B) \neq 0.$$

Tyto rovnosti ale očividně najednou nastat nemohou:

$$0 = P(A \cdot B) = P(A) \cdot P(B) \neq 0.$$

Proto je důležité tyto dvě vlastnosti jevů nezaměňovat.

■ **Řešený příklad 2.4** Dva střelci střílejí střídavě na terč. Každý má tři rány. Vítězí ten střelec, který první zasáhne cíl. Určete pravděpodobnost, že zvítězí:

a. první střelec,

b. druhý střelec.

Pravděpodobnost zásahu u obou střelců je stená $p = 0,7$.
Řešení:

Nejprve určíme pravděpodobnost, že zvítězí první střelec, kterého označíme A. Druhý střelec má označení B. Střelec A vítězí pokud jako první zasáhne cíl nebo pokud na první ránu ani jeden nezasáhne a na druhou ránu střelec A zasáhne nebo pokud první dvě rány ani jeden ze střelců nezasáhne a na třetí ránu střelec A zasáhne cíl. To, jestli střelec A uspěje nebo ne nijak neovlivní úspěšnost střelce B. Jevy jsou nezávislé a proto můžeme napsat:

$$\begin{aligned}
C &= A + \overline{A} \cdot \overline{B} \cdot A + \overline{A} \cdot \overline{B} \cdot \overline{A} \cdot \overline{B} \cdot A \\
P(C) &= P(A) + P\left(\overline{A}\right) \cdot P\left(\overline{B}\right) \cdot P(A) + P\left(\overline{A}\right) \cdot P\left(\overline{B}\right) \cdot P\left(\overline{A}\right) \cdot P\left(\overline{B}\right) \cdot P(A) = \\
P(C) &= 0,7 + (1-0,7) \cdot (1-0,7) \cdot 0,7 + \\
&\quad + (1-0,7) \cdot (1-0,7) \cdot (1-0,7) \cdot (1-0,7) \cdot 0,7 = 0,7687
\end{aligned}$$

Pravděpodobnost, že zvítězí druhý střelec pak získáme analogicky:

$$\begin{aligned}
C &= \overline{A} \cdot B + \overline{A} \cdot \overline{B} \cdot \overline{A} \cdot B + \overline{A} \cdot \overline{B} \cdot \overline{A} \cdot \overline{B} \cdot \overline{A} \cdot B = \\
P(C) &= P\left(\overline{A}\right) \cdot P(B) + P\left(\overline{A}\right) \cdot P\left(\overline{B}\right) \cdot P\left(\overline{A}\right) \cdot P(B) + \\
&\quad + P\left(\overline{A}\right) \cdot P\left(\overline{B}\right) \cdot P\left(\overline{A}\right) \cdot P\left(\overline{B}\right) \cdot P\left(\overline{A}\right) \cdot P(B) = \\
P(C) &= 0,3 \cdot 0,7 + 0,3^3 \cdot 0,7 + 0,3^5 \cdot 0,7 = 0,2306
\end{aligned}$$

2.2.2 Úplná pravděpodobnost a Bayesův vzorec

Setkáváme se také s případy, kdy chceme určit pravděpodobnost výskytu jevu A. Ale určit počet všech příznivých případů je příliš komplikované. Je-li možné zjistit pravděpodobnost tohoto jevu za různých podmínek a pravděpodobnost výskytu těchto podmínek, pak pro nás bude mnohem jednodušší určit pravděpodobnost jevu A pomocí úplné pravděpodobnosti.

Věta 2.1 (Úplná pravděpodobnost) Mějme úplnou skupinu vzájemně neslučitelných jevů $B_1, B_2, B_3, \ldots, B_n$ a libovolný jev A, jehož pravděpodobnost chceme určit. Potom pro pravděpodobnost jevu A platí:

$$P(A) \sum_{i=1}^{n} \left[P(A/B_i) \cdot P(B_i) \right]. \tag{2.22}$$

Důkaz: Protože skupina vzájemně neslučitelných jevů $B_1, B_2, B_3, \ldots, B_n$ je úplná můžeme říci, že platí:

$$B_1 + B_2 + B_3 + \ldots + B_n = I.$$

A jev A můžeme rozepsat takto:

$$A = A \cdot I = A \cdot (B_1 + B_2 + B_3 + \ldots + B_n) = A \cdot B_1 + A \cdot B_2 + A \cdot B_3 + \ldots + A \cdot B_n.$$

Díky tomu, že jevy $B_1, B_2, B_3, \ldots, B_n$ jsou vzájemně neslučitelné, jsou vzájemně neslučitelné i jevy $A \cdot B_1, A \cdot B_2, A \cdot B_3, \ldots, A \cdot B_n$. Potom pro libovolnou dvojici jevů platí:

$$(A \cdot B_i) \cdot (A \cdot B_j) = A \cdot (B_i \cdot B_j) = A \cdot \emptyset = \emptyset$$

pro všechna $i \neq j$.
Podle třetího axiomu je pak pravděpodobnost jevu A:

$$\begin{aligned} P(A) &= P(A \cdot B_1 + A \cdot B_2 + A \cdot B_3 + \ldots + A \cdot B_n) = \\ &= P(A \cdot B_1) + P(A \cdot B_2) + P(A \cdot B_3) + \ldots + P(A \cdot B_n). \end{aligned}$$

Využijeme ještě čtvrtý axiom, který říká, že $P(A \cdot B_i) = P(A/B_i) \cdot P(B_i)$. Proto platí:

$$\begin{aligned} P(A) &= P(A \cdot B_1 + A \cdot B_2 + A \cdot B_3 + \ldots + A \cdot B_n) = \\ &= P(A \cdot B_1) + P(A \cdot B_2) + P(A \cdot B_3) + \ldots + P(A \cdot B_n) = \\ &= P(A/B_1) \cdot P(B_1) + P(A/B_2) \cdot P(B_2) + P(A/B_3) \cdot P(B_3) + \ldots + \\ &\quad + P(A/B_n) \cdot P(B_n) = \\ &= \sum_{i=1}^{n} P(A/B_i) \cdot P(B_i). \end{aligned}$$

$\square$

I z důkazu je zřejmé, že pravděpodobnosti podmínek je potřeba znát již před provedením pokusu a proto je nazýváme apriorní. Podmíněné pravděpodobnosti $P(A/B_i)$ určují jak se změní pravděpodobnosti $P(B_i)$, když nastane jev A. Nazýváme je aposteriorní pravděpodobnosti. Tyto pravděpodobnosti pak určujeme pomocí Bayesovy věty.

Věta 2.2 (Bayesova věta) Mějme dánu úplnou skupinu neslučitelných jevů $B_1, B_2, B_3, \ldots, B_n$, které nazýváme hypotézy. Pro libovolný jev A pak pravděpodobnost hypotézy B_k podmíněné jevem A platí:

$$P(B_k/A) = \frac{P(A/B_k) \cdot P(B_k)}{P(A)} = \frac{P(A/B_k) \cdot P(B_k)}{\sum\limits_{i=1}^{n} \left[P(A/B_i) \cdot P(B_i)\right]}, \tag{2.23}$$

kde číslo $k = 1; 2; 3; \ldots; n$.

Důkaz: Z věty o násobení pravděpodobnosti nezávislých jevů víme, že platí:

$$\begin{aligned} P(B_k \cdot A) &= P(B_k/A) \cdot P(A) \\ P(B_k \cdot A) &= P(A/B_k) \cdot P(B_k), \end{aligned}$$

Z první rovnice vyjádříme pravděpodobnost součinu a dosadíme do ní druhou rovnici:

$$\begin{aligned} P(B_k/A) &= \frac{P(B_k \cdot A)}{P(A)} = \frac{P(A/B_k) \cdot P(B_k)}{P(A)} = \\ &= \frac{P(A/B_k) \cdot P(B_k)}{\sum_{i=1}^{n} \left[P(A/B_i) \cdot P(B_i)\right]}. \end{aligned}$$

□

■ **Řešený příklad 2.5** V laboratoři fyziky jsou tři druhy multimetrů. Student si vybere multimetr prvního druhu s pravděpodobností 0,4, druhého druhu s pravděpodobností 0,5 a třetího druhu s pravděpodobností 0,1. Spolehlivost jednohlivých multimetrů je pak po řadě 0,95; 0,98 a 0,9. Určete pravděpodobnost, že:

a. vybraný multimetr je spolehlivý,

b. vybraný multimetr je druhého, třetího druhu.

Nejprve určíme pravděpodobnost, že je vybraný výrobek kvalitní. Použijeme větu o úplné pravděpodobnosti. Vypíšeme si údaje ze zadání:

$$P(A/B_1) = 0,4 \qquad P(B_1) = 0,95 \qquad P(A/B_2) = 0,5 \qquad P(B_2) = 0,98$$

$$P(A/B_3) = 0,1 \qquad P(B_3) = 0,9$$

Tyto pravděpodobnosti dosadíme do rovnice úplné pravděpodobnosti:

$$\begin{aligned} P\left(A\right) &= \sum_{i=1} 3P\left(A/B_i\right)\cdot P\left(B_i\right) \\ P\left(A\right) &= 0,4\cdot 0,95+0,5\cdot 0,98+0,1\cdot 0,9=0,96 \end{aligned}$$

Když jsme určili pravděpodobnost $P\left(A\right)=0,96$, že vybereme kvalitní multimetr, můžeme pomocí Bayesovy věty určit pravděpodobnost, že je druhého a třetího druhu.

$$\begin{aligned} P\left(B_2/A\right) &= \frac{P\left(A/B_2\right)\cdot P\left(B_2\right)}{P\left(A\right)}=\frac{0,5\cdot 0,98}{0,96}=0,51 \\ P\left(B_3/A\right) &= \frac{P\left(A/B_3\right)\cdot P\left(B_3\right)}{P\left(A\right)}=\frac{0,1\cdot 0,9}{0,96}=0,094 \end{aligned}$$

2.2.3 Bernoulliho posloupnost nezávislých pokusů

Již jsme se zmiňovali o náhodných pokusech, které se opakují vícekrát. Při takovém pokusu je možné, aby se určitý sledovaný jev A opakoval právě n-krát. Jsou-li navíc pokusy vzájemně nezávislé (předchozí pokus neovlivní ten následující, např. do slosovacího zařízení vracíme tažená čísla) mluvíme pak o Bernoulliho posloupnosti nezávislých jevů. V takovém případě opakujícího se náhodného pokusu nás většinou nezajímá, v kterých pokusech nastal jev A, ale spíše určujeme kolikrát se daný jev uskutečnil, tedy určujeme pravděpodobnost, že se právě jev A v daném počtu n náhodných pokusů uskuteční právě k-krát.

Věta 2.3 (Bernoulliho schéma - binomické rodělení) Mějme právě n nezávislých pokusů, u kterých sledujeme pravděpodobnost výskytu jevu A. Nebo také říkáme, že pokus končí zdarem s pravděpodobností p a nezdarem s pravděpodobností q, kde $q=1-p$. Potom pravděpodobnost, že n náhodných pokusů skončí jevem A právě k-krát je dána takto:

$$P\left(A_k\right)=\binom{n}{k}\cdot p^k\cdot\left(1-p\right)^{n-k}. \tag{2.24}$$

Důkaz: Jestliže jev A nastane v jednom pokusu s pravděpodobností p a nenastane s pravděpodobností q, potom v posloupnosti n nezávislých pokusů jev A nastane právě k-krát v těchto případech $\binom{n}{k}$ s pravděpodobností úspěchu $\underbrace{p\cdot p\cdot p\cdot\ldots\cdot p}_{k}=p^k$ a zároveň nenastane s pravděpodbností $\underbrace{(1-p)\cdot(1-p)\cdot(1-p)\cdot\ldots\cdot(1-p)}_{n-k} = (1-p)^{n-k}$, proto:

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

$\square$

Všechny možné výsledky opakujícího se náhodného pokusu $A_0, A_1, A_2, \ldots, A_n$ tvoří úplnou skupinu vzájemně se nevylučujících jevů protože:

$$\begin{aligned} P(A_0) + P(A_1) + P(A_2) + \ldots + P(A_n) &= \binom{n}{0}\cdot p^0\cdot(1-p)^n + \binom{n}{1}\cdot p^1\cdot(1-p)^{n-1} + \\ &\quad + \binom{n}{2}\cdot p^2\cdot(1-p)^{n-2} + \ldots + \\ &\quad + \binom{n}{n}\cdot p^n\cdot(1-p)^0 = \\ &= (p + [1-p])^n = 1^n = 1 \end{aligned}$$

Proto se také Bernoulliho schéma nazývá binomické rozdělení.

Chceme-li místo pravděpodobnosti, že v posloupnosti n pokusů nastane jev A právě k-krát určit pravděpodobnost, že jev A nastane nejvýše k-krát, zjišťujeme vlastně všechny pravděpodobnosti, že jev A nastane právě $0, 1, 2, \ldots, k$-krát, tedy:

$$\begin{aligned} P(B_k) &= P(A_0) + P(A_1) + P(A_2) + \ldots + P(A_k) = \\ &= \sum_{i=0}^{k}\left[\binom{n}{i}\cdot p^i\cdot(1-p)^{n-i}\right]. \end{aligned}$$

Chceme-li určit pravděpodobnost, že jev A nastane nejméně m-krát, myslíme tím, že nastane právě m a více-krát:

$$\begin{aligned} P(B_m) &= P(A_m) + P(A_{m+1}) + P(A_{m+2}) + \ldots + P(A_n) = \\ &= \sum_{i=m}^{n}\left[\binom{n}{i}\cdot p^i\cdot(1-p)^{n-i}\right]. \end{aligned}$$

Chceme-li určit pravděpodobnost, že jev A nastane v n pokusech aspoň jednou je vlastně speciální případ předchozího příkladu, kdy $m = 1$:

$$\begin{aligned} P(B_1) &= P(A_1) + P(A_2) + P(A_3) + \ldots + P(A_n) = 1 - P(A_0) \\ &= 1 - \binom{n}{0}\cdot p^0\cdot(1-p)^{n-0} = 1 - (1-p)^n \end{aligned}$$

Poznámka: $P(A_0)$ je pravděpodobnost, že v n nezávislých pokusech nenastane jev A ani jednou.

Užitečné je také určit nejpravděpodobnější počet výskytu jevu A právě v n pokusech. Tento počet je celé číslo K, pro které platí:

$$np - (1 - p) \leq K \leq np + p.$$

■ **Řešený příklad 2.6** Střelec má 67% úspěšnost střelby na cíl. Střelec vystřelí na cíl desetkrát. Určete pravděpodobnost, že

a. zasáhne cíl právě devětkrát,

b. zasáhne cíl alespoň devětkrát,

c. zasáhne cíl nejvýše dvakrát.

d. Jaký je nepravděpodobnější počet zásahů?

Řešení:

a. Pravděpodobnost, že střelec zasáhne cíl právě děvětkrát spočítáme takto:

$$\begin{aligned} P(A) &= \binom{10}{9} \cdot 0,67^9 \cdot (1 - 0,67)^{10-9} = \frac{10!}{1! \cdot 9!} \cdot 0,67^9 \cdot 0,33^1 = \\ &= 0,0898 \end{aligned}$$

b. Pravděpodobnost, že střelec zasáhne cíl alespoň devětkrát znamená, že střelec zasáhne právě devětkrát nebo desetkrát.

$$P(B) = \binom{10}{9} \cdot 0,67^9 \cdot 0,33^1 + \binom{10}{10} \cdot 0,67^{10} \cdot 0,33^0 = 0,1080$$

c. Pravděpodobnost, že střelec zasáhne cíl nejvýše dvakrát určíme podobně jako předchozí.

$$\begin{aligned} P(C) &= \binom{10}{0} \cdot 0,67^0 \cdot 0,33^{10} + \binom{10}{1} \cdot 0,67^1 \cdot 0,33^9 + \binom{10}{2} \cdot 0,67^2 \cdot 0,33^8 = \\ &= 0,0032 \end{aligned}$$

d. Nejpravděpodobnější počet zásahů určíme takto:

$$\begin{aligned} np-(1-p) \leq \quad & K \quad \leq np+p \\ 10\cdot 0,67-(1-0,67) \leq \quad & K \quad \leq 10\cdot 0,67+0,67 \\ 6,37 \leq \quad & K \quad \leq 7,37 \qquad \Rightarrow K=7 \end{aligned}$$

Úlohy:

1. Jaká je pravděpodobnost výhry 1. ceny ve sportce při jednom vsazení? K výhře je potřeba uhodnout 6 čísel ze 46.

2. Kruhový terč má 3 pásma. Pravděpodobnost zásahu 1. pásma je 0,2, druhého 0,23 a třetího 0,15. Jaká je pravděpodobnost minutí cíle?

3. Do kolony bylo náhodně seřazeno 7 automobilů. 2 Škody, 2 Peugeoty a 3 BMW. Jaká je pravděpodobnost, že na prvním a posledním místě bude stát BMW?

4. V krabici je 6 bílých kuliček a 4 černé kuličky. Náhodně vylosujeme 2 kuličky. S jakou pravděpodobností:

   a. nebude vybrána ani jedna bílá kulička?

   b. bude vybrána jedna bílá a jedna černá kulička?

   c. obě kuličky budou bílé?

5. Ze 32 hracích karet vybíráme dvakrát za sebou jednu kartu. Určete pravděpodobnost, že:

   a. obě karty jsou esa, jestliže jsme první kartu nevrátili,

   b. obě karty jsou stejné barvy, jestliže jsme první vytaženou kartu opět vrátili zpět.

6. V telefoním seznamu náhodně vybereme jedno šestimístné číslo (může začínat nulou) a předpokládáme, že v seznamu jsou použita všechna šestimístná čísla. Jaká je pravděpodobnost, že telefoní číslo:

   a. neobsahuje 0,

   b. obsahuje jednu 3.

7. V dodávce 100 kusů křišťálových váz je 5 vadných. Při kontrole vybereme náhodně 4 kusy. S jakou pravděpodobností:

   a. je jedna vybraná váza vadná,

   b. alespoň jedna z vybraných váz vadná?

8. V urně je 5 bílých a 7 černých kuliček. Vytáhneme za sebou dvě kuličky. Jaká je pravděpodobnost vytažení dvou bílých kuliček, jestliže se po prvním tahu kulička:

   a. nevrátí,

   b. vrátí.

9. Z celkové produkce závodu jsou 4% zmetků a z dobrých je 75% standardních. Určete pravděpodobnost, že náhodně vybraný výrobek je standardní.

10. Tři závody vyrábí žárovky. První 45% celkové produkce, druhý 40% a třetí 15%. Z produkce prvního závodu je nepoškozených 70%, druhého 80% a třetího 81%. Určete pravděpodobnost, že si zákazník koupí standardní žárovku.

11. Součástky, ze kterých se montují stroje, dodávají tři závody. Je známo, že první má 0,3% zmetků, druhý 0,2% zmetků a třetí 0,4%. Přitom první závod dodal 1000, druhý 2000 a třetí 2500 součástek. Jaká je pravděpodobnost, že náhodně vybraná součástka bude zmetek?

12. Výrobek je postupně obráběn na dvou strojích. Pravděpodobnost kvalitního zpracování výrobku na prvním stroji je 0,8 a na druhém stroji 0,9. Stroje pracují nezávisle na sobě. Jaká je pravděpodobnost zhotovení kvalitního výrobku?

13. Jevy A, B a C jsou vzájemně nezávislé a všechny mají pravděpodobnost 0,8. S jakou pravděpodobností při jednom pokusu:

   a. nastanou všechny tři jevy současně,

   b. nenastane ani jeden z jevů,

   c. nastane pouze jev A,

   d. nastane právě jeden z těchto jevů.

14. Tři sportovci hází nezávisle jeden na druhém oštěpem. První překoná hranici 80 m průměrně v 80%, druhý v 70% a třetí v 50% hodů. Každý z nich jednou hodí. S jakou pravděpodobností bude překonaná hranice 80 m?

15. Střelec třikrát nezávisle vystřelil na cíl. Pravděpodobnost zásahů je prostupně 0,5; 0,6 a 0,8. S jakou pravděpodobností bude v cíli:

    a. právě jeden zásah,

    b. alespoň jeden zásah?

16. Pravděpodobnost, že dodávka bude mít více než 2% vadných výrobků, je 0,08. S jakou pravděpodobností bude ve třech dodávkách z dvaceti více než 2% vadných výrobků?

17. Dva střelci střílejí nezávisle na cíl. Pravděpodobnost zásahu prvního je 0,7 a druhého 0,8. S jakou pravděpodobností při současném výstřelu zasáhne cíl alespoň jeden z nich?

18. Dva hráči házejí postupně mincí. Vyhrává ten, komu padne jako první líc. S jakou pravděpodobností vyhraje první z hráčů?

19. Ve společnosti je 45% mužů a 55% žen. Vysokých nad 190 cm je 5% mužů a 1% žen. Náhodně vybraná osoba je vyšší než 190 cm. Jaká je pravděpodobnost, že je to žena?

20. S jakou pravděpodobností při 5 hodech kostkou padne alespoň jednou šestka?

21. Je známo, že určitý lék úspěšně léčí dané onemocnění v 90% případů. S jakou pravděpodobností alespoň čtyři z pěti pacientů budou tímto lékem vyléčeni?

22. Určete pravděpodobnost, že při pěti hodech kostou padne:

    a. šestka právě dvakrát,

    b. šestka při druhém a čtvrtém hodu.

23. Pravděpodobnost výhry hráče je 0,6. Určete, jaký je nejpravděpodobnější počet výher hráče v deseti odehraných partiích.

24. Na skladě je 70% přístrojů první jakosti a 30% druhé jakosti. Pravděpodobnost, že přístroj 1. jakosti pracuje bez poruchy je 0,95 a přístroj 2. jakosti 0,7. Organizace koupila jeden přístroj a ten pracoval bez poruchy. S jakou pravděpodobností byl přístroj 1. jakosti?

25. Při vyšetřování pacienta je podezření na tři navzájem se vylučující onemocnění. Pravděpodobnost výskytu první choroby je 0,3, druhé 0,5 a třetí 0,2. Laboratorní zkouška je pozitivní u 15% nemocných s první nemocí, 30% nemocných s druhou a 30% nemocných s třetí nemocí. Jaká je pravděpodobnost druhé nemoci, je-li po laboratotním vyšetření výsledek pozitivní?

26. V dílně pracuje 10 dělníků, kteří za směnu vyrobí stejný počet výrobků. Pět z nich vyrobí 96% dobrých výrobků, tři 90% a dva 85%. Náhodně vybereme jeden výrobek a ten je dobrý. S jakou pravděpodobností jej vyrobila první skupina dělníků?

Řešení:

1. $1,06 \cdot 10^{-7}$
2. $0,42$
3. $0,142$
4. a. $0,13$, b. $0,53$, c. $0,33$
5. a. $0,012$, b. $0,25$
6. a. $0,531$, b. $0,354$
7. a. $0,176$, b. $0,188$
8. a. $0,15$, b. $0,17$
9. $0,72$
10. $0,7565$
11. $0,003$
12. $0,72$
13. a. $0,512$, b. $0,008$, c. $0,032$, d. $0,096$
14. $0,97$
15. a. $0,26$, b. $0,96$
16. $0,14$
17. $0,94$
18. $0,67$
19. $0,196$
20. $0,598$
21. $0,918$
22. a. $0,16$, b. $0,016$

23. 6
24. $0,76$
25. $0,588$
26. $0,52$

Kapitola 3

Náhodná veličina

3.1 Náhodná veličina

Výsledky náhodných pokusů označujeme jako jevy. Některé jevy mají kvantitativní charakter. To znamená můžeme je označit čísly. Nejjednodušším příkladem je hod kostkou - padne číslo 1, 2,. . . , 6; losování čísel z osudí, atd. Ostatní jevy mají kvalitativní povahu. Za kvalitativní jev považujeme takový jev, který popisuje vlastnost nebo stav výsledku, např. fungující - rozbitá součástka, barva na kostce, atd. V takových případech se snažíme každému výsledku přiřadit nějaké číslo = hodnotu. Mějme třeba situaci, kdy určujeme, že se narodí syn nebo dcera. Výsledky těchto jevů můžeme rozlišit přiřazením reálného čísla: syn = 0, dcera = 1. Tato přiřazená reálná čísla potom považujeme za hodnotu náhodné veličiny. Na základě předchozích úvah pak náhodnou veličinu můžeme definovat následujícím způsobem.

Definice 3.1 (Náhodná veličina X) Náhodnou veličinou chápeme funkci, která každému jevu z množiny všech elemetárních jevů (úplné množiny jevů) přiřazuje reálné číslo.

Definice 3.2 (Obor hodnot náhodné veličiny X) Množinu všech reálných čísel, kterých může náhodná velčina nabýt nazýváme obor hodnot H náhodné veličiny X.

Poznámka:

- Náhodné veličiny obvykle značíme velkými písmeny např. X, Y a hodnoty těchto veličin značíme malými písmeny x, y.

- Náhodnou veličinu nelze určit bez provedení náhodného pokusu.

■ **Řešený příklad 3.1** Náhodnou veličinu X definujeme přiřazením přirozených čísel na různobarevné kostce. Určete všechny hodnoty, které může náhodná veličina X získat.
Řešení:
Každá kostka tvaru krychle má šest stěn. To znamená máme šest barev. Pravděpodobnost, že na kostce padne právě jedna strana je stále stejná $\frac{1}{6}$. Protože všechny stěny mají stejnou pravděpodobnost výskytu je obor hodnot náhodné veličiny, která je dána barvou stěny kostky je $H = \{\frac{1}{6}\}$.

■ **Řešený příklad 3.2** Náhodná veličina X je definována jako vzdálenost, kterou urazí částice než narazí do jiné částice nebo překážky. Určete obor hodnot náhodné veličiny X.
Řešení:
Vzdálenost, kterou sledovaná částice může urazit je teoreticky nekonečně dlouhá a zároveň může okamžitě při začátku sledování narazit do jiné částice nebo na překážku. Proto můžeme říci, že obor hodnot H náhodné veličiny X je interval $H = \langle 0; \infty)$.

Z předchozích příkladů si můžeme všimnout, že obory hodnod se mohou různit dvěma způsoby. Buď získáme spočetnou množinu, nebo nespočetnou mnoužinu reálných čísel. Podle toho také náhodné veličiny rozdělujeme na:

Diskérní náhodná veličina může nabývat nejvýše spočetně mnoha reálných čísel a jejím oborem hodnot je posloupnost reálných čísel.

Spojitá náhodná veličina nabývá nespočetně mnoha reálných hodnot a jejím oborem hodnot je buď uzavřený nebo otevřený interval reálných čísel.

3.2 Frekvenční a distribuční funkce

3.2.1 Distribuční funkce

Dále k popisu náhodné veličiny kromě jejího oboru hodnot používáme pravděpodobnost výskytu jednotlivých hodnot náhodné veličiny. Z různých pokusů např. házením kostkou zjistíme, že výskyt hodnot náhodné veličiny je nahodilý i přesto je možné nalézt jisté zákonitosti v těchto výskytech. Tyto zákonistosti si ukážeme na distribuční a frekvenční funkci náhodné veličiny. Jako první si ukážeme distribuční funkci, která ukazuje jakým způsobem je rozdělena pravděpodobnost hodnoty náhodné veličiny.

Definice 3.3 (Distribuční funkce náhodné veličiny X) Distribuční funkce nebo také rozdělení pravděpodobnosti náhodné veličiny je funce, která je definována na všech reálných číslech a vyjadřuje pravděpodobnost, s kterou náhodná veličina nabude hodnoty menší než a, platí:

$$F(x) = P(X < a) \tag{3.1}$$

Poznámka: Distribuční funkce definujeme stejným způsobem jak pro diskrétní, tak i pro spojitou náhodnou veličinu X.

Z definice pak přímo vyplívají následující vlastnosti.

Vlastnosti 3.1

1. *Distribuční funkce je v oboru všech reálných čísel neklesající a platí:*

$$(\forall x_1; x_2 \in \mathbb{R})\,(x_1 < x_2 \Rightarrow F(x_1) \leq F(x_2)) \tag{3.2}$$

2. *V nevlastních bodech $-\infty$ a ∞ platí:*

$$F(-\infty) = \lim_{x\to-\infty} F(x) = 0 \tag{3.3}$$
$$F(\infty) = \lim_{x\to\infty} F(x) = 1 \tag{3.4}$$

3. *Z vlastnosti 2. přímo vyplývá, že pro všechna $x \in \mathbb{R}$ platí:*

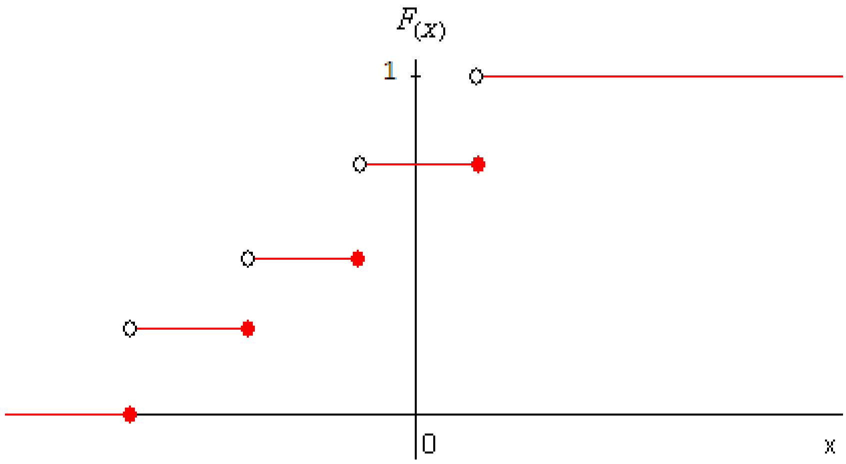
$$0 \leq F(x) \leq 1 \tag{3.5}$$

4. *Pro pravděpodobnost výskytu hodnot z intervalu $\langle a; b)$, kde $a; b \in \mathbb{R}$, platí:*

$$P(a \leq X < b) = F(b) - F(a) \tag{3.6}$$

5. *Pro pravděpodobnost hodnot z intervalu $\langle a; \infty)$, kde $a \in \mathbb{R}$, platí:*

$$P(X \geq a) = 1 - F(a) \tag{3.7}$$



Obrázek 3.1: Distribuční funkce diskrétní náhodné veličiny

6. *Pro pravděpodobnost, že náhodná veličina X má hodnotu právě $a \in \mathbb{R}$, platí:*

$$P(X = a) = \lim_{x \to a} F(x) = F(a) \tag{3.8}$$

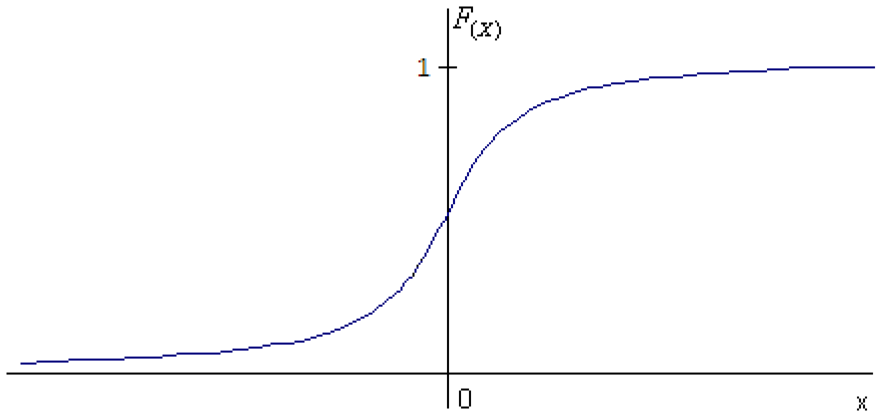
Distribuční funkce spojité a diskrétní náhodné veličiny X se liší pouze ve spojitosti a nespojitosti distributivní funkce. Distribuční funkce je u diskrétní náhodné veličiny konstantí a její hodnoty se mění „skokem“ na další hodnotu. Zároveň je v místě „skoku“ zprava spojitá. Distribuční funkci spojité náhodné veličiny pak přiřadíme vlastnost spojitá funkce. Nejlépe je tento rozdíl patrný na obrázcích (3.1) a (3.2):

3.2.2 Frekvenční funkce

Frekvenční funkce obecně vyjadřují rozložení pravděpodobností náhodných veličin. Frekvenční funkci definujeme pro diskrétní a spojitou náhodnou veličinu X vzlášť. K odlišení také slouží jiné názvy těchto funkcí. Pro diskrétní náhodnou veličinu ji nazveme Pravděpodobnostní funkce náhodné veličiny a pro spojitou náhodnou veličinu ji nazveme Hustota pravděpodobnosti náhodné veličiny.

Pravděpodobností funkce diskrétní náhodné veličiny

Definice 3.4 (Pravděpodobnostní funkce) Reálnou funkci $p(x) = P(X = x)$ nazveme pravděpodobnostní funkce diskrétní náhodné veličiny X, kde výraz $P(X = x)$ přiřazuje



Obrázek 3.2: Distribuční funkce spojité náhodné veličiny

každému číslu x z oboru hodnot H náhodné veličny pravděpodobnost, že náhodná veličina nabývá hodnoty x.

Z definice pak přímo vyplívají následující vlastnosti.

Vlastnosti 3.2

1. *Pro všechny hodnoty z oboru hodnot, platí:*

$$p\left(x\right) \geq 0 \tag{3.9}$$

2. *Pro všechny náhodné veličiny X, platí:*

$$\sum_{i} p_i\left(x\right) = 1 \tag{3.10}$$

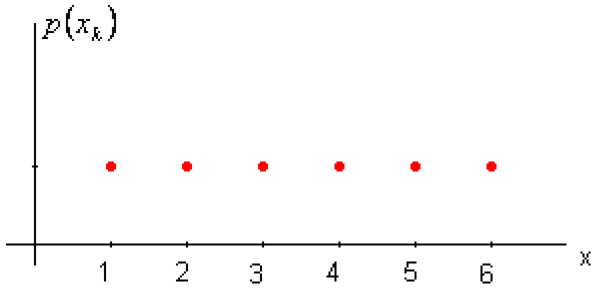
3. *Pro distribuční funkci diskrétní náhodné veličiny, platí:*

$$F\left(x\right) = \sum_{x_i < x} p\left(x_i\right) \tag{3.11}$$

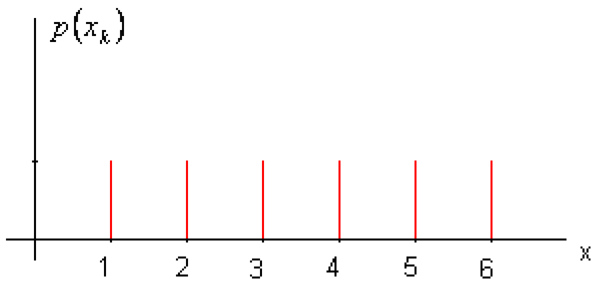
Poznámka:

- Třetí vlastnost využíváme především k výpočtům distribuční funkce diskrétní náhodné veličiny X.

- Pravděpodobnostní funkci můžeme také chápat jako počet výskytu různých pravděpodobností náhodné veličiny X.
- Ke grafickému znázornění pravděpodobnostní funkce nejčastěji využíváme bodový graf (Obrázek (3.3)), úsečkový graf (Obrázek (3.4)) nebo histogram (Obrázek (3.5)). Dnešní výpočetní technika umožňuje využít i jiné druhy grafů.



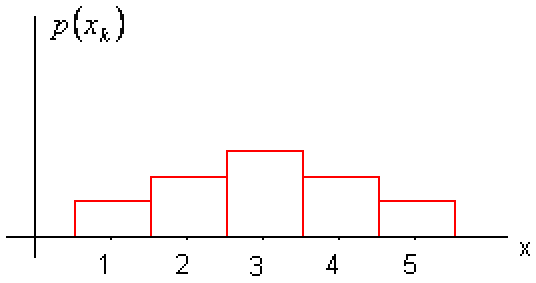
Obrázek 3.3: Bodobý graf pravděpodobností funkce



Obrázek 3.4: Úsečkový graf pravděpodobností funkce

Hustota pravděpodobnosti spojité náhodné veličiny X

Hustotu pravděopodobnosti spojité náhodné veličiny obvykle značíme $f(x)$. Jak již víme, obor hodnot spojité náhodné veličiny je interval reálných čísel. Zvolme náhodnou veličinu X tak, že její obor hodnot H je uzavřený interval $\langle a; b\rangle$. Otázkou je jaká



Obrázek 3.5: Histogram pravděpodobnostní funkce

je pravděpodobnost, že náhodná velčina X nabude hodnoty x, neboli $P(X = x) = ?$. Vzhledem k tomu, že oborem hodnot je uzavřený interval, těchto hodnot je nekonečně mnoho, proto když chceme určit pravděpodobnost $P(X = x)$ jedná se o případ, kdy zjišťujeme pravděpodobnost výskytu jednoho výsledku mezi nekonečně mnoha výsledky. Je takřka nemožné ho najít a proto:

$$P(X = x) = 0$$

Z této ukázky je vidět, že hustotu pravděpodobnosti spojité náhodné veličiny X budeme muset definovat jinak než u diskrétní náhodné veličiny. Díky spojitosti náhodné veličiny X v jejím oboru hodnot nám pomohou limity.

Definice 3.5 (Hustota pravděpodobnosti spojité náhodné veličiny X) Mějme nezápornou reálnou funkci jedné proměnné definovanou na intervalu $\langle a; b\rangle$, pro kterou platí:

$$f(x) = \lim_{h \to 0} \frac{P(x \le X < x + h)}{h}. \tag{3.12}$$

Takovou funkci nazveme hustota pravděpodobnosti spojité náhodné veličiny pro všechna $x; x+h \in \langle a; b\rangle$ (číslo h považujeme za přírustek hodnoty náhodné veličiny). Pro všechna ostatní $x \notin \langle a; b\rangle$ platí:

$$f(x) = 0 \tag{3.13}$$

Z definice pak přímo vyplívají následující vlastnosti.

Vlastnosti 3.3

1. *Hustota pravděpodobnosti náhodné veličiny vždy nabývá hodnot* $f(x) \geq 0$ *pro všechna* $x \in \mathbb{R}$.
2. *Náhodná veličina nemusí být definována pouze na uzavřeném intervalu* $\langle a; b\rangle$, *ale i na množině všech reálných čísel a proto platí:*

$$\int\limits_a^b f(x)\,\mathrm{d}x = \int\limits_{-\infty}^{\infty} f(x)\,\mathrm{d}x = 1 \tag{3.14}$$

3. *Pro distribuční funkci náhodné veličiny platí:*

$$F(x) = P(X < x) = \int_a^x f(x)\,\mathrm{d}x = \int_{-\infty}^x f(x)\,\mathrm{d}x \tag{3.15}$$

4. *Pro distribuční funkci náhodné veličiny platí:*

$$f(x) = F'(x) \tag{3.16}$$

5. *Pro pravděpodobnost, že hodnota náhodné velčiny je z intervalu* $(c; d)$ *platí:*

$$P(c < X < d) = F(d) - F(c) \tag{3.17}$$

3.3 Charakteristiky rozložení pravděpodobnosti

Dostaneme-li náhodnou veličinu u níž je příliš složité nebo pracné určit vlastnosti frekvenční a distribuční funkce, snažíme se její charakteristiky popsat jiným způsobem. Za vhodný způsob považujeme číselné charakteristiky, které nám pomohou si o rozložení náhodné veličiny udělat obrázek a výstižně charakterizovat její důležité vlastnosti a navíc nám takové hodnoty pomohou lépe porovnávat různé náhodné veličiny.

Tyto číselné charakteristiky rozložení pravděpodobnosti náhodných veličin samozřejmě pro lepší přehlednost dělíme a ty nejdůležitější si ukážeme. Podle popisované vlastnosti rozdělíme charakteristiky takto:

Charakteristiky polohy Jsou to hodnoty, které můžeme považovat za jistý druh středu, kolem kterého všechny ostatní hodnoty náhodné veličiny kolísají. Mezi nejpoužívanější charakteristiky polohy řadíme střední hodnotu, modus, medián a různé kvantily.

Charakteristiky variability Určují míru odchýlení hodnot náhodné veličny od její střední hodnoty. Mezi nejpoužívanější charakteristiky variability řadíme rozptyl, směrodatnou odchylku, střední odchylku, pravděpodobnou odchylku, kvartilovou odchylku, variační koeficient a variační rozpětí.

Chrakteristiky šikmosti a špičatosti Jsou to charakteristiky, které popisují tvar křivky frekvenční funkce náhodné veličiny.

Podle toho jakým způsobem pak charakteristiky tvoříme je ještě navíc rozdělujeme na:

Charakteristiky momentu Momentové charakteristiky jsou funkcemi všech hodnot, kterých náhodná veličina může nabývat. Patří mezi ně centrované (centrální) a normované (standardizované) momenty.

Charakteristiky kvantilu Kvantilové charakteristiky jsou hodnoty x_p náhodné veličiny.

3.3.1 Charakteristiky polohy

Střední hodnota náhodné veličiny

Jedná se o základní charakteristiku polohy s jejíž pomocí určujeme další charakteristiky a to nejen charakteristiky polohy. Také se můžeme setkat s označením očekávaná hodnota z anglického ekvivalentu Expected value, z kterého pak vychází její značení.

Definice 3.6 (Střední hodnota náhodné veličiny X) Střední hodnota $E(X)$ náhodné veličiny X je reálné číslo, pro které paltí:

- Jesliže X je diskrétní náhodná veličina:

$$E(X) = \sum_{i=1}^{n} x_i \cdot P(X = x_i) = \sum_{i} x_{i=1}^{n} \cdot p(x_i) \tag{3.18}$$

- Jestliže X je spojitá náhodná veličina:

$$E(X) = \int_{-\infty}^{\infty} x \cdot f(x)\,\mathrm{d}x \tag{3.19}$$

Toto tvrzení platí ze předpokladu, že uvedená řada, respektive integrál absolutně konvergují.

Z definice pak přímo vyplívají následující vlastnosti.

Vlastnosti 3.4

Mějme dvě libovolné náhodné veličiny X, Y a reálnou konstantu k a l. Potom pro střední hodnotu platí:

1. *Střední hodnota konstanty.*
$$E(k) = k \tag{3.20}$$

2. *Střední hodnota součinu konstanty a náhodné veličiny X.*
$$E(k \cdot X) = k \cdot E(X) \tag{3.21}$$

3. *Střední hodnota součinu a součtu konstanty s náhodnou veličinou X.*
$$E(k \cdot X + l) = k \cdot E(X) + l \tag{3.22}$$

4. *Střední hodnota součtu dvou náhodných veličin X a Y.*
$$E(X + Y) = E(X) + E(Y) \tag{3.23}$$
Tuto vlastnost můžeme zobecnit na n náhodných veličin $X_1, X_2, X_3, \ldots, X_n$.
$$\begin{aligned} E(X_1 + X_2 + X_3 + \ldots + X_n) &= E\left(\sum_{i=1}^{n} X_i\right) = \\ &= \sum_{i=1}^{n} E(X_i) = E(X_1) + E(X_2) + E(X_3) + \ldots + E(X_n) \end{aligned}$$

5. *Střední hodnota součinu dvou náhodných veličin X a Y.*
$$E(X \cdot Y) = E(X) \cdot E(Y) \tag{3.24}$$
Tuto vlastnost můžeme zobecnit na n náhodných veličin $X_1, X_2, X_3, \ldots, X_n$.
$$E(X_1 \cdot X_2 \cdot X_3 \cdot \ldots \cdot X_n) = E\left(\prod_{i=1}^{n} X_i\right) = \prod_{i=1}^{n} E(X_i) \tag{3.25}$$
Tato vlastnost ovšem platí pouze v případě, že jsou všechny náhodné veličiny $X_1, X_2, X_3, \ldots, X_n$ vzájemně nezávislé.

Střední hodnota náhodné veličiny je většinou doprovázena další důležitou charakteristikou rozptylem náhodné veličiny, který definujeme až v části Charakteristiky variability.

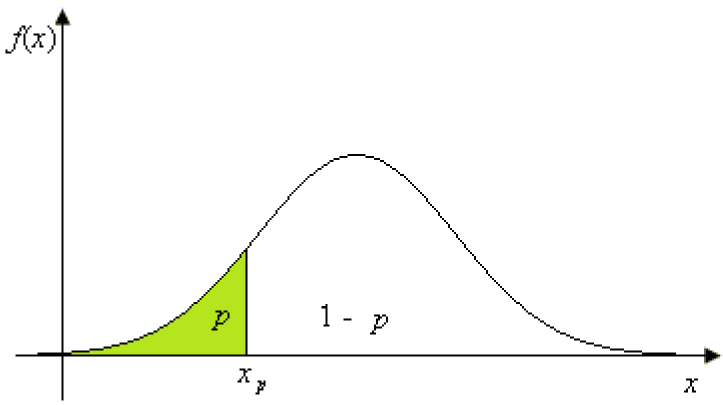
Pokračujme ve výčtu charakteristik polohy, z kterých si ukážeme kvantily a to především speciálními případy - modus, medián, dolní, horní kvartil a percentil.

P-kvantily náhodné veličiny X

Definice 3.7 (P-kvantily náhodné veličiny X) Mějme distribuční funkci $F(x)$ spojité náhodné veličiny X. Potom hodnota x_p, která je řešením rovnice $F(x_p) = p$, nazýváme p-kvantilem rozložení spojité náhodné veličiny X.

P-kvantily jsou tedy čísla, která dělí plochu pod grafem hustoty pravděpodobnosti spojité náhodné veličiny X, právě v poměru:

$$\frac{p}{1-p}$$



Obrázek 3.6: P-kvantil

Nyní slíbené p-kvantily se speciálními názvy.

Medián

Medián je p-kvantil, kde $p = \frac{1}{2}$.

Definice 3.8 (Medián) Medián $Me(X) = \widetilde{x}$ je p-kvantil, kde řešením rovnice $F(\widetilde{x}) = \frac{1}{2}$. Medián $Me(X) = \widetilde{x}$ rozděluje plochu pod křivkou hustoty pravděpodobnosti náhodné veličiny na dvě poloviny.

Definice 3.9 (Dolní kvartil) Dolní kvartil je p-kvantil, kde $p = \frac{1}{4}$. Značíme $x_{\frac{1}{4}}$. Dolní kvartil je tedy řešením rovnice:

$$F\left(x_{\frac{1}{4}}\right) = \frac{1}{4}. \tag{3.26}$$

Dolní kvartil tak dělí plochu pod grafem hustoty pravděpodobnosti v poměru $1:3$.

Definice 3.10 (Horní kvartil) Horní kvartil je p-kvantil, kde $p = \frac{3}{4}$. Značíme $x_{\frac{3}{4}}$. Horní kvartil je tedy řešením rovnice:

$$F\left(x_{\frac{3}{4}}\right) = \frac{3}{4}. \tag{3.27}$$

Horní kvartil tak dělí plochu pod grafem hustoty pravděpodobnosti v poměru $3 : 1$.

Nejspíše nejznámější p-kvantil, který je hojně používaný při porovnávání úspěšnosti testování je percentil.

Definice 3.11 (Percentil) Percentil je p-kvantil, kde $p = \frac{k}{100}$. Značíme $x_{\frac{k}{100}}$. Percentil je tedy řešením rovnice:

$$F\left(x_{\frac{k}{100}}\right) = \frac{k}{100}. \tag{3.28}$$

Percentil tak dělí plochu pod grafem hustoty pravděpodobnosti na 100 stejných částí z nichž vybereme právě k.

Modus

Modus náhodné veličiny X, který značíme $Mod\,(X)$, musíme definovat zvlášť pro diskrétní a spojitou náhodnou veličinu X.

Definice 3.12 (Modus diskrétní náhodné veličiny X) Modus $Mod\,(X)$ diskrétní náhodné veličiny X je hodnota v níž pravděpodobnostní funkce $p\,(x)$ náhodné veličiny X nabývá maxima.

$$Mod\,(X) = x_m \Leftrightarrow P\,(X = x_m) \geq P\,(X = x_i)\,, \tag{3.29}$$

kde $i = 1; 2; 3; \ldots; n$.

Definice 3.13 (Modus spojité náhodné veličiny X) Modus $Mod\,(X)$ spojité náhodné veličiny X je hodnota v níž hustota pravděpodobnosti $f\,(x)$ nabývá lokální maximum.

$$Mod\,(X) = x_m \Leftrightarrow f\,(x_m) \geq f\,(x_i)\,, \tag{3.30}$$

kde $x_i \in \mathbb{R}$.

Navíc modus můžeme jetště rozdělit na:

Unimodální rozložení pravděpodobnosti - náhodná veličina X má právě jeden modus.

Polymodální rozložení pravděpodobnosti - náhodná veličina X má více než jeden modus.

3.3.2 Charakteristika momentů a variability

Základní charakteristikou variability je rozptyl, který většinou doprovází střední hodnotu náhodné veličiny. Definujeme jej komplikovaněji pomocí momentů náhodné veličiny, které zase vycházejí ze střední hodnoty $E\left(X\right)$. Momenty náhodné veličiny pak ještě dělíme na počáteční, centrované a normované.

Definice 3.14 (Počáteční moment r-tého stupně μ_r) Střední hodnota r-té mocniny náhodné veličiny X je r-tý počáteční moment μ_r náhodné veličiny X.

Pro diskrétní náhodnou veličinu X platí:

$$\mu_r = E\left(X^r\right) = \sum_{i=1}^{n} x_i^r \cdot p\left(x_i\right). \tag{3.31}$$

Pro spojitou náhodnou veličinu X platí:

$$\mu_r = E\left(X^r\right) = \int_{\mathbb{R}} x^r \cdot f\left(x\right) \mathrm{d}x. \tag{3.32}$$

Z definice r-tého počátečního momentu dále definujeme střední hodnotu náhodné veličiny X, právě jako počáteční moment prvního stupně:

$$\mu_1 = E\left(X^1\right) = E\left(X\right) \tag{3.33}$$

Definice 3.15 (Centrální moment r-tého stupně ν_r) Mějme střední hodnotu náhodné veličiny X, jejíž střední hodnota $E\left(X\right) = \mu_1$, je zároveň střední hodnota mocniny $\left(X - \mu_1\right)^r$ nazýváme centrální moment r-tého stupně ν_r náhodné veličiny, kdy:

Pro diskrétní náhodnou veličinu X platí:

$$\nu_r = E\left(\left(X - \mu_1\right)^r\right) = \sum_{i=1}^{n} \left(x_i - \mu_1\right)^r \cdot p\left(x_i\right). \tag{3.34}$$

Pro spojitou náhodnou veličinu X platí:

$$\nu_r = E\left(\left(X - \mu_1\right)^r\right) = \int_{\mathbb{R}} \left(x - \mu_1\right)^r \cdot f\left(x\right) \mathrm{d}x. \tag{3.35}$$

Nyní prozkoumáme konkrétní centrální momenty a některé z nich konkrétně pojmenujeme pro jejich užitečné vlastnosti. Získáme je tak, že budeme dosazovat za stupeň momentu.

Centrální moment prvního stupně

Následující platí pro diskrétní i spojitou náhodnou veličinu X.

$$\nu_1 = E\left([X-\mu_1]^1\right) = E\left(X\right) - E\left(\mu_1\right) = \mu_1 - \mu_1 = 0$$

Vzhledem k tomu, že ve všech případech nabývá hodnoty 0 se tento moment prvního stupně v podstatě nepoužívá.

Centrální moment druhého stupně

Diskrétní náhodná veličina X

$$\nu_2 = E\left([X-\mu_1]^2\right) = \sum_{i=1}^{n} \left(x_i - \mu_1\right)^2 \cdot p\left(x_i\right) \tag{3.36}$$

$$rov445\nu_2 = E\left([X-\mu_1]^2\right) = \int_{\mathbb{R}} \left(x - \mu_1\right)^2 \cdot f\left(x\right) \mathrm{d}x \tag{3.37}$$

Centrální moment druhého stupně nazýváme rozptyl náhodné veličiny X a značíme jej σ^2. Také je možné setkat se s označením $D\left(X\right)$, tento název je odvozen z anglického ekvivalentu dispersion. Centrální moment druhého stupně má velice důležité vlastnosti:

Spojitá náhodná veličina X Vlastnosti 3.5

Mějme náhodnou veličinu X, její rozptyl σ^2 a reálné konstanty a, b, c. Pak řekneme, že:

1. *Rozptyl konstanty je nula, protože všechny hodnoty jsou rovny střední hodnotě. To znamená, že žádná hodnota se neodchyluje a tím pádem $\sigma^2 = 0$.*
2. *Rozptyl součinu náhodné veličiny X a konstanty.*
$$D\left(a \cdot X\right) = a^2 \cdot D\left(X\right) \tag{3.38}$$
3. *Rozptyl součinu náhodné veličiny X s konstantou a součtu konstanty.*
$$D\left(a \cdot X + b\right) = a^2 \cdot D\left(X\right) + D\left(b\right) = a^2 \cdot D\left(X\right) + 0 = a^2 \cdot D\left(X\right) \tag{3.39}$$
4. *Rozptyl součtu dvou náhodných veličin X a Y.*
$$D\left(X + Y\right) = D\left(X\right) + D\left(Y\right) \tag{3.40}$$

Tuto vlastnost zle také zobecnit pro součet náhodných veličin $X_1; X_2; X_3; \ldots; X_n$.

$$D\left(X_1+X_2+X_3+\ldots+X_n\right)=\sum_{i=1}^{n} D\left(X_i\right)=D\left(X_1\right)+D\left(X_2\right)+D\left(X_3\right)+\ldots+D\left(X_n\right) \tag{3.41}$$

Zároveň musí platit, že náhodné veličiny $X_1; X_2; X_3; \ldots; X_n$ musí být nezávislé.

5. *Výpočet rozptylu náhodné veličiny X pomocí střední hodnoty nebo pomocí počátečního momentu.*

$$\sigma^2=E\left(X^2\right)-[E(X)]^2=E\left(X^2\right)-E^2(X)=\mu_2-\mu_1^2 \tag{3.42}$$

Odmocněním rozptylu získáme směrodatnou odchylku $\pm\sigma$, kterou v praxi používáme častěji, protože na rozdíl od rozptylu udává vzdálenost a ne plochu čtverce jako rozptyl. Zjednodušeně můžeme říci, že směrodatná odchylka nám dá odchylku se správnou jednotkou - např. směrodatná odchylka délky je také délka, ale rozptyl nás informuje v plošných jednotkách.

Výpočet centrálních momentů vyšších stupňů je poměrně komplikovaný. Proto si jejich výpočet ukážeme pomocí počátečních momentů:

$$\begin{aligned}
\nu_r &= \sum_{k=0}^{r}(-1)^k \cdot \binom{r}{k} \cdot \mu_{r-k} \cdot \mu_1^k \\
\nu_3 &= \mu_3-3\mu_2 \cdot \mu_1+2\mu_1^3 \\
\nu_4 &= \mu_4-4\mu_1 \cdot \mu_3+6\mu_1^2 \cdot \mu_2-3\mu_1^4
\end{aligned}$$

Z charakteristik momentu ještě prozkoumáme normovaný moment r-tého stupně, jehož konkrétní případy si ukážeme až u dané charakteristiky, kterou vyjadřuje.

Definice 3.16 (Normovaný moment r-tého stupně $\overline{\nu_r}$) Normovaný moment r-tého stupně $\overline{\nu_r}$ náhodné veličiny X je definován vztahem:

$$\overline{\nu_r}=\frac{\nu_r}{\sigma^r}, \tag{3.43}$$

kde ν_r je centrální moment r-tého stupně a σ^r je r-tá mocnina směrodatné odchylky náhodné veličiny X.

Nyní tedy konkrétní normované momenty náhodné veličiny, stejně jako u centrálních momentů je získáme dosazením za stupeň momentu.

Normovaný moment prvního a druhého stupně náhodné veličiny X

$$\overline{\nu_1} = \frac{\nu_1}{\sigma^1} = \frac{0}{\sigma} = 0$$
$$\overline{\nu_2} = \frac{\nu_2}{\sigma^2} = \frac{\nu_2}{\nu_2} = 1$$

Jak vidíme hodnoty normovaného momentu prvního a druhého stupně náhodné veličiny X budou mít vždy hodnotu 0 a 1. Proto se při zpracování naměřených dat příliš často nepoužívají. Další stupně normovaného momentu si více přiblížíme až v části Charakteristiky šikmosti a špičatosti.

Průměrná odchylka střední hodnoty náhodné veličiny X

Průměrná odchylka nás informuje o tom, jak se průměrně odchylují hodnoty náhodné veličiny X od její střední hodnoty.

Definice 3.17 (Průměrná odchylka) Průměrnou nebo také střední odchylkou od střední hodnoty náhodné veličiny X nazveme nezáporné číslo, pro které platí:

$$\nu = E\left(\left|X - E\left(X\right)\right)\right|) = E\left(\left|x - \mu_1\right|\right) \tag{3.44}$$

Definice 3.18 (Pravděpodobná odchylka) Pravděpodobná odchylka od střední hodnoty náhodné veličiny X nazveme nezápornou veličinu ϵ, pro kterou platí rovnice:

$$P\left(\mu_1 - \epsilon < X < \mu_1 + \epsilon\right) = \frac{1}{2}. \tag{3.45}$$

Poznámka: Pravděpodbnou odchylku chápeme jako hodnotu, která se rovná mediánu absolutních hodnot odchylek náhodné veličiny X.

Definice 3.19 (Kvartilová odchylka) Kvartilová odchylka od střední hodnoty náhodné veličiny X je rozdíl horního a dolního kvartilu náhodné veličiny X.

$$Ko\left(X\right) = x_{\frac{3}{4}} - x_{\frac{1}{4}}. \tag{3.46}$$

Poznámka: Analogicky můžeme definovat i decilovou odchylku a percetilovou odchylku náhodné veličiny. Kvartilová odchylka je pak užitečnou alternativou směrodatné odchylky. Její výhodou je, že není tak velkým způsobem ovlivněna příliš malými nebo velkými hodnotami náhodné veličiny X.

Definice 3.20 (Variační koeficient) Variační koeficient náhodné veličiny definujeme pro nenulovou střední hodnotu jako podíl směrodatné odchylky a střední hodnoty náhodné veličiny X.

$$\vartheta = \frac{\sigma}{\mu} = \frac{\sqrt{D\left(X\right)}}{E\left(X\right)} \tag{3.47}$$

Variační koeficient určuje míru spolehlivosti střední hodnoty. Většinou je udáván v procentech. Obvykle hodnotíme střední hodnotu podle variačního koeficientu tak, že pokud $\vartheta < 0,10$ je střední hodnota spolehlivá. V opačném případě řekneme, že střední hodnota není příliš spolehlivá (Necharakterizuje příliš dobře hodnoty náhodné veličiny X).

Poslední zde zmíněnou charakteristikou variability je variační rozpětí, které nás informuje o tom, jak moc se maximální a minimální hodnota náhodné veličiny vzdaluje od její střední hodnoty. Také z něj poznáme jak daleko jsou od sebe tyto hodnoty vzdálené.

Definice 3.21 (Variační rozpětí) Variační rozpětí náhodné veličiny X definujeme jako rozdíl maximální a minimální hodnoty náhodné veličiny X.

$$R = x_{\text{max}} - x_{\text{min}} \tag{3.48}$$

Touto variační charakteristikou ukončíme výčet těchto charakteristik a projdeme si poslední skupinu charakteristik náhodné veličiny X.

3.3.3 Charakteristiky šikomosti a špičatosti náhodné veličiny X

Jak jsme již uvedli tyto charakteristiky vyplívají z normovaných momentů náhodné veličiny X. Zatím jsme prozkoumali pouze normované momenty prvního a druhého stupně. Ty se pro jejich konstantní hodnoty příliš nepoužívají. Naopak normované momenty vyšších stupňů pak využijeme a přiřadíme jim vlastní názvy.

Koeficient šikmosti

Tento koeficient nás informuje o nesymetrii hodnot náhodné veličiny X.

Definice 3.22 (Koeficient šikmosti náhodné veličiny X) Normovaný moment třetího stupně náhodné veličiny X:

$$\overline{\nu_3} = \frac{\nu_3}{\sigma^3} = \frac{E\left(\left|X - E\left(X\right)\right|^3\right)}{\sigma^3} = \gamma_1 \tag{3.49}$$

nazveme koeficient šikmosti $\gamma_1 \in \mathbb{R}$ nebo také $A_3(X)$.

Z definice pak přímo vyplívají následující vlastnosti.

Vlastnosti 3.6

1. *Pro koeficient šikmosti platí* $\gamma_1 = \overline{\nu_3} = 0$, *jestliže je frekvenční funkce náhodné veličiny* X *symetrická.*
2. *Pro koeficient šikmosti platí* $\gamma_1 = \overline{\nu_3} < 0$, *jestliže je frekvenční funkce náhodné veličiny* X *zešikmená doprava.*
3. *Pro koeficient šikmosti platí* $\gamma_1 = \overline{\nu_3} > 0$, *jestliže je frekvenční funkce náhodné veličiny* X *zešikmená doleva.*
4. *Mějme libovolnou náhodnou veličinu* X *a reálná čísla* a, b, *pak platí:*

$$A_3(a \cdot X + b) = A_3(X). \tag{3.50}$$

Poznámka: Podle uvedených vlastností je možné předem odhadnout znaménko koeficientu šikmosti, jestliže si načrtneme graf frekvenční funkce a rozhodneme, na kterou stranu jsou zešikmené odchylky střední hodnoty.

Koeficient špičatosti

Tento koeficient nás informuje o tom, jak je rozdělení náhodné veličiny špičaté (strmé) nebo naopak ploché.

Definice 3.23 (Koeficient špičatosti) Normovaný moment čtvrtého stupně náhodné veličiny X:

$$\overline{\nu_4} - 3 = \frac{\nu_4}{\sigma^4} - 3 = \frac{E\left(|X - E(X)|^4\right)}{\sigma^4} - 3 = \gamma_2 \tag{3.51}$$

nazveme koeficient šikmosti $\gamma_2 \in \mathbb{R}$ nebo také $A_4(X)$.

Z definice pak přímo vyplívají následující vlastnosti.

Vlastnosti 3.7

1. *Pro koeficient špičatosti platí $\gamma_2 = \overline{\nu_4} - 3 = 0$, jestliže má náhodná veličina X normální rozdělení.*
2. *Pro koeficient špičatosti platí $\gamma_1 = \overline{\nu_4} - 3 < 0$, jestliže je maximum frekvenční funkce náhodné veličiny X „tupé“.*
3. *Pro koeficient špičatosti platí $\gamma_2 = \overline{\nu_4} - 3 > 0$, jestliže je maximum frekvenční funkce náhodné veličiny X „špičaté“.*
4. *Mějme libovolnou náhodnou veličinu X a reálná čísla a, b, pak platí:*

$$A_4\,(a \cdot X + b) = A_4\,(X) \tag{3.52}$$

Koeficient šikmosti a špičatosti většinou počítáme kvůli jednoduchosti z počátečních momentů:

$$\overline{\nu_r} = \sum_{k=0}^{r} (-1)^k \cdot \binom{r}{k} \cdot \mu_{r-k} \cdot \mu_1^k \tag{3.53}$$

$$\gamma_1 = \frac{\mu_3 - 3\mu_1\mu_2 + 2\mu_1^3}{\sigma^3} \tag{3.54}$$

$$\gamma_2 = \frac{\mu_4 - 4\mu_1\mu_3 + 6\mu_1\mu_2 - 3\mu_1^4}{\sigma^4} - 3 \tag{3.55}$$

■ **Řešený příklad 3.3** Náhodná veličina X je určena tabulkou (3.1) pravděpodobnostní funkce.

Tabulka 3.1: Pravděpodobnostní funkce k příkladu (3.3)

x_i	1	2	3	4	5	6	7	8	9	10
$p\,(x_i)$	0,050	K	0,125	0,080	0,095	0,105	0,055	0,135	0,195	0,085

Určete hodnotu pravděpodobnostní funkce $p\,(2) = K$, distribuční funkci, pravděpodobnost jevu, že náhodná veličina X nabude hodnoty $P\,(x < 6)$, $P\,(x > 4)$, $P\,(3 < x < 7)$ a charakteristiky rozdělení pravděpodobnosti.

Řešení:

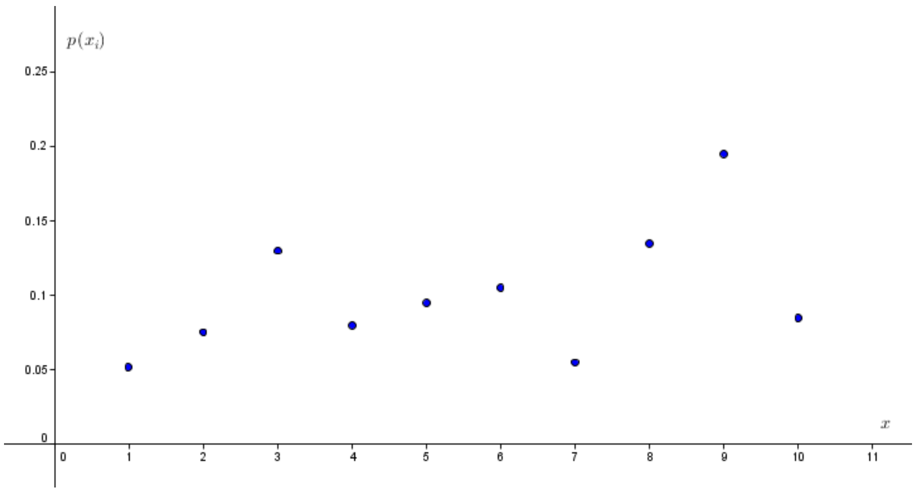
Podle vlastností pravděpodobnostní funkce určíme hodnotu pravděpodobnostní funkce $p\,(0) = K$.

$$
\begin{aligned}
1 &= \sum_{i=1}^{n} p\,(x) \\
p\,(0) &= 1 - p\,(1) - p\,(3) - p\,(4) - p\,(5) - p\,(6) - p\,(7) - p\,(8) - p\,(9) - p\,(10) = \\
&= 0,075
\end{aligned}
$$

Nyní můžeme určit distribuční funkci náhodné veličiny X tabulka (3.2).

Tabulka 3.2: Distribuční funkce k příkladu (3.3)

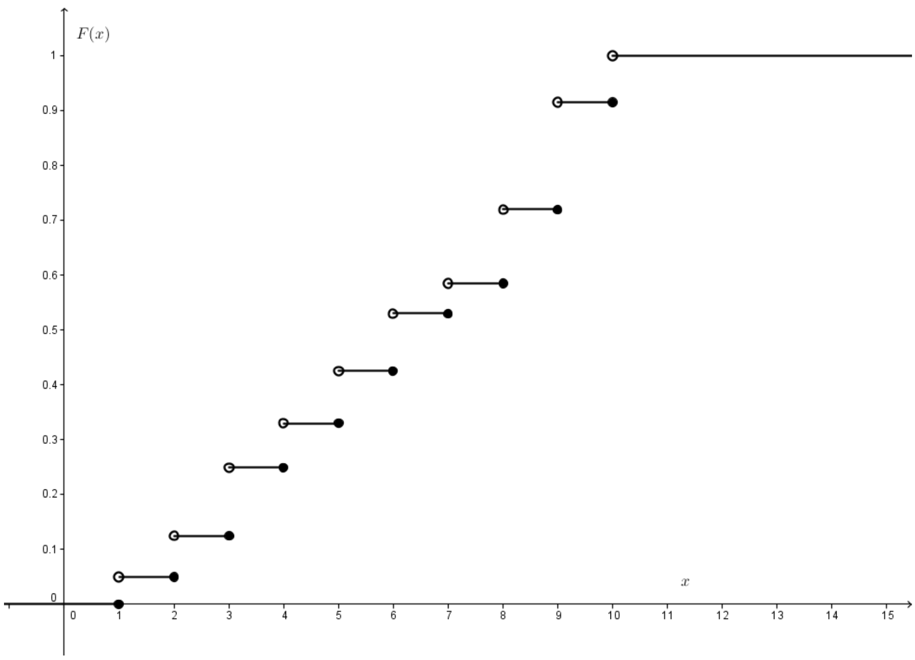
x_i	$x < 1$	1	2	3	4	5
$F\,(x_i)$	0	0	0,050	0,125	0,250	0,330
-	6	7	8	9	10	$x > 10$
-	0425	530	0,585	0,720	0,915	1



Obrázek 3.7: Pravděpodobnostní funkce

Určíme pravděpodobnost $P\,(x < 6)$, $P\,(x > 4)$ a $P\,(3 \leq x < 7)$:

$$
\begin{aligned}
P\,(x < 6) &= F\,(6) = P\,(1) + P\,(2) + P\,(3) + P\,(4) + P\,(5) = 0,425 \\
P\,(x > 4) &= F\,(x > 10) - F\,(5) = P\,(5) + P\,(6) + P\,(7) + P\,(8) + P\,(9) + P\,(10) = \\
&= 0,670 \\
P\,(3 \leq x < 7) &= F\,(7) - F\,(3) = P\,(3) + P\,(4) + P\,(5) + P\,(6) = 0,405
\end{aligned}
$$



Obrázek 3.8: Distribuční funkce

Nejprve určíme střední hodnotu a rozptyl náhodné veličiny X podle definice (3.6) a podle rovnice (3.36). K výpočtu se nám bude hodit vyplnit si tabulku (3.3) s počátečními momenty.

Střední hodnota, rozptyl a směrodatná odchylka.

$$\begin{aligned} E\left(X\right) &= \sum_{i=1}^{n} x_i p\left(x_i\right) = 6,07 \\ D\left(X\right) &= E\left(X^2\right) - \left[E\left(X\right)\right]^2 = 7,6951 \\ \sigma &= \sqrt{D\left(X\right)} = 2,774 \end{aligned}$$

Koeficient šikmosti a špičatosti určíme podle definic (3.22) a (3.23). Ještě před samotným výpočtem si vypíšeme hodnoty počátečních momentů.

$$\mu_1 = 6,07 \qquad \mu_2 = 44,54 \qquad \mu_3 = 358,84 \qquad \mu_4 = 3041,72$$

Koeficient šikmosti a špičatosti pomocí počátečních momentů:

Tabulka 3.3: Tabulka s počátečními momenty

x_i	1	2	3	4	5
$p(x_i)$	0,050	0,075	0,125	0,080	0,095
$x_i \cdot p(x_i)$	0,050	0,150	0,375	0,320	0,475
$x_i^2 \cdot p(x_i)$	0,050	0,300	1,125	1,280	2,375
$x_i^3 \cdot p(x_i)$	0,050	0,600	3,375	5,120	11,875
$x_i^4 \cdot p(x_i)$	0,050	1,200	10,125	20,480	59,375
$\lvert x_i - \mu_1 \rvert \cdot p(x_i)$	0,254	0,305	0,384	0,166	0,102
-	6	7	8	9	10
-	0,105	0,055	0,135	0,195	0,085
-	0,630	0,385	1,080	1,755	0,850
-	2,780	2,695	8,640	15,795	8,500
-	22,680	18,865	69,120	142,155	85
-	136,080	132,055	552,960	1279,395	850
-	0,007	0,051	0,261	0,571	0,334

$$
\begin{aligned}
\gamma_1 &= \frac{\nu_3}{\sigma^3} = \frac{\mu_3 - 3\mu_1\mu_2 + 2\mu_1^3}{\sigma^3} = \frac{358,84 - 3 \cdot 44,54 \cdot 6,07 + 2 \cdot 6,07^3}{2,774^3} = \\
&= -0,231 \\
\gamma_2 &= \frac{\nu_4}{\sigma^4} - 3 = \frac{\mu_4 - 4\mu_1\mu_3 + 6\mu_1^2\mu_2 - 3\mu_1^4}{\sigma^4} - 3 = \\
&= \frac{3041,72 - 4 \cdot 6,07 \cdot 358,84 + 6 \cdot 6,07^2 \cdot 44,54 - 3 \cdot 6,07^4}{2,774^4} - 3 = -1,2626
\end{aligned}
$$

Ještě určíme průměrnou odchylku, variační koeficient a variační rozpětí podle definic (3.17), (3.20) a (3.21).

$$
\begin{aligned}
\nu &= E(|x - \mu_1|) = \sum_{i=1}^{n} |x_i - \mu_1| \cdot p(x_i) = 2,435 \\
\vartheta &= \frac{\sigma}{\mu_1} = \frac{2,774}{6,07} = 0,457 \\
R &= x_{\text{max}} - x_{\text{min}} = 10 - 1 = 9
\end{aligned}
$$

Vyhodnotíme získané charakteristiky. Střední hodnota $E(X) = 6,07$ se směrodatnou odchylkou $\sigma = 2,774$ nevystihuje náhodnou veličinu příliš dobře, protože variační koeficient $\vartheta = 0,457$ je velký. Pravděpodobnostní funkce je zešikmená vpravo a je spíše tupá.

■ **Řešený příklad 3.4** Spojitá náhodná veličina je dána distribuční funkcí:

$$\begin{aligned} F(x) &= 0 \qquad x \in \left(-\infty; \frac{\sqrt{6}}{3}\right\rangle \\ F(x) &= ax^2 + b \qquad x \in \left(\frac{\sqrt{6}}{3}; \sqrt{2}\right) \\ F(x) &= 1 \qquad x \in \left\langle\sqrt{2}; \infty\right) \end{aligned}$$

Určete koeficienty a a b, hustotu pravděpodobnosti, pravděpodobnost $P\left(x < \frac{5}{6}\right)$, $P(x > 1)$, $P\left(1 < x < \frac{5}{4}\right)$, střední hodnotu, rozptyl, směrodatnou odchylku, horní a dolní kvartil, kvartilovou odchylku, medián, modus, koeficient šikmosti a špičatosti, variační koeficient a variační rozptyl.

Řešení:

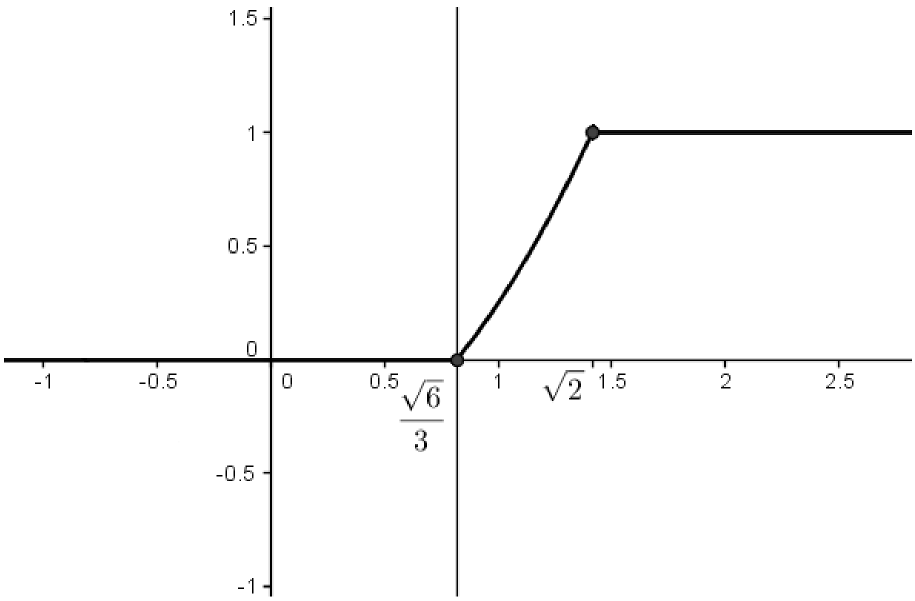
Z vlastností distribuční funkce určíme hodnoty koeficientů a a b.

$$\begin{aligned} F\left(\frac{\sqrt{6}}{3}\right) &= 0 = a \cdot \left(\frac{\sqrt{6}}{3}\right)^2 + b \\ F\left(\sqrt{2}\right) &= 1 = a \cdot \left(\sqrt{2}\right)^2 + b \qquad \Rightarrow \qquad b = 1 - 2a \\ 0 &= a\frac{2}{3} + 1 - 2a \\ \frac{4}{3} &= 1 \\ a &= \frac{3}{4} \qquad \Rightarrow \qquad b = -\frac{1}{2} \end{aligned}$$

Distribuční funkce má tvar:

$$\begin{aligned} F(x) &= 0 \qquad x \in \left(-\infty; \frac{\sqrt{6}}{3}\right\rangle \\ F(x) &= \frac{3}{4}x^2 - \frac{1}{2} \qquad x \in \left(\frac{\sqrt{6}}{3}; \sqrt{2}\right) \\ F(x) &= 1 \qquad x \in \left\langle\sqrt{2}; \infty\right) \end{aligned}$$

Pomocí vlastností hustoty pravděpodobnosti $f(x) = F'(x)$ ji určíme:

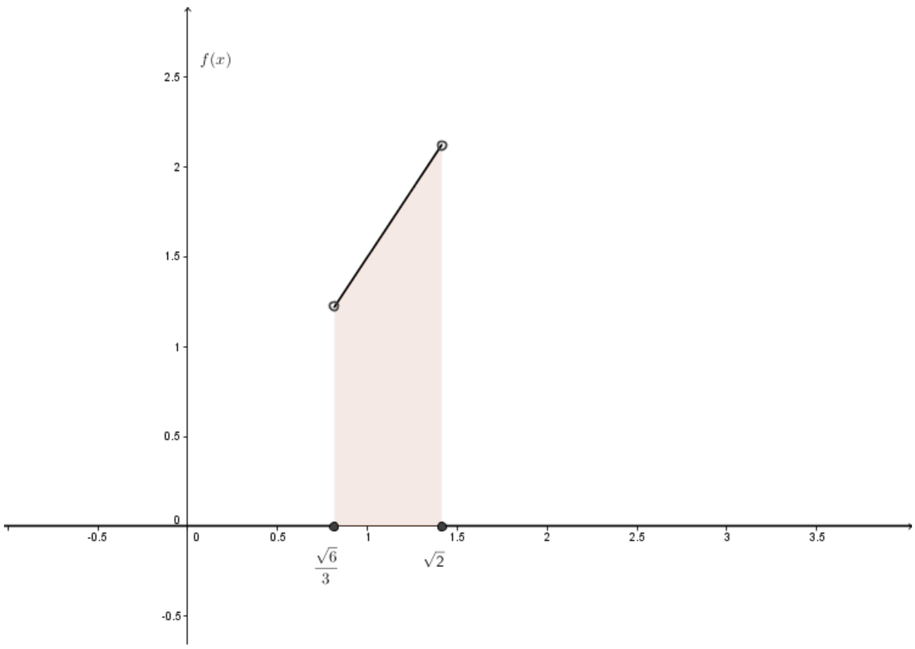


Obrázek 3.9: Distribuční funkce

$$
\begin{aligned}
f(x) &= \frac{\mathrm{d}\left(\frac{3}{4}x^2 - \frac{1}{2}\right)}{\mathrm{d}x} = \frac{3}{2}x \qquad x \in \left(\frac{\sqrt{6}}{3}; \sqrt{2}\right) \\
f(x) &= 0 \qquad x \notin \left\langle \frac{\sqrt{6}}{3}; \sqrt{2} \right\rangle
\end{aligned}
$$

Určíme pravděpodobnost výskytu veličiny X - $P\left(x < \frac{5}{6}\right)$, $P(x > 1)$ a $P\left(1 < x < \frac{5}{4}\right)$. Můžeme je spočítat dvěma způsoby.

$$
\begin{aligned}
P\left(x < \frac{5}{6}\right) &= F\left(\frac{5}{6}\right) = \frac{3}{4} \cdot \left(\frac{\sqrt{5}}{6}\right)^2 - \frac{1}{2} = 0,0208 \\
&= \int\limits_{-\infty}^{\frac{5}{6}} f(x)\,\mathrm{d}x = \underbrace{\int\limits_{-\infty}^{\frac{\sqrt{6}}{3}} f(x)\,\mathrm{d}x}_{0} + \int\limits_{\frac{\sqrt{6}}{3}}^{\frac{5}{6}} f(x)\,\mathrm{d}x = \frac{3}{2} \cdot \int\limits_{\frac{\sqrt{6}}{3}}^{\frac{5}{6}} x\mathrm{d}x = \\
&= = \frac{3}{2} \cdot \left[\frac{x^2}{2}\right]_{\frac{\sqrt{6}}{3}}^{\frac{5}{6}} = \frac{3}{4} \cdot \left(\frac{25}{36} - \frac{6}{9}\right) = 0,0208 \\
P(x > 1) &= F\left(\sqrt{2}\right) - F(1) = 1 - \frac{3}{4} \cdot 1 - \frac{1}{2} = 0,75
\end{aligned}
$$



Obrázek 3.10: Hustota pravděpodobnosti

$$P\left(1 < x < \frac{5}{4}\right) = F\left(\frac{5}{4}\right) - F(1) = \frac{3}{4}\cdot\frac{25}{16} - \frac{1}{2} - 1 + \frac{1}{2} = 0,5625$$

K určení střední hodnoty, rozptylu, koeficientu šikmosti a špičatosti potřebujeme počáteční momenty:

$$\mu_1 = \int_{-\infty}^{\infty} x\cdot f(x)\,\mathrm{d}x = \frac{3}{2}\cdot\int_{\frac{\sqrt{6}}{3}}^{\sqrt{2}} x^2\mathrm{d}x = \frac{3}{2}\cdot\left[\frac{x^3}{3}\right]_{\frac{\sqrt{6}}{3}}^{\sqrt{2}} = \frac{1}{2}\cdot\left(\sqrt{2}^3 - \left[\frac{\sqrt{6}}{3}\right]^3\right) =$$

$$= \sqrt{2} - \frac{\sqrt{6}}{3} = 1,142$$

$$\mu_2 = \int_{-\infty}^{\infty} x^2\cdot f(x)\,\mathrm{d}x = \frac{3}{2}\cdot\int_{\frac{\sqrt{6}}{3}}^{\sqrt{2}} x^3\mathrm{d}x = \frac{3}{2}\cdot\left[\frac{x^4}{4}\right]_{\frac{\sqrt{6}}{3}}^{\sqrt{2}} = \frac{3}{8}\cdot\left(4 - \frac{4}{9}\right) = \frac{4}{3}$$

$$\mu_3 = \int_{-\infty}^{\infty} x^3\cdot f(x)\,\mathrm{d}x = \frac{3}{2}\cdot\int_{\frac{\sqrt{6}}{3}}^{\sqrt{2}} x^4\mathrm{d}x = \frac{3}{10}\cdot\left[\frac{x^5}{5}\right]_{\frac{\sqrt{6}}{3}}^{\sqrt{2}} = \frac{3}{10}\cdot\left(\sqrt{2}^5 - \left(\frac{\sqrt{6}}{3}\right)^5\right) = 1,588$$

$$\mu_4 = \int_{-\infty}^{\infty} x^4 \cdot f(x)\,\mathrm{d}x = \frac{3}{2} \cdot \int_{\frac{\sqrt{6}}{3}}^{\sqrt{2}} x^5 \mathrm{d}x = \frac{3}{2} \cdot \left[\frac{x^6}{6}\right]_{\frac{\sqrt{6}}{3}}^{\sqrt{2}} = \frac{3}{12} \cdot \left(8 - \frac{8}{27}\right) = 1,926$$

Z počátečních momentů určíme střední hodnotu, rozptyl, směrodatnou odchylku, koeficient šikmosti a špičatosti.

$$\begin{aligned}
E(X) &= \mu_1 = 1,142 \\
D(X) &= \mu_2 - \mu_1^2 = \frac{4}{3} - 1,142^2 = 0,029 \\
\sigma &= \sqrt{D(X)} = 0,17 \\
\gamma_1 &= \frac{\mu_3 - 3\mu_1\mu_2 + 2\mu_1^3}{\sigma^3} = \frac{1,588 - 3 \cdot 1,142 \cdot \frac{4}{3} + 2 \cdot 1,142^3}{0,17^3} = -0,262 \\
\gamma_2 &= \frac{m_4 - 4\mu_1\mu_3 + 6\mu_1^2\mu_2 - \mu_1^4}{\sigma^4} = \\
&= \frac{1,926 - 4 \cdot 1,142 \cdot 1,588 + 6 \cdot 1,142^2 \cdot \frac{4}{3} - 3 \cdot 1,142^4}{0,17^4} = 3,349
\end{aligned}$$

Pomocí distribuční funkce můžeme určit kvartily a modus.

$$\begin{aligned}
F(x_{0,25}) &= \frac{1}{4} \\
\frac{3x_{0,25}^2 - 2}{4} &= \frac{1}{4} \\
3x_{0,25}^2 &= 3 \\
x_{0,25} &= 1 \\
F(x_{0,5}) &= \frac{1}{2} \\
\frac{3x_{0,5}^2 - 2}{4} &= \frac{1}{2} \\
3x_{0,5}^2 &= 4 \\
x_{0,5} &= \frac{2\sqrt{3}}{3} = Me(X) \\
F(x_{0,75}) &= \frac{3}{4} \\
\frac{3x_{0,75}^2 - 2}{4} &= \frac{3}{4} \\
3x_{0,75}^2 &= 5
\end{aligned}$$

$$x_{75} = \frac{\sqrt{15}}{3}$$

K nalezení modusu je potřeba nalézt maximum hustoty pravděpodobnosti. Vzhledem k tomu, že hustota pravděpodobnosti je lineární funkce, je maximum v bodě $x = \sqrt{2} = Mod\,(X)$.

Poslední charakteristiky, které musíme určit kvartilová odchylka, variační koeficient a variační rozpětí podle definic (3.19), (3.20) a (3.21).

$$\begin{aligned} Ko &= x_{0,75} - x_{0,25} = \frac{\sqrt{15}}{3} - 1 = 0,29 \\ \vartheta &= \frac{\sigma}{\mu_1} = \frac{0,17}{1,142} = 0,149 \\ R &= \sqrt{2} - \frac{\sqrt{6}}{3} = 0,5977 \end{aligned}$$

Vyhodnotíme získané charakteristiky. Střední hodnota $E\,(X) = 1,142$ se směrodatnou odchylkou $\sigma = 0,17$ s jistou tolerancí vystihuje náhodnou veličiny celkem dobře, protože variační koeficient $\vartheta = 0,149$ je skoro v mezích únosnosti. Pravděpodobnostní funkce je zešikmená vpravo a je špičatá.

Úlohy:

1. Náhodná veličina X je dána tabulkou (3.4) pravděpodobnostní funkcí. Určete hod-

Tabulka 3.4: Tabulka pravděpodobnostní funkce k úloze (1.)

x_i	1	2	3	4
$p(x_i)$	0,3	K	0,4	0,2

notu K, střední hodnotu, rozptyl, koeficient šikmosti a špičatosti náhodné veličiny X.

2. Náhodná veličina je dána pravděpodobnostní funkcí.

$$\begin{aligned} f(x) &= 2x \qquad x \in \langle 0; 1\rangle \\ &= 0 \qquad x \notin \langle 0; 1\rangle \end{aligned}$$

Určete střední hodnotu, rozptyl, koeficient šikmosti a špičatosti náhodné veličiny X.

3. Náhodná veličina X je dána distribuční funkcí:

$$\begin{aligned} F(x) &= 0 \qquad x < 3 \\ &= \frac{x}{3} - 1 \qquad x \in \langle 3; 6\rangle \\ &= 1 \qquad x \geq 6 \end{aligned}$$

Určete hustotu pravděpodobnosti, a pravděpodobnost $P\left(\frac{3}{2} \leq x \leq 4\right)$.

4. Náhodná veličina X je dána hustotou pravděpodobnosti:

$$\begin{aligned} f(x) &= x - \frac{1}{2} \qquad x \in \langle 1; 2) \\ f(x) &= 0 \qquad x \notin 1; 2) \end{aligned}$$

Určete distribuční funkci náhodné veličiny X.

5. Náhodná veličina X je dána tabulkou (3.5) pravděpodobnostní funkce. Určete ko-

Tabulka 3.5: Tabulka distribuční funkce k úloze (5.)

x_i	1	2	3
$p(x_i)$	0,5	0,3	0,2

eficienty šikmosti a špičatosti náhodné veličiny X.

Tabulka 3.6: Tabulka distribuční funkce k úloze (6.)

x_i	2	3	4	5
$p(x_i)$	0,2	0,3	0,3	0,2

6. Náhodná veličina X je dána tabulkou (3.6) pravděpodobnostní funkce. Určete koeficienty šikmosti a špičatosti náhodné veličiny X.

7. Náhodná veličina X je dána hustotou pravděpodobnosti:

$$\begin{aligned} f(x) &= \frac{3}{4} \cdot \left(2x - x^2\right) \qquad x \in \langle 0; 2\rangle \\ &= 0 \qquad x \notin 0; 2\rangle \end{aligned}$$

   Určete modus a medián náhodné veličiny X.

---

1. $K = 0,1$, $E(X) = 2,5$, $D(X) = 1,25$, $\gamma_1 = -0,21$, $\gamma_2 = -1,36$
2. $E(X) = \frac{2}{3}$, $D(X) = \frac{1}{18}$, $\gamma_1 = -0,43$, $\gamma_2 = -0,4$
3. $f(x) = \frac{1}{2} \qquad x \in \langle 3; 6)$, $f(x) = 0 \qquad x \notin \langle 3; 6)$, $P\left(\frac{3}{2} \leq x \leq 4\right) = \frac{1}{2}$
4. $F(x) = 0 \qquad x \in (-\infty; 1)$, $F(x) = \frac{x}{2} - \frac{1}{2} \qquad x \langle 1; 2)$, $F(x) = 1 \qquad x \in \langle 2; \infty)$
5. $\gamma_1 = 0,579$, $\gamma_2 = -1,1357$
6. $\gamma_1 = 0$, $\gamma_2 = -1,129$
7. $Mod(X) = 1$, $Me(X) = 1$

Kapitola 4

Rozdělení pravděpodobností diskrétní náhodné veličiny

Jak jsme se již dozvěděli pro každou náhodnou veličinu X existují pravidla, která každému jevu přiřazují určitou pravděpodobnost výskytu. Výše jsme si ukázali jak kvalitativně popsat a charakterizovat libovolnou náhodnou veličinu. Většinu náhodných veličin je možné rozlišit podle toho, jaké hodnoty pravděpodobnosti nabývá. Takových rozdělení je velké množství a proto si ukážeme především ta základní. Zároveň je velice důležité rozlišit diskrétní a spojitou náhodnou veličinu X.

4.1 Binomické rozdělení

Mezi všemi pokusy, které vykonáme je možné nalézt velké množství takových u nichž existují pouze dva výsledky. Tím myslíme úspěšný pokus nebo neúspěšný. Zároveň pro ně platí, že každý pozorovatel může výsledky daného pokusu rozlišit jiným způsobem, ale stále se jedná o dvě možnosti výsledku. Vezměme si příklad. Při hodu mincí je pro pozorovatel A úspěchem, když padne líc, naopak pro pozorovatele B je úspěch, když padne rub. Z příkladu je vidět, že ať padne jakákoli strana mince, vždy bude pokus pro jednoho z pozorovatelů úspěchem a pro druhého neúspěchem. Toto označení, ale nijak neovlivní samotné rozdělení pravděpodobnosti. Proto můžeme klidně říci, že nezáleží na označení výsledku: Zdar × Nezdar.

Ve většině případů se také setkáme s tím, že výsledku, který označíme jako úspěch přiřadíme hodnotu 1 a neúspěchu 0. Máme tedy náhodnou veličinu jejíž obor hodnot je dvouprvková množina $H = \{0; 1\}$. Zároveň jsme schopni určit pravděpodobnost, že nastane úspěch a neúspěch. Pro potřeby určení pravděpodobnosti označme úspěch jako jev A a neúspěch $\overline{A}$. Potom platí:

$$P(A) = P(X = 1) = p \tag{4.1}$$

$$P\left(\overline{A}\right) = P(X = 0) = 1 - p \tag{4.2}$$

Můžeme-li navíc pokus opakovat n-krát a každé provedení takového pokusu neovlivní výsledek dalších pokusů. Potom můžeme celému souboru takových pokusů přiřadit číslo x, které v celé sérii určuje právě počet příznivých výsledků. Takto získanou náhodnou veličinu nazýváme Binomická náhodná veličina X a její rozdělení pravděpodobností nazveme binomické.

Definice 4.1 (Binomické rozdělení náhodné veličiny X) Náhodná veličina X s binomickým rozdělením $Bi(n; p)$ má pravděpodobnostní funkci:

$$p(x) = \binom{n}{x} \cdot p^x \cdot (1 - p)^{n-x}, \tag{4.3}$$

kde $x = 1; 2; 3; \ldots; n$ udávající počet sledovaných příznivých výsledků, n je počet všech nezávislých pokusů a p je pravděpodobnost úspěchu v každém pokusu.

V předchozí kapitole jsme se naučili přiřazovat každé náhodné veličině její číselné charakteristiky. Ukažme si tedy nejčastěji užívané charakteristiky pro náhodnou veličinu X s binomickým rozdělením pravděpodobnosti.

Vlastnosti 4.1

1. *Střední hodnota*
$$E(X) = n \cdot p. \tag{4.4}$$
2. *Rozptyl*
$$\sigma^2 = n \cdot p \cdot (1 - p) \tag{4.5}$$
3. *Koeficient šikmosti*
$$\gamma_1 = \frac{1 - 2p}{\sqrt{n \cdot p \cdot (1 - p)}} \tag{4.6}$$
4. *Koeficient špičatosti*
$$\gamma_2 = \frac{1 - 6p \cdot (1 - p)}{n \cdot p \cdot (1 - p)} \tag{4.7}$$

5. *Koeficient koncentrace*

$$\overline{\nu_4} = \gamma_2 + 3 \tag{4.8}$$

6. *Nejpravděpodobnější počet příznivých pokusů je přirozené číslo x_i, které je řešením rovnice:*

$$(n+1) \cdot p - 1 \leq x_i \leq (n+1) \cdot p \tag{4.9}$$

Binomické rozdělení náhodné veličiny X nejčastěji nalezneme u náhodného výběru s vracením. Tím myslíme, že zkoumaný (vybraný) prvek vrátíme zpět do souboru prvků. Tetno prvek je možné zkontrolovat. Tak jsme zajistili, že pravděpodobnost úspěchu, respektive neúspěchu, je stále stejná.

Speciální příklad. Mějme náhodnou veličinu s binomickým rozdělením, u které vykonáme pouze jeden pokus. Takové rozdělení pak zapíšeme takto $Bi\,(1; p)$. Vzhledem k tomu, že se jedná o speciální případ binomického rozdělení pravděpodobnosti, tak ho tak i pojmenujeme Alternativní rozdělení $A\,(p)$.

■ **Řešený příklad 4.1** Na základě statistického měření v mlžné komoře bylo zjištěno, že částice α se vyskytne s pravděpodobností $p = 0,875$ a částice β s pravděpodobností $p = 0,125$. Výskyt kosmického a jiného záření je tak malý, že jeho pravděpodobnost je zanedbatelná. Pro sto sledovaných částic určete střední hodnotu, rozptyl, koeficient šikmosti, koeficient špičatosti a nejpravděpodobnější počet α a β částic. Určete pravděpodobnost, že mezi stovkou pozorovaných částic bude právě 82 z nich α
Řešení

Dostali jsme pokus u něhož jsou možné dva výsledky: Pozorujeme částici α nebo β. Binomické rozdělení $Bi\,(n = 100; p = 0,875)$. Pomocí vlastností binomického rozdělení určíme číselné charakteristiky výskytu obou částic:

$$\begin{aligned}
E\,(\alpha) &= n \cdot p = 100 \cdot 0,875 = 87,5 \\
E\,(\beta) &= n \cdot (1-p) = 100 \cdot 0,125 = 12,5 \\
D\,(\alpha) &= n \cdot p \cdot (1-p) = 100 \cdot 0,875 \cdot (1-0,875) = 10,9375 \\
D\,(\beta) &= n \cdot p \cdot (1-p) = 100 \cdot 0,125 \cdot (1-0,875) = 10,9375 \\
\gamma_1\,(\alpha) &= 0,4708 \\
\gamma_1\,(\beta) &= 0,9244 \\
\gamma_2\,(\alpha) &= 0,94 \\
\gamma_2\,(\beta) &= 0,94
\end{aligned}$$

Nejpravděpodobnější počet výskytu částic α a β musí splňovat:

$$\begin{aligned} (n+1)\cdot p - 1 \leq & \quad x_i \quad \leq (n+1)\cdot p \\ 101\cdot 0,875 - 1 \leq & \quad x_i \quad \leq 101\cdot 0,875 \\ 87,375 \leq & \quad x_i \quad \leq 88,375 \Rightarrow x_i = 88 \text{ alfa částic.} \\ 11,625 \leq & \quad x_i \quad \leq 12,625 \Rightarrow x_i = 12 \text{ beta částic.} \end{aligned}$$

Pravděpodobnost, že mezi stem částic jich bude právě 82 α určíme:

$$p(82) = \binom{100}{82}\cdot 0,875^{82}\cdot(1-0,875)^{100-82} = 0,0299$$

4.2 Alternativní rozdělení

Definice 4.2 (Alternativní rozdělení náhodné veličiny X) Náhodná veličina X s alternativním rozdělením $A(p)$ má pravděpodobnostní funkci:

$$p_1 = p \tag{4.10}$$

$$p_0 = 1 - p, \tag{4.11}$$

kde index 0 a 1 označuje neúspěch, respektive úspěch a číslo p pravděpodobnost úspěchu v daném pokusu.

Grafické znázornění distribuční $F(X)$ obrázek (4.1) a pravděpodobnostní $p(X)$ obrázek (4.2) funkce binomické náhodné veličiny X.

V předchozí kapitole jsme se naučili přiřazovat každé náhodné veličině její číselné charakteristiky. Ukažme si tedy nejčastěji užívané charakteristiky pro náhodnou veličinu X s alternativním rozdělením pravděpodobnosti.

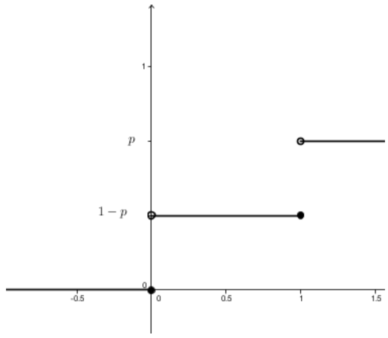
Vlastnosti 4.2

1. *Střední hodnota*

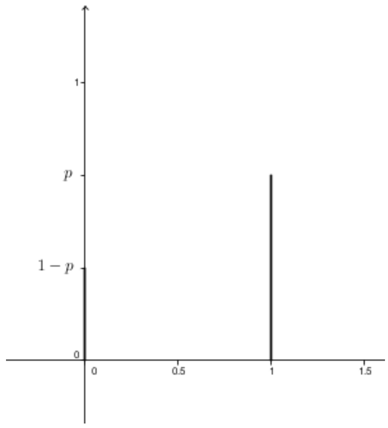
$$E(X) = p \tag{4.12}$$

2. *Rozptyl*

$$\sigma^2 = p\cdot(1-p) \tag{4.13}$$



Obrázek 4.1: Distribuční funkce binomického rozdělení



Obrázek 4.2: Pravděpodobnostní funkce binomického rozdělení

3. *Koeficient šikmosti*

$$\gamma_1 = \frac{1-2p}{\sqrt{p\cdot(1-p)}} \tag{4.14}$$

4. *Koeficient špičatosti*

$$\gamma_2 = \frac{1-6p\cdot q}{p\cdot(1-p)} \tag{4.15}$$

5. *Koeficient koncentrace*

$$\gamma_2 + 3 \tag{4.16}$$

4.3 Hypergeometrické rozdělení

Nyní mějme n-krát opakující se pokus se stejnými předpoklady jako u binomického rozdělení s jediným rozdílem a to, že pokusy se vzájemně ovlivňují. Tím myslíme, že výsledek následujícího pokusu závisí na provedení předcházejícího pokusu. Označme počet všech pokusů v řadě N, počet všech pokusů, které dávají výsledek V označíme M, počet všech uskutečněných pokusů označíme n a počet všech pokusů, které nedávají výsledek V označíme $N-M$. Potom můžeme hypergeometrické rozdělení definovat takto:

Definice 4.3 (Hypergeometrické rozdělení náhodné veličiny X) Hypergeometrické rozdělení $H\left(N;M;n\right)$ má pravděpodobnostní funkci:

$$P\left(X\right) = \frac{\binom{M}{x}\cdot\binom{N-M}{n-x}}{\binom{N}{n}}, \tag{4.17}$$

kde $x = 1;2;3,\ldots;n$.

Hypergeometrické rozdělení náhodné veličiny X je co se do výpočtů týká poměrně náročné, proto si v některých případech, kdy je jistá tolerance k přesnosti výsledku můžeme dovolit hypergeometrické rozdělení aproximovat binomickým rozdělením.

Podmínka pro aproximaci hypergeometrického rozdělení. Máme-li dostatečně velký soubor pokusů N a uskutečníme pouze malé množství pokusů n můžeme si dovolit tvrdit, že tyto uskutečněné pokusy příliš neovliňují ještě neprovedené pokusy a pak řekneme, že jednotlivé pokusy jsou nezávislé. V tomto případě provedeme aproximaci hypergeometrického rozdělení binomickým rozdělením náhodné veličiny X:

$$H\left(N;M;n\right) \approx Bi\left(n;\frac{M}{N}\right) \Rightarrow \frac{\binom{M}{x}\cdot\binom{N-M}{n-x}}{\binom{N}{n}} \approx \binom{n}{x}\cdot\left(\frac{M}{N}\right)^{x}\cdot\left(1-\frac{M}{N}\right)^{n-x} \tag{4.18}$$

Aproximace na binomické rozdělení funguje dobře, jestliže platí:

$$n \leq \frac{N}{10}$$

V předchozí kapitole jsme se naučili přiřazovat každé náhodné veličině její číselné charakteristiky. Ukažme si tedy nejčastěji užívané charakteristiky pro náhodnou veličinu X s hypergeometrickým rozdělením pravděpodobnosti.

Vlastnosti 4.3

1. *Střední hodnota*
$$E(X) = n \cdot \frac{M}{N} \tag{4.19}$$
2. *Rozptyl*
$$\sigma^2 = n \cdot \frac{M}{N} \cdot \left(1 - \frac{M}{N}\right) \cdot \frac{N-n}{N-1} \tag{4.20}$$
3. *Nejpravděpodobnější počet příznivých pokusů x_i je řešením rovnice:*
$$a - 1 \leq x_i \leq a, \tag{4.21}$$

kde $a = \frac{(M+1)\cdot(n+1)}{N+2}$.

■ **Řešený příklad 4.2** V dodávce 100 výrobků je 8 poškozených. Náhodně vybereme 5 z nich k dalšímu zpracování. Jaká je pravděpodobnost, že mezi vybranými výrobky bude nejvýše jeden poškozený? Určete charakteristiky rozdělení a nejpravděpodobnější počet vybraných poškozených výrobků.

Řešení:

Máme $N = 100$ výrobků, vybíráme $n = 5$ výrobků a pravděpodobnost, že vytáhneme jeden vadný výrobek je $\frac{M}{N} = \frac{8}{100}$. Jedná se o hypergeometrické rozdělení $H(N = 100; M = 8; n = 5)$.

Pravděpodobnost, že mezi taženými výrobky bude nejvýše jeden poškozený získáme tak, že určíme pravděpodobnost vytažení jednoho vadného výrobku a žádného vadného výrobku.

$$\begin{aligned} p(0) &= \frac{\binom{8}{0}\cdot\binom{100-8}{5-0}}{\binom{100}{5}} = 0,6532 \\ p(1) &= \frac{\binom{8}{1}\cdot\binom{100-8}{5-1}}{\binom{100}{5}} = 0,3 \\ p(x \leq 1) &= p(0) + p(1) = 0,9534 \end{aligned}$$

Charakteristiky hypergeometrického rozdělení:

$$\begin{aligned} E(X) &= n \cdot \frac{M}{N} = 5 \cdot \frac{8}{100} = 0,4 \\ \sigma^2 &= n \cdot \frac{M}{N} \cdot \left(1 - \frac{M}{N}\right) \cdot \frac{N-n}{N-1} = 0,3531 \end{aligned}$$

Nejpravděpodobnější počet vytažených vadných výrobků je řešením rovnice:

$$\begin{aligned} \frac{(M+1)\cdot(n+1)}{N+2} - 1 \leq & \quad x_i \quad \frac{(M+1)\cdot(n+1)}{N+2} \\ -0,47 \leq & \quad x_i \quad \leq 0,53 \Rightarrow x_i = 0 \end{aligned}$$

Toto hypergeometrické rozdělení můžeme nahradit binomickým rozdělením $Bi\left(n; \frac{M}{N}\right)$, protože platí podmínka pro aproximaci $n = 5 \leq \frac{N=100}{10}$. Výpočet se pak značně zjednoduší:

$$\begin{aligned} p(x \leq 1) &= \binom{5}{0} \cdot \left(\frac{8}{100}\right)^0 \cdot \left(1 - \frac{8}{100}\right)^{5-0} + \binom{5}{1} \cdot \left(\frac{8}{100}\right)^1 \cdot \left(1 - \frac{8}{100}\right)^{5-1} = \\ &= 0,9456 \end{aligned}$$

4.4 Poissonovo rozdělení

Binomické a hypergeometrické rozdělení náhodné veličiny X fungují velice dobře a celkem snadno se s nimi počítá až do chvíle, když dostaneme k analýze skutečně velký počet pokusů. V takovém případě je výpočet kombinačních čísel $\binom{n}{k}$ problematický i kapesním kalkulátorem a musíme buď přistoupit k výpočtům na pc nebo použijeme jiné rozdělení náhodné veličiny X. Bude-li soubor náhodné veličiny X skutečně velký a nezávislý, tím myslíme teoreticky nekonečný počet pokusů což znamená, že obor hodnot náhodné veličiny je $H = \{0; 1; 2; 3; \ldots; n\}$. V takovém případě se binomického a hypergeometrického rozdělení náhodné veličiny vzdáme a použijeme nové Poissonovo reozdělení náhodné veličiny X.

Shrňme si předchozí úvahu. Poissonovské rozdělení pravděpodobnosti má náhodná veličina X, která vyjdřuje počet výskytů málo pravděpodobných jevů v nějaké míře (časová, délková, plošná, objemová, ...). Tímto jsme získali podmínky pro Poissonovské rozdělení:

1. Musíme znát střední počet výskytu daného jevu na úseku jednotkového jevu. Označíme jej λ.

2. Jednotkový úsek musíme umět rozdělit na n dílčích částí o velikosti Δt, pro které bude dále platit:

- Pravděpodobnost, že na libovolné dílčí části úseku Δt nastane sledovaný jev více než jednou je zanedbatelná.
- Pravděpodobnost nastoupení daného jevu na libovolné dílčí části úseku je přímo úměrná délce tohoto úseku (čím větší úsek, tím větší pravděpodobnost nastoupení jevu a naopak).
- Výskyty jevů v různých dílčích úsecích jsou na sobě nezávislé.

Definice 4.4 (Poissonovo rozdělení náhodné veličiny X) Náhodná veličina má Poissonovo rozdělení pravděpodobnosti $Po\left(\lambda\right)$, kde λ je kladný reálný parametr má pravděpodobnostní funkci:

$$P\left(X\right)=\frac{\lambda^{x}}{x!}\cdot e^{-\lambda}, \tag{4.22}$$

kde $x=1;2;3;\ldots;n$.

Může se stát, že u některých náhodných veličin nebudeme schopni určit pravděpodobnost výskytu jevu v dílčí části úseku δt, ale dokážeme určit pravděpodobnost v úseku, jehož míru označíme t (čas, délka, povrch, objem, ...). V takovém případě bude dán střední počet výskytů parametrem $\lambda\cdot t$ a pro pravděpodobnostní funkci náhodné veličiny platí:

$$P\left(X\right)=\frac{\left(\lambda\cdot t\right)^{x}}{x!}\cdot e^{-\lambda\cdot t}. \tag{4.23}$$

V předchozí kapitole jsme se naučili přiřazovat každé náhodné veličině její číselné charakteristiky. Ukažme si tedy nejčastěji užívané charakteristiky pro náhodnou veličinu X s poissonovským rozdělením pravděpodobnosti.

Vlastnosti 4.4

1. *Střední hodnota*

$$E\left(X\right)=\lambda \tag{4.24}$$

2. *Rozptyl*

$$\sigma^{2}=\lambda \tag{4.25}$$

3. *Koeficient šikmosti*

$$\gamma_1 = \frac{1}{\sqrt{\lambda}} \tag{4.26}$$

4. *Koeficient špičatosti*

$$\gamma_2 = \frac{1}{\lambda} \tag{4.27}$$

5. *Nepravděpodobnější počet výskytu přiznivých jevů x_i je řešení rovnice:*

$$\lambda - 1 \leq x_i \leq \lambda \tag{4.28}$$

Poissonovo rozdělení můžeme užít například u náhodných veličin, které popisují:

- Počty poruch, nehod nebo přírodních katastrof (zemětřesení, erupce sopky, dopad meteorů, ...).
- Počet mezinárodních nebo mezikontinentálních hovorů, atd.

Nahrazuje-li poissonovské rozdělení binomické, které má velký počet pokusů n s malou pravděpodobností úspěchu p. Pak pro parametr λ poissonovského rozdělení platí vztah:

$$\lambda = n \cdot p.$$

Takové nahrazení dává téměř stejné výsledky už aspoň pro 30 pokusů, u nichž je pravděpodobnost úspěchu $p \leq 0,1$.

Nahradit hypergeometrické rozdělení poissonovským má smysl, jestliže platí:

$$M \leq \frac{N}{10} \text{ a } n \leq \frac{N}{10}.$$

■ **Řešený příklad 4.3** Z radioaktivního materiálu vyzařuje průměrně 30 částic α za jednu minutu. Určete pravděpodobnost, že během jedné sekundy bude vyzářena právě jedna částice, žádná částice, nejvýše dvě částice, více než dvě částice. Určete pravděpodobnost, že během 15 sekund bude vyzářeno 6 částic α. Navíc určete charakteristiky vyzařování částice α a nejpravděpodobnější počet výskytů částice α v čase 5 s.

Řešení:

Známe střední počet výskytu sledovaného jevu, který je:

$$\lambda = \frac{30}{60}$$

Pravděpodobnost, že částice α bude za jednu sekundu vyzářena více než jednou se zdá poměrně malá. Ověříme při řešení úkolů.

Podmínky pro označení jevu za poissonovský jsou splněny. Platí $Po\left(\lambda = \frac{30}{60}\right)$.
Určíme pravděpodobnost, že bude vyzářena právě jedna částice $x = 1$:

$$p\left(1\right) = \frac{\left(\frac{1}{2}\right)^1}{1!} \cdot e^{-\frac{1}{2}} = 0,3033$$

Analogicky určíme i pravděpodobnost, že za dobu jedné sekundy nebude vyzářena žádná částice $x = 0$:

$$p\left(0\right) = \frac{\left(\frac{1}{2}\right)^0}{0!} \cdot e^{-\frac{1}{2}}$$

Pravděpodobnost, že budou vyzářeny nejvýše dvě částice získáme jesltiže sečteme pravděpodobnosti vyzáření žádné, jedné a dvou částic α.

$$\begin{aligned} p\left(0\right) &= 0,3033 \\ p\left(1\right) &= 0,6065 \\ p\left(2\right) &= \frac{\left(\frac{1}{2}\right)^2}{2!} \cdot e^{-\frac{1}{2}} = 0,0758 \\ P\left(x \leq 2\right) &= 0,3033 + 0,6065 + 0,0758 = 0,9856 \end{aligned}$$

Pravděpodobnost, že budou vyzářeny více než dvě částice učíme pomocí opačného jevu:

$$\begin{aligned} p\left(x > 2\right) &= p\left(3\right) + p\left(4\right) + \ldots = 1 - \left(p\left(0\right) + p\left(1\right) + p\left(2\right)\right) = \\ &= 1 - 0,6065 - 0,3033 - 0,0758 = 0,0144 \end{aligned}$$

Pro časový úsek délky $t = 15$ s platí, že pravděpodobnostní funkce má tvar:

$$\begin{aligned} p\left(x\right) &= \frac{\left(\lambda \cdot t\right)^x}{x!} \cdot e^{-\lambda \cdot t} \\ p\left(6\right) &= \frac{\left(\frac{1}{2} \cdot 15\right)^6}{6!} \cdot e^{-\frac{1}{2} \cdot 15} = 0,1367 \end{aligned}$$

Charakteristicky veličiny α jsou:

$$\begin{aligned} E\left(\alpha\right) &= \lambda = \frac{1}{2} \\ \sigma^2 &= \lambda = \frac{1}{2} \\ \gamma_1 &= \frac{1}{\sqrt{\lambda}} = \frac{1}{\sqrt{\frac{1}{2}}} = \sqrt{2} \\ \gamma_2 &= \frac{1}{\lambda} = \frac{1}{\frac{1}{2}} = 2 \end{aligned}$$

Pro nejpravděpodobnější výskyt vyzáření částice α za dobu $t = 5$ s upravíme a určíme takto:

$$\begin{aligned} \lambda \cdot t - 1 \leq \quad & x_i \quad \leq \lambda \cdot t \\ \frac{1}{2} \cdot 5 - 1 \leq \quad & x_i \quad \leq \frac{1}{2} \cdot 5 \Rightarrow x_i = 2 \end{aligned}$$

Úlohy:

1. Házíme třikrát mincí. Náhodná veličina X je určena počtem padnutí líce. Určete pravděpodobnostní funkci a charakteristiky veličiny X.
2. V zásilce je 100 výrobků. 80 výrobků je první a 20 druhé jakosti. Vybereme tři výrobky po jednom a po každém výběru výrobek vrátíme. Určete pravděpodobnost, že všechny tři vybrané výrobky budou první jakosti.
3. Dlouhodobým pozorováním stavu vody v řece byla určena pravděpodobnost jarní povodně na $\frac{4}{12}$. Určete střední hodnotu a rozptyl počtu povodní v nejbližších 100 letech.
4. Pravděpodobnost, že bude vyroben vadný kondenzátor je 0,05. S jakou pravděpodobností budou mezi 80 vyrobenými kondenzátory 4 vadné?
5. Bylo statisticky zjištěno, že na tisíc metrů určité látky připadá průměrně pět kazů. Pro dodávku bylo připraveno 50 stometrových balíků. Jaký počet balíků bez vady můžeme v očekávat v dodávce?
6. Průměrná poruchovost výrobního stroje za 10000 hodin provozu je 10. Určete pravděpodobnost, že se stroj porouchá za 100 hodin práce.
7. Ve výrobě se provádí kontrola sta součástek připravených k expedici. Vybereme náhodně 20 výrobků. Dlouhodobým sledováním bylo zjištěno, že zmetkovitost jsou 3 %. Určete střední hodnotu a rozptyl počtu zmetků.
8. Za jasné letní noci můžeme v průměru každých 10 minut vidět padat hvězdu. Jaká je pravděpodobnost, že během 25 minut uvidíme 3 padající hvězdy?
9. Jaká je pravděpodobnost, že mezi 750 lidmi budou dva, kteří se narodili ve stejný den?
10. Korektura o rozsahu 500 stran obsahuje 500 tiskových chyb. Jaká je pravděpodobnost, že na jedné straně bude nejvýše jedna chyba?

Řešení:

1. $p\left(x\right)=\binom{3}{x}\cdot\left(\frac{1}{2}\right)^{x}\cdot\left(\frac{1}{2}\right)^{3-x}$, $E\left(X\right)=\frac{3}{2}$, $\sigma^2=\frac{3}{4}$, $\gamma_1=0$, $\gamma_2=-\frac{2}{3}$
2. $p\left(3\right)=0,512$
3. $E\left(X\right)=26,6$, $\sigma^2=19,5$
4. $0,195$
5. 30
6. $0,095$
7. $E\left(X\right)=0,6$, $\sigma^2=0,470$
8. $p\left(3\right)=0,2138$
9. $p\left(2\right)=0,2705$
10. $p\left(x\leq 1\right)=0,7358$

Kapitola 5

Rozdělení pravděpodobností spojité náhodné veličiny

Máme-li spojitou náhodnou veličinu zajímáme se, s jakou pravděpodobností se daná veličina realizuje v určeném konečném nebo nekonečném intervalu $\langle a; b\rangle$. Již víme, že náhodná veličina X má spojité rozdělení pokud existuje nezáporná funkce $f(x)$, kterou nazýváme hustota pravděpodobnosti náhodné veličiny X. V takovém případě je pravděpodobnost realizace náhodné veličiny v intevralu $(a; b)$ dána vztahem:

$$P(X \in (a; b)) = \int_a^b f(x)\,\mathrm{d}x. \tag{5.1}$$

Hustota pravděpodobnosti musí stále splňovat podmínku:

$$\int_{-\infty}^{\infty} f(x)\,\mathrm{d}x = 1. \tag{5.2}$$

Již víme, že spojité náhodné veličiny mohou mít stejně jako diskrétní náhodné veličiny různá rozdělení pravděpodobnosti a pak je můžeme roztřídit. Jednotlivá rozdělení se liší pouze různou volbou parametrů a jejich hodnot.

Jako první příklad rozložení spojité náhodné veličiny uvedeme rovnoměrné rozdělení.

5.1 Rovnoměrné rozdělení

Definice 5.1 (Rovnoměrné rozdělení $R(a;b)$) Náhodná veličina X má rovnoměrné rozdělení $R(a;b)$, pro jejíž hustotu pravděpodobnosti platí:

$$f(x) = \frac{1}{b-a}, \tag{5.3}$$

kde $x \in \langle a;b \rangle$.

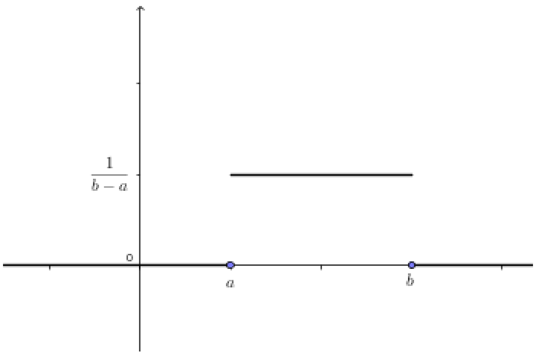
$$f(x) = 0, \tag{5.4}$$

kde $x \notin \langle a;b \rangle$. Pro oba případy platí $a; b \in \mathbb{R}$.

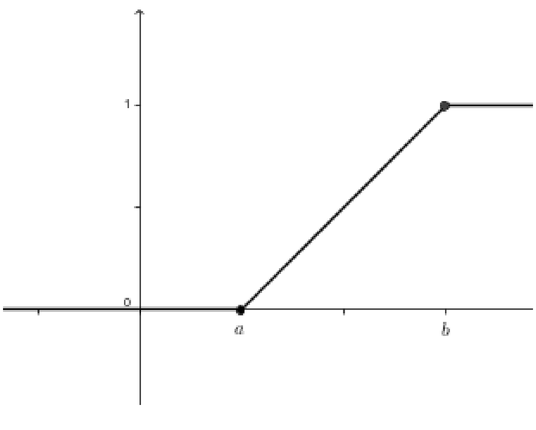
Pro náhodnou veličinu X s rovnoměrným rozdělením platí následující vlastnosti a charakteristiky.

Vlastnosti 5.1

1. *Pro distribuční funkci:*
   - $F(x) = 0$ *pro všechna* $x \in (-\infty; a\rangle$.
   - $F(x) = \frac{x-a}{b-a}$ *pro všechna* $x \in (a;b)$.
   - $F(x) = 1$ *pro všechna* $x \in \langle b; \infty)$.
2. *Střední hodnota*
$$E(X) = \frac{a+b}{2} \tag{5.5}$$
3. *Rozptyl*
$$\sigma^2 = \frac{(b-a)^2}{12} \tag{5.6}$$
4. *Koeficient šikmosti*
$$\gamma_1 = 0 \tag{5.7}$$
5. *Koeficient špičatosti*
$$\gamma_2 = -\frac{6}{5} \tag{5.8}$$
6. *Pravděpodobnost výskytu sledovaného jevu v intervalu* $(c;d) \subset \langle a;b \rangle$, *kde* $a;b;c;d \in \mathbb{R}$ *a zároveň* $a < c < d < b$.
$$P(c < X < d) = \frac{d-c}{b-a} \tag{5.9}$$



Obrázek 5.1: Hustota pravděpodobnosti rovnoměrného rozdělení



Obrázek 5.2: Distribuční funcke rovnoměrného rozdělení

Graf hustoty pravděpodobnosti obrázek (5.1) a graf distribuční funkce (5.2).

■ **Řešený příklad 5.1** Vlaky metra odjíždějí ze zastávky v desetiminutových intervalech. Cestující může na zastávku přijít v libovolném okamžiku. Určete hustotu pravděpodobnosti, distribuční funkci, číselné charakteristiky a pravděpodobnost, že cestující bude na vlak čekat déle než 7 minut a pravděpodobnost, že příjde v časovém intervalu dvě minuty až pět minut po odjezdu vlaku?

Řešení:

Náhodná veličina X čekání na příjezd vlaku je spojitá v intervalu $(0;10)$. Zároveň je pravděpodobnost příchodu cestujícího na zastávku všude stejná. Volíme rovnoměrné rozdělení $R\,(0;10)$.

Nejprve určíme hustotu pravděpodobnosti podle definice (5.1) a distribuční funkci určíme podle vlastností, které vyplívají vlastností hustoty pravděpodobnosti rovnoměrného rozdělení.

$$\begin{aligned}
f\,(x) &= \frac{1}{b-a} = \frac{1}{10-0} = \frac{1}{10} \qquad x \in \langle 0;10\rangle \\
f\,(x) &= 0 \qquad \text{jinde.} \\
F\,(x) &= \frac{x-a}{b-a} = \frac{x-0}{10-0} = \frac{x}{10} \qquad x \in \langle 0;10\rangle \\
F\,(x) &= 0 \qquad \text{jinde.}
\end{aligned}$$

Číselné charakteristiky určíme podle vlastností rovnoměrného rozdělení:

$$\begin{aligned}
E\,(X) &= \frac{a+b}{2} = \frac{0+10}{2} = 5 \\
\sigma^2 &= \frac{(b-a)^2}{12} = \frac{(10-0)^2}{12} = \frac{25}{3} \\
\gamma_1 &= 0 \\
\gamma_2 &= -\frac{6}{5}
\end{aligned}$$

Nyní určíme pravděpodobnost, že bude cestující čekat déle než sedm minut $P\,(x>7)$. Tu určíme pomocí pravděpodobnosti opačného jevu. Nejprve najdeme pravděpodobnost, že bude cestující čekat méně než sedm minut a pomocí této hodnoty určíme hledanou pravděpodobnost:

$$\begin{aligned} P(x<7) &= F(x<7) = \int\limits_0^7 f_x \mathrm{d}x = \frac{1}{10}\int\limits_0^7 \mathrm{d}x = \frac{1}{10}[x]_0^7 = \frac{7}{10} \\ P(x>7) &= 1-P(x<7) = 1-\frac{7}{10} = \frac{3}{10} \end{aligned}$$

Ještě určíme pravděpodobnost, že cestující příjde na zastávku mezi druhou a pátou minutou po odjezdu vlaku.

$$P(2<x<5) = \frac{d-c}{b-a} = \frac{5-2}{10-0} = \frac{3}{10}$$

5.2 Normální - Gaussovo rozdělení $N\left(\mu;\sigma^2\right)$

Normální rozdělení je jedno z nejdůležitějších rozdělení spojité náhodné veličiny X. Nejedná se o rozdělení, které by vyhovovalo většině spojitých veličin, ale za určitých podmínek můžeme normálním rozdělením aproximovat některá jiná rozdělení a to jak spojitých tak i diskrétních náhodných veličin. Především se používá k aproximaci binomického, poissonova a studentova rozdělení pravděpodobnosti. Ta se používají u náhodné veličiny, která je výsledkem velkého počtu na sobě nezávislých pokusů.

Definice 5.2 (Normální rozdělní pravděpodobnosti $N(\mu;\sigma^2)$) Náhodná veličina X má normální rozdělení pravděpodobnosti $N(\mu;\sigma^2)$ s hustotou pravděpodobnosti:

$$f(x) = \frac{1}{\sigma\cdot\sqrt{2\pi}}\cdot e^{-\frac{(x-\mu)^2}{2\cdot\sigma^2}}, kde$$

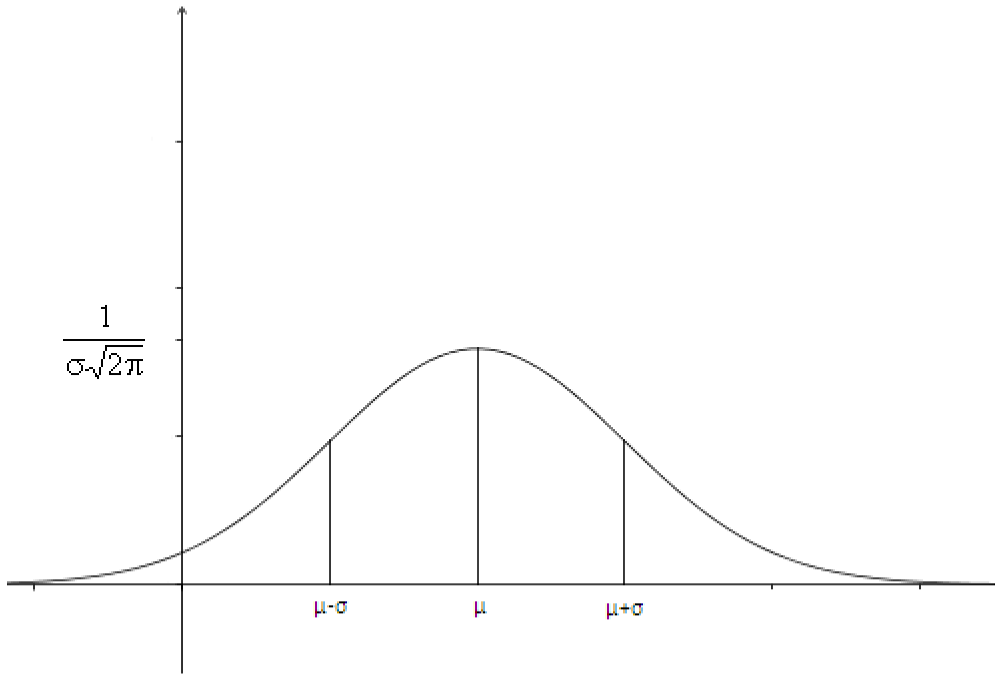
x$\in\mathbb{R}$, parametr μ je střední hodnota náhodné veličiny parametr σ^2 je rozptyl náhodné veličiny. (5.10)

Pro náhodnou veličinu X s rovnoměrným rozdělením platí následující vlasntosti a charakteristiky.

Vlastnosti 5.2

1. *Střední hodnota normálního rozdělení veličiny je shodná s hodnotou mediánu.*

$$E(X) = x_{0,5} \tag{5.11}$$



Obrázek 5.3: Hustota pravděpodobnosti - Gaussova křivka

2. *Koeficient šikmosti*

$$\gamma_1 = 0 \tag{5.12}$$

3. *Koeficient špičatosti*

$$\gamma_2 = 0 \tag{5.13}$$

4. *Distribuční funkce náhodné veličiny* X

$$F(x) = \int_{-\infty}^{\infty} \frac{1}{\sigma \cdot \sqrt{2\pi}} \cdot e^{-\frac{(x-\mu)^2}{2\cdot\sigma^2}} \mathrm{d}x \tag{5.14}$$

Graf hustoty pravděpodobnosti názýváme Gaussova křivka, její maximum najdeme v bodě $x = E(X)$. Inflexní body Gaussovy křivky jsou v bodech $x = \mu \pm \sigma$. Gaussova křivka je v intervalu $x \in (\mu - \sigma; \mu + \sigma)$ konkávní a ve zbylých bodech $x \notin (\mu - \sigma; \mu + \sigma)$ konvexní. Hodnoty v nevlastních bodech Gaussovy křivky jsou nulové.

Pro lepší představu o rozptýlení náhodné veličiny X okolo střední hodnoty si ukážeme její některé pravděpodobnosti:

$$P(X \in (\mu - \sigma; \mu + \sigma)) = 0,689$$

$$\begin{aligned} P\left(X \in (\mu - 2\sigma; \mu + 2\sigma)\right) &= 0,954 \\ P\left(X \in (\mu - 3\sigma; \mu + 3\sigma)\right) &= 0,997 \end{aligned}$$

Přesto, že se náhodná veličina X s normálním rozdělením realizuje pro všechna $x \in R$, v praxi většinou její realizaci zkoumáme pouze v intevalu $(\mu - 3\sigma; \mu + 3\sigma)$, kde je pravděpodobnost výskuty příznivého výsledku pokusu téměř rovna 1 (nastane skoro vždy).

I když má náhovná veličina X s normální rozdělením velké množství výhod, tak problémem může být výpočet jednotlivých hodnot pravděpodobnostní funkce. Nicméně je možné ji spočítat. Mnohem větším problémem bude výpočet konkrétních hodnot distribuční funkce, kterou spočteme integrací pravděpodobnostní funkce. Bohužel tato funkce není analyticky integrovatelná, musíme integrovat numericky. Numerická integrace sama o sobě není příliš komplikovaná, nicméně by zabrala další kapitolu. Zájemci o numerickou intergraci se doporučuje buď kurz numerických metod. Kvůli tomuto problému s integrací zavádíme další rozdělení pravděpodobnosti: Základní normované normální rozdělení pravděpodobnosti. Jedná se o speciální případ normálního rozdělení pravděpodobnosti, kde střední hodnota nabývá hodnoty 0 a rozptyl má hodnotu 1. Jednoduše řečeno náhodnou veličinu vhodnou transformací vycentrujeme na střed soustavy souřadnic.

5.2.1 Normované normální rozdělení $N(0;1)$

Definice 5.3 (Základní normované normální rozdělení $N(0;1)$) Náhodná veličina má základní normované normální rozdělení $N(0;1)$, jestliže její střední hodnota je $\mu = 0$ a rozptyl $\sigma^2 = 1$, pak hustota pravděpodobnosti je dána vztahem:

$$\varphi(x) = \frac{1}{\sqrt{2 \cdot \pi}} \cdot e^{-\frac{x^2}{2}}, \tag{5.15}$$

kde $x \in \mathbb{R}$.

Pro náhodnou veličinu X se základním normovaným normálním rozdělením platí následující vlastnosti a charakteristiky:

Vlastnosti 5.3

1. *Distribuční funkce normovaného normálního rozdělení* $N(0;1)$ *značíme* $\Phi(x)$ *a je dána vztahem:*

$$\Phi(x) = \frac{1}{\sqrt{2\cdot\pi}} \cdot \int\limits_{-\infty}^{x} e^{\frac{t^2}{2}}\,\mathrm{d}t. \tag{5.16}$$

Poznámka: *Ani distribuční funkce normovaného normálního rozdělení* $N(0;1)$ *není analyticky integrovatelná, nicméně její numerická integrace je značně jednodušší. Některé hodnoty numerické integrace této distribuční funkce uvádí tabulka na straně (116).*

2. *Záporné hodnoty distribuční funkce* $\Phi(x)$ *v tabulce na staně (116) nejsou uvedeny, protože pro ně platí vztah:*

$$\Phi(-x) = 1 - \Phi(x) \tag{5.17}$$

Poznámka: *Tato vlastnost přímo vyplývá ze symetrie Gaussovy křivky.*

3. *Má-li náhodná veličina* X *normální rozdělení* $N(\mu;\sigma^2)$ *s hustotou pravděpodobnosti:*

$$f(x) = \frac{1}{\sigma\cdot\sqrt{2\cdot\pi}} \cdot e^{-\frac{(x-\mu)^2}{2\cdot\sigma^2}},$$

pak transformace náhodné veličiny X *do náhodné veličiny s normovaným normálním rozdělením* $N(0;1)$ *vypadá takto:*

$$Y = \frac{X-\mu}{\sigma}. \tag{5.18}$$

Hustota pravděpodobnosti náhodné veličiny Y *je:*

$$\varphi(y) = \frac{1}{\sqrt{2\cdot\pi}} \cdot e^{\frac{y^2}{2}} \tag{5.19}$$

a distribuční funkci náhodné veličiny X *určíme takto:*

$$F(x) = \Phi\left(\frac{x-\mu}{\sigma}\right) \tag{5.20}$$

■ **Řešený příklad 5.2** Náhodná veličina X má normální rozdělení $N(-5;16)$. Určete hustotu pravděpodobnosti, distribuční funkci a pravděpodobnost $P(-1 < x < 2)$ a $P(-6 < x < 5)$.

Řešení:

Hustotu pravděpodobnosti a distribuční funkci určíme dosazením rozptylu a střední hodnoty do definice (5.3) a rovnice (5.16).

$$\begin{aligned}
\varphi(x) &= \frac{1}{\sigma \cdot \sqrt{2 \cdot \pi}} \cdot e^{-\frac{(x-\mu)^2}{2 \cdot \sigma^2}} \\
\varphi(x) &= \frac{1}{4 \cdot \sqrt{2 \cdot \pi}} \cdot e^{-\frac{(x+5)^2}{32}} \\
\Phi(x) &= \int\limits_{-\infty}^{x} \frac{1}{\sigma \cdot \sqrt{2 \cdot \pi}} \cdot e^{-\frac{(t-\mu)^2}{2 \cdot \sigma^2}} \mathrm{d}t \\
\Phi(x) &= \frac{1}{4 \cdot \sqrt{2 \cdot \pi}} \int\limits_{-\infty}^{x} e^{-\frac{(t+5)^2}{32}} \mathrm{d}t
\end{aligned}$$

Abychom mohli určit pravděpodobnosti $P(-1 < x < 2)$ a $P(-6 < x < 5)$ budeme veličinu X transformovat na veličinu T, pro kterou platí:

$$T = \frac{X - \mu}{\sigma}$$

Touto transformací převedeme náhodnou veličinu X s rozdělením pravděpodobnosti $N(\mu; \sigma^2)$ na náhodnou veličinu T s rozdělením pravděpodobnosti $N(0; 1)$. Tuto transformaci provdeme kvůli tomu, že se s normálním normovaným rozdělením lépe počítá. K tomu abychom mohli dopočítat hledané pravděpodobnosti je potřeba transformovat i meze hledaných pravděpodobností.

$$\begin{aligned}
t_1 &= \frac{x_1 - \mu}{\sigma} = \frac{-1+5}{4} = 1 \\
t_2 &= \frac{x_2 - \mu}{\sigma} = \frac{2+5}{4} = 1,75 \\
t_3 &= \frac{x_3 - \mu}{\sigma} = \frac{-6+5}{4} = -0,25 \\
t_4 &= \frac{x_4 - \mu}{\sigma} = \frac{5+5}{4} = 2,5
\end{aligned}$$

Nyní můžeme konečně určit hledané pravděpodobnosti, ke kterým potřebujeme nalézt hodnoty distribuční funkce normálního normovaného rozdělení $N(0; 1)$. Tyto hodnoty nalezneme v tabulce hodnot distribuční funkce na staně (116).

$$P(-1 < x < 2) \quad \rightarrow \quad P(1 < t < 1,75) = \Phi(1,75) - \Phi(1) = 0,9599 - 0,8413 = 0,1186$$

$$\begin{aligned} P(-6;5) \quad &\to \quad P(-0,25 < t < 2,5) = \Phi(2,5) - \Phi(-0,25) = \\ &= \quad \Phi(2,5) - (1 - \Phi(0,25)) = 0,9938 - 1 + 0,5987 = 0,5925 \end{aligned}$$

5.3 Exponenciální rozdělení

Exponenciální rozdělení je úzce spojeno s Poissonovým rozdělením. Připomeňme si, že Poissonovo rozdělení určuje počet událostí, které nastanou v nějakém časovém intervalu (i jiná míra). Exponenciálním reozdělením se řídí večina, která je dána dobou nebo nějakou jinou mírou čekání na výskyt určité události.

Definice 5.4 (Exponenciální rozdělení $E(\lambda)$) Náhodná veličina X má exponenciální rozdělení $E(\lambda)$ s hustotou pravděpodobnosti:

$$f(x) = \lambda \cdot e^{-\lambda \cdot x}, \tag{5.21}$$

kde $x \in \mathbb{R}^+$.

$$f(x) = 0, \tag{5.22}$$

kde $x \in \mathbb{R}_0^-$ a parametr λ je kladná reálná konstanta určená středním počtem poissonovských náhodných jevů v určitém jednotkovém intervalu dané míry.

Pro náhodnou veličinu s exponenciálním rozdělením $E(\lambda)$ platí následující vlastnosti a charakteristiky:

Vlastnosti 5.4

1. *Distribuční funkce*
$$F(x) = 1 - e^{-\lambda \cdot x} \tag{5.23}$$
2. *Střední hodnota*
$$E(X) = \frac{1}{\lambda} \tag{5.24}$$
3. *Rozptyl*
$$\sigma^2 = \frac{1}{\lambda^2} \tag{5.25}$$

■ **Řešený příklad 5.3** Doba čekání na padající hvězdu je náhodná veličina X s exponenciálním rozdělením $E(\lambda)$. Sledováním bylo zjištěno, že průměrně uvidíme padat hvězdu čtyřikrát do hodiny. Určete hustostu pravděpodobnosti, distribuční funkci, střední hodnotu čekání a rozptyl. Určete dobu x čekání na padající hvězdu tak, aby její pravděpodobnost byla 0,8.

Řešení:

Průměrná doba čekání na hvězdu v jedné hodině je $\lambda = 4$. Hustotu pravděpodobnosti a distribuční funkci náhodné veličiny X s exponenciálním rozdělením pravděpodobnosti určíme podle definice (5.4).

$$\begin{aligned} f(x) &= \lambda \cdot e^{-\lambda \cdot x} = 4 \cdot e^{-4x} \\ F(x) &= 1 - e^{-\lambda \cdot x} = 1 - e^{-4x} \end{aligned}$$

Pro číselné charakteristiky platí:

$$\begin{aligned} E(X) &= \frac{1}{\lambda} = \frac{1}{4} \\ \sigma^2 &= \frac{2}{\lambda^2} - \left(\frac{1}{\lambda}\right)^2 = \frac{1}{16} \end{aligned}$$

Nakonec určíme hodnotu a pro kterou platí, že náhodná veličina pro ní má pravděpodobnost $P(0 < x < a) = 0,8$.

$$\begin{aligned} 0,8 &= P(0 < x < a) = F(x < a) = \int_0^a f(x)\,\mathrm{d}x = \int_0^a 4 \cdot e^{-4x}\mathrm{d}x = 4 \cdot \left[\frac{e^{-4x}}{-4}\right]_0^a \\ 0,8 &= -e^{-4a} + e^0 \\ e^{-4a} &= 0,2 \\ \ln e^{-4a} &= \ln 0,2 \\ -4a &= \ln 0,2 \\ a &= 0,4023 \end{aligned}$$

Doba a čekání na pozorování hvězdy s pravděpodobností 0,8 je 24 minut a 8,5 sekund.

5.4 Další rozdělení spojitých náhodných veličin

5.4.1 Weibullovo rozdělení $W(\delta; c)$

Toto rozdělení spojité náhodné veličiny se používá pro rozdělení veličiny, kterému nevyhovuje exponenciální rozdělení pravděpodobnosti. Parametr δ je materiálová konstanta a parametr $c > 0$.

Pro hustotu pravděpodobnosti platí:

$$\begin{aligned} f(x) &= \frac{c \cdot x^{c-1}}{\delta^c} \cdot e^{-\left(\frac{x}{\delta}\right)^c} \qquad x > 0 && (5.26) \\ &= 0 \qquad x \leq 0. && (5.27) \end{aligned}$$

Pro distribuční funkci platí:

$$\begin{aligned} F(x) &= 1 - e^{-\left(\frac{x}{\delta}\right)^c} \qquad x > 0 && (5.28) \\ &= 0 \qquad x \leq 0. && (5.29) \end{aligned}$$

5.4.2 Pearsonovo rozdělení χ_n^2

Pearsonovo rozdělení χ_n^2 čteme chí kvadrát s n stupni volnosti. Toto rozdělení pravděpodobnosti užijeme, jestliže $X_1; X_2; \ldots; X_n$ náhodných veličin má normální normované rozdělení $N(0; 1)$, pak veličina $X = X_1^2 + X_2^2 + \ldots + X_n^2$ má Pearsonovo rozdělení.

Pro hustotu pravděpodobnosti platí:

$$\begin{aligned} f(x) &= \frac{x^{\frac{n}{2}-1} \cdot e^{-\frac{x}{2}}}{2^{\frac{n}{2}} \cdot \Gamma\left(\frac{n}{2}\right)} \qquad x > 0 && (5.30) \\ &= 0 \qquad x \leq 0 && (5.31) \end{aligned}$$

Kde $\Gamma(x)$ je gama funkce:

$$\Gamma(x) = \int_0^\infty e^{-t} \cdot t^{x-1} \mathrm{d}t$$

5.4.3 Studentovo rozdělení t_n

Náhodná veličina X má studentovo rozdělení s n stupni volnosti, jestliže platí:

$$X = \frac{X_1}{X_2} \cdot \sqrt{n}.$$

Kde náhodná veličina X_1 má normované normální rozdělení $N(0;1)$ a veličina X_2 má Pearsonovo rozdělení χ^2_n.

Pro hustotu pravděpodobnosti platí:

$$f(x) = \frac{1}{\sqrt{n \cdot \pi}} \cdot \frac{\Gamma\left(\frac{n+1}{2}\right)}{\Gamma\left(\frac{n}{2}\right)} \cdot \left(1 + \frac{x^2}{n}\right)^{\frac{n+1}{2}} \qquad x \in \mathbb{R} \quad n \in \mathbb{N}. \tag{5.32}$$

Kde $\Gamma(x)$ je gama funkce z předchozího rozdělení.

5.4.4 Fischerovo-Snedecorovo rozdělení

Náhodná veličina X má Fischerovo-Snedecorovo rozdělení s m a n stupni volnosti, jestliže platí:

$$X = \frac{\frac{X_1}{m}}{\frac{X_2}{n}}.$$

Kde náhodná veličina X_1 má Pearsonovo rozdělení χ^2_m a veličina X_2 má Pearsonovo rozdělení χ^2_n.

Pro hustotu pravděpodobnosti platí:

$$\begin{aligned} f(x) &= \frac{\Gamma\left(\frac{m+n}{2}\right)}{\Gamma\left(\frac{m}{2}\right) \cdot \Gamma\left(\frac{n}{2}\right)} \cdot \left(\frac{m}{n}\right)^{\frac{m}{2}} \cdot x^{\frac{m}{2}-1} \cdot \left(1 + \frac{mx}{n}\right)^{-\frac{m+n}{2}} \qquad x > 0 && (5.33) \\ &= 0 \qquad x \leq 0 && (5.34) \end{aligned}$$

Úlohy:

1. Jaká je pravděpodobnost, že náhodná veličina X, která má rozdělení $N\left(10;9\right)$, nabude hodnoty:
   a. menší než 16,
   b. větší než 10.
2. Náhodná veličina má rozdělení pravděpodobnosti:
   a. $N\left(0;1\right)$ určete $P\left(|X|<0,7\right)$
   b. $N\left(0;4\right)$ určete $P\left(x<-0,5\right)$
   c. $N\left(1;4\right)$ určete $P\left(x<-0,5\right)$
3. Doba čekání zákazníka na obsluhu je náhodná veličina s exponenciálním rozdělením $E\left(\lambda=0,5\right)$. Určete hustotu pravděpodobnosti a distribuční funkci náhodné veličiny X. Určete průměrnou dobu čekání na obsluhu a dobu čekání, během níž bude zákazník obsloužen s pravděpodobností 0,9.
4. Pokud necháme studenta samotného v laboratoři fyziky, dovede poškodit průměrně 4 fyzikální pomůcky za hodinu. Jak dlouho může učitel fyziky nechat studenta bez dozoru, jestliže chceme aby student nic nepoškodil s pravděpodobností 0,9.
5. Počítač má průměrně poruchu jednou za 2000 hodin. Předpokládejme, že doba čekání na poruchu je náhodná veličina s exponenciálním rozdělením. Určete dobu x tak, aby pravděpodobnost, že počítač bude fungovat bez poruchy delší dobu než x byla 0,99
6. Náhodná veličina má hustotu pravděpodobnosti:
$$\begin{aligned} f\left(x\right) &= \frac{1}{10}\cdot e^{-\frac{x}{10}} \qquad x>0 \\ &= 0 \qquad x\leq 0 \end{aligned}$$
   Určete její střední hodnotu a rozptyl.
7. Měření je zatíženo chybou -0,3 cm. Náhodné chyby měření mají normální rozdělení pravděpodobnosti se směrodatnou odchylkou $\sigma=0,5$ cm. Jaká je pravděpodobnost, že chyba měření nepřekročí v absolutní hodnotě trojnásobek směrodatné odchylky?

Řešení:

1. a. $0,9773$, b. $0,5$
2. a. $0,8413$, b. $0,1587$, c. $0,2913$
3. $f(x) = \frac{1}{2} \cdot e^{-\frac{x}{2}}$ kde $x \geq 0$ a $f(x) = 0$ kde $x < 0$, $F(x) = 1 - e^{-\frac{x}{2}}$ kde $x \geq 0$ a $F(x) = 0$ kde $x < 0$, $E(X) = 2$, $\sigma^2 = 4$ $x_1 = 4,6$
4. $x = 0,0263$ h $= 1,58$ min.
5. $x = 20,5$
6. $E(X) = 10$, $\sigma^2 = 100$
7. $0,9916$

x	0,00	0,01	0,02	0,03	0,04	0,05	0,06	0,07	0,08	0,09
0,0	0,5000	0,5039	0,5079	0,5119	0,5159	0,5199	0,5239	0,5279	0,5318	0,5358
0,1	0,5398	0,5438	0,5478	0,5517	0,5557	0,5596	0,5636	0,5675	0,5714	0,5754
0,2	0,5793	0,5832	0,5871	0,5910	0,5948	0,5987	0,6026	0,6064	0,6103	0,6141
0,3	0,6179	0,6217	0,6255	0,6293	0,6331	0,6368	0,6406	0,6443	0,6480	0,6517
0,4	0,6554	0,6591	0,6628	0,6664	0,6700	0,6736	0,6772	0,6808	0,6844	0,6879
0,5	0,6915	0,6950	0,6985	0,7019	0,7054	0,7088	0,7123	0,7157	0,7190	0,7224
0,6	0,7258	0,7291	0,7324	0,7357	0,7389	0,7422	0,7454	0,7486	0,7518	0,7549
0,7	0,7580	0,7612	0,7642	0,7673	0,7704	0,7734	0,7764	0,7794	0,7823	0,7852
0,8	0,7881	0,7910	0,7939	0,7967	0,8000	0,8023	0,8051	0,8079	0,8106	0,8133
0,9	0,8159	0,8186	0,8212	0,8238	0,8264	0,8289	0,8315	0,8340	0,8365	0,8389
1,0	0,8413	0,8438	0,8461	0,8485	0,8508	0,8531	0,8554	0,8577	0,8599	0,8621
1,1	0,8643	0,8665	0,8686	0,8708	0,8729	0,8749	0,8770	0,8790	0,8810	0,8830
1,2	0,8849	0,8869	0,8888	0,8907	0,9825	0,8944	0,8962	0,8980	0,8997	0,9015
1,3	0,9032	0,9049	0,9066	0,9082	0,9099	0,9115	0,9131	0,9147	0,9162	0,9177
1,4	0,9192	0,9207	0,9222	0,9236	0,9251	0,9265	0,9279	0,9292	0,9306	0,9319
1,5	0,9332	0,9345	0,9357	0,9370	0,9382	0,9394	0,9406	0,9418	0,9430	0,9441
1,6	0,9452	0,9463	0,9474	0,9485	0,9495	0,9505	0,9515	0,9525	0,9535	0,9545
1,7	0,9554	0,9564	0,9573	0,9581	0,9591	0,9599	0,9608	0,9616	0,9625	0,9633
1,8	0,9641	0,9649	0,9656	0,9664	0,9671	0,9678	0,9686	0,9693	0,9700	0,9706
1,9	0,9713	0,9719	0,9726	0,9732	0,9738	0,9744	0,9750	0,9756	0,9762	0,9767
2,0	0,9773	0,9778	0,9783	0,9788	0,9793	0,9798	0,9803	0,9808	0,9812	0,9817
2,1	0,9821	0,9826	0,9830	0,9834	0,9838	0,9842	0,9846	0,9850	0,9854	0,9857
2,2	0,9861	0,9865	0,9868	0,9871	0,9875	0,9878	0,9881	0,9884	0,9887	0,9890
2,3	0,9893	0,9896	0,9898	0,9901	0,9904	0,9906	0,9909	0,9911	0,9913	0,9916
2,4	0,9918	0,9920	0,9922	0,9925	0,9927	0,9929	0,9931	0,9932	0,9934	0,9936
2,5	0,9938	0,9940	0,9941	0,9943	0,9945	0,9946	0,9948	0,9949	0,9951	0,9952
2,6	0,9953	0,9955	0,9956	0,9957	0,9959	0,9960	0,9961	0,9962	0,9963	0,9964
2,7	0,9965	0,9966	0,9967	0,9968	0,9969	0,9970	0,9971	0,9972	0,9973	0,9974
2,8	0,9975	0,9975	0,9976	0,9977	0,9977	0,9978	0,9979	0,9980	0,9980	0,9981
2,9	0,9981	0,9982	0,9983	0,9983	0,9984	0,9984	0,9985	0,9985	0,9986	0,9986
3,0	0,9987	0,9987	0,9987	0,9988	0,9988	0,9989	0,9989	0,9989	0,9990	0,9990
3,1	0,9990	0,9991	0,9991	0,9991	0,9992	0,9992	0,9992	0,9992	0,9993	0,9993
3,2	0,9993	0,9993	0,9994	0,9994	0,9994	0,9994	0,9994	0,9995	0,9995	0,9995
3,3	0,9995	0,9995	0,9996	0,9996	0,9996	0,9996	0,9996	0,9996	0,9996	0,9997
3,4	0,9997	0,9997	0,9997	0,9997	0,9997	0,9997	0,9997	0,9997	0,9998	0,9998

Tabulka 5.1: $\Phi\left(x\right) = \frac{1}{\sqrt{2\pi}} \int\limits_{-\infty}^{x} e^{-\frac{u^2}{2}} \mathrm{d}x$

Kapitola 6

Náhodný vektor a jeho charakteristiky

6.1 Náhodný vektor

Zatím jsme zkoumali pouze náhodné pokusy u nichž byla výsledkem právě jedna veličina. Vezměme si nyní náhodný pokus, při němž měříme každý den ve stejnou dobu teplotu a atmosférický tlak. Výsledkem měření není pouze jedna, ale dvě náhodné veličiny. Každou z veličin jsme schopni velice dobře vyhodnotit zvlášť, ale budou-li nás zajímat společné vlastnosti, závislosti a charakteristiky těchto dvou veličin, budeme mít problém. Z tohoto důvodu zavedeme **Náhodný vektor** dvou náhodných veličin, který zadefinujeme právě tak, aby bylo možné zkoumat i vztah mezi naměřenými veličinami. Náhodný vektor dále zavedeme tak, aby bylo možné společně vyhodnotit libovolný počet náhoných veličin. Nicméně vzhledem k horším možnostem představy o n-rozměrném náhodném vektoru budeme všechny vlastnosti a charakteristiky ukazovat právě na dvourozměrném náhoném vektoru, jehož charakteristiky a vlastnosti můžeme lehce ukázat i graficky. Na druhou stranu je možné tyto vlastnosti zpětně zobecnit na každý n-rozměrný náhodný vektor.

Definice 6.1 (Náhodný vektor $\mathbf{X} = (X; Y)$) Pojmem n-rozměrný náhodný vektor $\mathbf{X}$ rozumíme vektor $\mathbf{X} = (X_1; X_2; \ldots; X_n)$, kde jeho složky $X_1; X_2; \ldots; X_n$ jsou náhodné veličiny. Speciálně pak dvourozměrný náhodný vektor $\mathbf{X} = (X; Y)$.

Náhodný vektor můžeme vytvořit mnoha způsoby a tím pádem můžeme získat náhodný vektor diskrétních a spojitých náhodných veličin, které budeme popisovat. Ovšem získat můžeme i náhodný vektor jehož všechny složky nebudou pouze diskrétní nebo spojité. Takovým náhodným vektorům se v dalším studiu vyhneme a tím si ho značně zjednodušíme.

Pro správnou představu o tom jak náhodný vektor vypadá, zavedeme stejně jako u náhodné veličiny popisné funkce a to distribuční a frekvenční funkce náhodného vektoru. V analogii s náhodnou veličinou bude také platit, že definice frekvenčních funkcí se budou lišit u diskrétního a spojitého náhodného vektoru.

Definice 6.2 (Distribuční funkce náhodného vektoru $F(X;Y)$) Reálnou funkci dvou prouměnných:

$$F(x;y) = P(X < x; Y < y), \tag{6.1}$$

nazveme distribuční (simultánní, sdružená) funkce náhodného vektoru $\mathbf{X} = (X;Y)$. Definiční obor distribuční funkce diskrétního náhodného vektoru je spočetná množina uspořádaných dvojic $[x;y]$. Definiční obor distribuční funkce spojitého náhodného vektoru je nespočetná množina uspořádaných dvojic $[x;y]$.

Poznámka: Distribuční funkce $F(x;y)$ přiřazuje náhodnému vektoru $\mathbf{X} = (X;Y)$ pravděpodobnost, že náhodná veličina X nabude hodnoty menší než x a zároveň náhodná veličina Y nabude hodnoty menší než y.

Stejným způsobem jako u náhodné veličiny přiřadíme distribuční funkci náhodného vektoru základní vlastnosti:

Vlastnosti 6.1

1. *Obor hodnot distribuční funkce $F(x;y)$, pro všechny dvojice $[x;y]$ platí:*

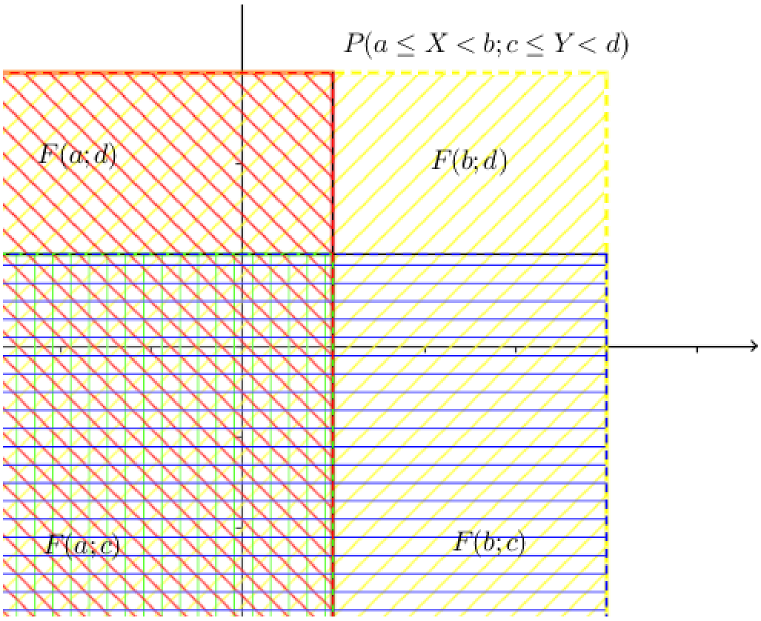
$$F(x;y) \in \langle 0;1 \rangle \tag{6.2}$$

2. *Monotonie distribuční funkce - distribuční funkce je neklesající funkce každé své proměnné.*
3. *Spojitost - distribuční funkce je v každém bodě zleva spojitá.*
4. *Významné hodnoty distribuční funkce:*

$$\begin{aligned} F(-\infty;-\infty) &= 0 \\ F(-\infty;y) &= 0 \\ F(x;\infty) &= 0 \\ F(\infty;\infty) &= 1 \end{aligned}$$

5. *Pravděpodobnost výskytu náhodného vektoru v podmnožině jeho definičního oboru:*

$$P\left(a \leq X < b; c \leq Y < d\right) = F\left(b; d\right) - F\left(a; d\right) - F\left(b; c\right) + F\left(a; c\right) \tag{6.3}$$



Obrázek 6.1: $P\left(a \leq X < b; c \leq Y < d\right)$

Marginální distribuční funkce

Setkáváme se také s názvem okrajová distribuční funkce, která vyjadřuje rozložení pravděpodobnosti jedné ze složek náhodného vektoru.

Definice 6.3 (Marginální distribuční funkce složek náhodného vektoru $\mathbf{X} = (X;Y)$) Marginální distribuční funkce $F_X\left(x\right)$ veličiny X nebo Y náhodného vektoru $\mathbf{X} = (X;Y)$ je dána vztahem:

$$F_X\left(x\right) = P\left(X < x\right) = \lim_{y\to\infty} F\left(x;y\right) = F\left(x;\infty\right) \tag{6.4}$$

$$F_Y\left(y\right) = P\left(Y < y\right) = \lim_{x\to\infty} F\left(x;y\right) = F\left(\infty;y\right). \tag{6.5}$$

Marginální distribuční funkce vyjadřují pravděpodobnost nabytí hodnoty menší než x nebo y složky X nebo Y náhodného vektoru bez omezení druhou veličinou.

Pro další popis náhodného vektoru musíme u náhodného vektoru rozlišit podle toho, jestli jsou jeho složky spojité nebo diskrétní.

Náhodný vektor s diskrétními složkami

Frekvenční funkci náhodného vektoru s diskrétními složkami nazýváme, v analogii s náhodnou veličinou, pravděpodobnostní funkce.

Definice 6.4 (Pravděpodobnostní (simultánní) funkce náhodného vektoru)
Pravděpodobnostní funkce náhodného vektoru $\mathbf{X} = (X; Y)$ s dikrétními složkami určuje pravděpodobnost, že náhodná veličina X nabude hodnotu x a zároveň náhodná veličina Y nabude hodnotu y:

$$p(x; y) = P(X = x; Y = y)\,. \tag{6.6}$$

Z definice přímo vyplívají následující vlastnosti:

Vlastnosti 6.2

1. *Obor hodnot pravděpodobnostní funkce* $p(x; y) = \langle 0; 1\rangle$.(6.6)
2. *Pro všechny složky náhodného vektoru platí:*
$$\sum_{i=1}^{n}\sum_{j=1}^{m} p(x_i; y_j) = 1 \tag{6.7}$$
3. *Pro distribuční funkci náhodného vektoru platí:*
$$F(x; y) = \sum_{x_i<x}^{n}\sum_{y_j<y}^{m} p(x_i; y_j). \tag{6.8}$$

Samozřejmě zavedeme také frekvenční funkce jednotlivých složek náhodného vektoru $\mathbf{X} = (X; Y)$.

Definice 6.5 (Marginální pravděopodobnostní funkce) *Marginální pravděpodobnostní funkce složky X nebo Y náhodného vektoru $\mathbf{X} = (X; Y)$ je dána vztahem:*

$$p_X(x) = P(X = x) = \sum_{j=1}^{m} p(x; y_j) \tag{6.9}$$

$$p_Y(y) = P(Y = y) = \sum_{i=1}^{n} p(x_i; y) \tag{6.10}$$

Hodnoty pravděpodobnostní funkce diskrétního náhodného zapisujeme do následující tabulky:

V posledním řádku a sloupci jsou marginální hodnoty složek X a Y.

Tabulka 6.1: Hodnoty pravděpodobnostní funkce

x$\backslash y$	y_1	y_2	y_3	$\ldots$	y_m	$p_X(x)$
x_1	$p(x_1; y_1)$	$p(x_1; y_2)$	$p(x_1; y_3)$	$\ldots$	$p(x_1; y_m)$	$p_X(x_1)$
x_2	$p(x_2; y_1)$	$p(x_2; y_2)$	$p(x_2; y_3)$	$\ldots$	$p(x_2; y_m)$	$p_X(x_2)$
x_3	$p(x_3; y_1)$	$p(x_3; y_2)$	$p(x_3; y_3)$	$\ldots$	$p(x_3; y_m)$	$p_X(x_3)$
$\vdots$	$\vdots$	$\vdots$	$\vdots$	$\ddots$	$\vdots$	$\vdots$
x_n	$p(x_n; y_1)$	$p(x_n; y_2)$	$p(x_n; y_3)$	$\ldots$	$p(x_n; y_m)$	$p_X(x_n)$
$p_Y(y)$	$p_Y(y_1)$	$p_Y(y_2)$	$p_Y(y_3)$	$\ldots$	$p_Y(y_m)$	

Náhodný vektor se spojitými složkami

U spojitého náhodného vektoru nelze počítat s pravděpodobnsotní funkcí. Stejně jako u spojité náhodné veličiny má náhodný vektor nespočetně mnoho hodnot a proto je pravděpodobnost, že nastane z tak velkého počtu výledků právě uspořádaná dvojice hodnot $[x_i; y_i]$:

$$P(X = x; Y = y) = 0$$

Pak stejně jako u náhodné veličiny zavedeme hustotu pravděpodobnosti.

Definice 6.6 (Hustota pravděpodobnosti náhodného vektoru X $= (X; Y)$) Hustota pravděpodobnosti je funkce $f(x; y)$, pro kterou platí:

$$F(x; y) = \int_{-\infty}^{x} \int_{-\infty}^{y} f(s; t)\, \mathrm{d}s\mathrm{d}t, \tag{6.11}$$

kde $x; y \in \mathbb{R}$

Z definice přímo vyplívají následující vlastnosti:

Vlastnosti 6.3

1. *Obor hodnot hustoty pravděpodobnosti:*

$$f(x; y) \geq 0 \tag{6.12}$$

2. *Pro všechny složky náhodného vektoru platí:*

$$\int_{-\infty}^{\infty} \int_{-\infty}^{\infty} f(x; y)\, \mathrm{d}x\mathrm{d}y = 1$$

3. *Ve všech bodech, kde existují derivace platí:*

$$f(x;y) = \frac{\partial^2 F(x;y)}{\partial x \partial y} \tag{6.13}$$

4. *Pravděpodobnost výskytu náhodného vektoru v podmnožině jeho definičiního oboru:*

$$P(a \leq X < b; c \leq Y < d) = \int_a^b \int_c^d f(s;t)\,\mathrm{d}s\mathrm{d}t \tag{6.14}$$

Jak je vidět, jedná se o plochu obdélníku ohraničeného hodnotami a, b, c a d. Do obdélníku můžeme klidně započítat i hranici, protože je tvořena jednotlivými body a jejich pravděpodobnost je nulová.

Samozřejmě zavedeme také frekvenční funkce složek náhodného vektoru $\mathbf{X} = (X;Y)$.

Definice 6.7 (Marginální hustota pravděpodobnosti složek náhodného vektoru) *Marginální hustota pravděpodobnosti $f_X(x)$ nebo $f_Y(y)$ složek náhodného vektoru $\mathbf{X} = (X;Y)$ je:*

$$f_X(x) = \int_{-\infty}^{\infty} f(x;y)\,\mathrm{d}y \tag{6.15}$$

$$f_Y(y) = \int_{-\infty}^{\infty} f(x;y)\,\mathrm{d}x \tag{6.16}$$

6.2 Podmíněná rozdělení a nezávislost náhodných veličin

Při studiu pravděpodobnosti jevů jsme narazili na problematiku podmíněné pravděpodobnosti. Připomeňme si co vyjadřovala. Podmíněná pravděpodobnost $P(A/B)$ určuje pravděpodobnost, že nastane jev A za podmínky, že také nastane jev B. Tu jsme určili vztahem:

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

kde $P(B) \neq 0$.

Stejnou pravděpodobnost, že nastane náhodná veličina X za podmínky uskutečnění veličiny Y můžeme pozorovat u náhodného vektoru $\mathbf{X} = (X;Y)$. V tomto případě nebudeme od sebe ve studiu oddělovat náhodné vektory se spojitými a diskrétními složkami vzhledem k jejich velké podobnosti.

Začneme podmíněnou pravděpodobností frekvenčních funkcí. Vezměme si příklad, kdy chceme určit hodnotu pravděpodobnostní funkce nebo hustotu pravděpodobnosti pro veličinu X jestliže chceme, aby zároveň veličina Y měla hodnotu y.

Definice 6.8 (Podmíněná pravděpodobností funkce diskrétní náhodné veličiny X nebo Y) Mějme náhodní vektor s diskrétními složkami $\mathbf{X} = (X; Y)$. Podmíněnou pravděpodobností funkcí veličiny X nebo Y nazveme funkci, pro kterou platí:

$$p\,(x/y) = \frac{p\,(x;y)}{p_Y\,(y)}, \tag{6.17}$$

$$p\,(y/x) = \frac{p\,(x;y)}{p_X\,(x)}, \tag{6.18}$$

kde funkce $p_Y\,(y)$ a $p_X\,(x)$ jsou marginální pravděpodobnostní funkce náhodných veličin X a Y, které zároveň musí být různé od nuly.

Definice 6.9 (Podmíněná pravděpodobností funkce spojité náhodné veličiny X nebo Y) Mějme náhodní vektor se spojitými složkami $\mathbf{X} = (X; Y)$. Podmíněnou hustotu pravděpodobnosti veličiny X nebo Y nazveme funkci, pro kterou platí:

$$f\,(x/y) = \frac{f\,(x;y)}{f_Y\,(y)}, \tag{6.19}$$

$$f\,(y/x) = \frac{f\,(x;y)}{f_X\,(x)}, \tag{6.20}$$

kde funkce $p_Y\,(y)$ a $p_X\,(x)$ jsou marginální pravděpodobnostní funkce náhodných veličin X a Y, které zároveň musí být různé od nuly.

Vzhledem k tomu, že umíme určit podmíněné frekvenční funkce náhodného vektoru $\mathbf{X} = (X; Y)$, pak z vlastnotí, které jsem si u nich uvedli umíme určit i podméněné distribuční funkce $F\,(x/y)$ a $F\,(y/x)$ spojitého i diskrétního náhodného vektoru.

Definice 6.10 (Podmíněná distribuční funkce diskrétní náhodné veličiny X nebo Y) Mějme náhodný vektor s diskrétními složkami $\mathbf{X} = (X; Y)$ a jeho pravděpodobnostní funkci $p\,(x; y)$. Podmíněnou distribuční funkci veličiny X nebo Y nazveme funkci, pro kterou platí:

$$F\,(x/y) = \frac{\sum_{x<x_i} p\,(x_i;y)}{p_Y\,(y)}, \tag{6.21}$$

$$F\,(y/x) = \frac{\sum_{y<y_i} p\,(x;y_i)}{p_X\,(x)}, \tag{6.22}$$

kde $p_Y(y)$ nebo $p_X(x)$ jsou marginální pravděpodobnostní funkce náhodné veličiny X nebo Y a zároveň musí být různá od nuly.

Definice 6.11 (Podmíněná distribuční funkce spojité náhodné veličiny X nebo Y) Mějme náhodný vektor se spojitými složkami $\mathbf{X} = (X;Y)$ a jeho hutotu pravděpodobnosti $f(x;y)$. Podmíněnou distribuční funkci veličiny X nebo Y nazveme funkci, pro kterou platí:

$$F(x/y) = \frac{1}{p_Y(y)} \cdot \int_{-\infty}^{x} f(t;y)\,\mathrm{d}t, \tag{6.23}$$

$$F(y/x) = \frac{1}{p_X(x)} \cdot \int_{-\infty}^{y} f(x;t)\,\mathrm{d}t, \tag{6.24}$$

kde $f_Y(y)$ nebo $f_X(x)$ jsou marginální hustoty pravděpodobnosti náhodné veličiny X nebo Y a zároveň musí být různá od nuly.

Nyní když jsme zavedli všechny podmíněné funkce můžeme na jejich základě zavést i nezávislost náhodných veličin. Nezávislost stejně jako u podmíněných pospisných funkcích zavedeme jako analogii k nezávislosti jevů z teorie pravděpodobnosti.

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

Definice 6.12 (Nezávislost složek náhodného vektoru $\mathbf{X} = (X;Y)$) Mějme náhodný vektor $\mathbf{X} = (X;Y)$. Náhodné veličiny X a Y jsou nezávislé jestliže pro jejich marginální distiribuční funkce paltí:

$$F(x;y) = F_X(x) \cdot F_Y(y)\,. \tag{6.25}$$

Konkrétně pro náhodný vektor s diskrétními složkami platí, že jsou nezávislé pokud splňují rovnici:

$$p(x;y) = p_X(x) \cdot p_Y(y) \tag{6.26}$$

Konkrétně pro náhodný vektor se spojitými složkami platí, že jsou nezávislé pokud splňují rovnici:

$$f(x;y) = f_X(x) \cdot f_Y(y) \tag{6.27}$$

Pro všechny hodnoty veličin X a Y.

6.3 Charakteristiky náhodného vektoru

Díky tomu, že složky náhodného vektoru jsou náhodné veličiny, můžeme i u náhodného vektoru studovat jeho významné charakteristiky, které pro lepší přehlednost dělíme do tří skupin:

Marginální charakteristiky - podle názvu se jedná o charakteristiky marginálních funkcí, které budou popisovat jejich polohu, variabilitu, šikmost a špičatost.

Podmíněné charakteristiky - podle názvu se jedná o charakteristiky podmíněných funkcí. Ukážeme si podobné charakteristiky jako u marginálních funkcí.

Vztah mezi veličinami náhodného vektoru X $= (X;Y)$ - tyto charakteristiky se týkají kovariance a koeficientu korelace. Tyto charakteristiky nás budou informovat o vzájemném vztahu složek náhodného vektoru.

6.3.1 Marginální charakteristiky

Stejně jako u náhodné veličiny začneme charakteristikou polohy = střední hodnota. Střední hodnotu háhodného vektoru budeme definovat podobně jako u náhodné veličiny a proto musíme rozlišit náhodný vektor se spojitými a diskrétními složkami **X** $= (X;Y)$

Definice 6.13 (Střední hodnota diskrétní náhodné veličiny X a Y) Mějme náhodný vektor s diskrétními složkami **X** $= (X;Y)$ a jeho pravděpodobnostní funkce. Střední hodnotu náhodné veličiny X nebo Y definujeme vztahy:

$$E\left(X\right) = \mu_x = \sum_{i=1}^{n} x_i \cdot p_X\left(x_i\right), \tag{6.28}$$

$$E\left(Y\right) = \mu_y = \sum_{i=1}^{m} y_i \cdot p_Y\left(y_i\right), \tag{6.29}$$

kde $p_X\left(x\right)$ a $p_Y\left(y\right)$ jsou marginální pravděpodobnostní funkce.

Definice 6.14 (Střední hodnota spojité náhodné veličiny X a Y) Mějme náhodný vektor se spojitými složkami **X** $= (X;Y)$ a jeho hustotu pravděpodobnosti. Střední hodnotu náhodné veličiny X nebo Y definujeme vztahy:

$$E\left(X\right) = \mu_x = \int_{-\infty}^{\infty} x \cdot f_X\left(x\right) \mathrm{d}x, \tag{6.30}$$

$$E(Y) = \mu_y = \int_{-\infty}^{\infty} y \cdot f_Y(y)\,\mathrm{d}y, \tag{6.31}$$

kde $f_X(x)$ a $f_Y(y)$ jsou marginální pravděpodobnostní funkce.

Definice 6.15 (Rozptyl diskrétní náhodné veličiny X a Y) Mějme náhodný vektor s diskrétními složkami $\mathbf{X} = (X; Y)$, jeho diskrétní pravděpodobnostní funkce a střední hodnoty náhodných veličin X a Y. Rozptyl náhodné veličiny X a Y definujeme vztahy:

$$\sigma_x^2 = \sum_{i=1}^{n} (x_i - \mu_x)^2 \cdot p_X(x_i) = \sum_{i=1}^{n} x_i^2 \cdot p_X(x_i) - \mu_x^2 \tag{6.32}$$

$$\sigma_y^2 = \sum_{i=1}^{m} (y_i - \mu_y)^2 \cdot p_Y(y_i) = \sum_{i=1}^{m} y_i^2 \cdot p_Y(y_i) - \mu_y^2 \tag{6.33}$$

Definice 6.16 (Rozptyl spojité náhodné veličiny X a Y) Mějme náhodný vektor se spojitými složkami $\mathbf{X} = (X; Y)$, jeho hustoty pravděpodobnosti a střední hodnoty náhodných veličin X a Y. Rozptyl náhodné veličiny X a Y definujeme vztahy:

$$\sigma_x^2 = \int_{-\infty}^{\infty} (x - \mu_x)^2 \cdot f_X(x)\,\mathrm{d}x = \int_{-\infty}^{\infty} x^2 \cdot f_X(x)\,\mathrm{d}x - \mu_x^2 \tag{6.34}$$

$$\sigma_y^2 = \int_{-\infty}^{\infty} (y - \mu_y)^2 \cdot f_Y(y)\,\mathrm{d}y = \int_{-\infty}^{\infty} y^2 \cdot f_Y(y)\,\mathrm{d}y - \mu_y^2 \tag{6.35}$$

K výpočtům většinou používáme druhou část předchozích rovnic.

6.3.2 Podmíněné charakteristiky

Stejně jako u marginálních charakteristik začneme charakteristikou polohy = střední hodnota.

Definice 6.17 (Střední hodnota diskrétní podmíněné náhodné veličiny X a Y) Mějme náhodný vektor s diskrétními složkami $\mathbf{X} = (X; Y)$ a jeho podmíněnou pravděpodobnostní funkci. Podmíněnou střední hodnotu náhodné veličiny X nebo Y definujeme vztahy:

$$E(X) = \sum_{i=1}^{n} x_i \cdot p(x_i/y), \tag{6.36}$$

$$E(Y) = \sum_{i=1}^{m} y_i \cdot p(y_i/x). \tag{6.37}$$

Definice 6.18 (Střední hodnota spojité podmíněné náhodné veličiny X a Y) Mějme náhodný vektor se spojitými složkami $\mathbf{X} = (X;Y)$ a jeho podmíněnou hustotu pravděpodobnosti. Střední hodnotu náhodné veličiny X nebo Y definujeme vztahy:

$$E(X) = \int_{-\infty}^{\infty} x \cdot f(x/y)\,\mathrm{d}x, \tag{6.38}$$

$$E(Y) = \int_{-\infty}^{\infty} y \cdot f(y/x)\,\mathrm{d}y. \tag{6.39}$$

U podmíněné střední hodnoty náhodné veličiny se také setkáme s názvem regresní funkce veličiny X vzhledem k veličině Y a naopak. Důvodem tohoto označení je závislost frekvenční funkce na veličině Y respektive X.

Další podmíněnou charakteristikou je podmíněný rozptyl veličin náhodného vektoru.

Definice 6.19 (Podmíněný rozptyl diskrétní náhodné veličiny X a Y) Mějme náhodný vektor s diskrétními složkami $\mathbf{X} = (X;Y)$, jeho podmíněnou pravděpodobnostní funkcí a střední hodnotou. Podmíněné rozptyly veličin X a Y definujeme takto:

$$D(x/y) = \sum_{i=1}^{n} (x_i - E(x/y))^2 \cdot p(x_i/y), \tag{6.40}$$

$$D(y/x) = \sum_{i=1}^{m} (y_i - E(y/x))^2 \cdot p(y_i/x). \tag{6.41}$$

Definice 6.20 (Podmíněný rozptyl spojité náhodné veličiny X a Y) Mějme náhodný vektor se spojitými složkami $\mathbf{X} = (X;Y)$, jeho podmíněnou hustotu pravděpodobnosti a střední hodnotou. Podmíněné rozptyly veličin X a Y definujeme takto:

$$D(x/y) = \int_{-\infty}^{\infty} (x - E(x/y))^2 \cdot f(x/y), \tag{6.42}$$

$$D(y/x) = \int_{-\infty}^{\infty} (y - E(y/x))^2 \cdot f(y/x). \tag{6.43}$$

K výpočtu podmíněného rozptylu se používá podmíněná střední hodnota, proto i podmíněný rozptyl veličiny X nebo Y je závislý na veličině Y respektive X. Podmíněný rozptyl nazveme také skedastická funkce. Dostaneme-li náhodný vektor jehož podmíněný rozptyl je konstantní nazveme ho homoskedastická funkce.

6.3.3 Vztah mezi veličinami náhodného vektoru $\mathbf{X} = (X;Y)$

Vztah mezi veličinami náhodného vektoru $\mathbf{X} = (X;Y)$ popisuje kovariance a koeficient korelace. Výpočet kovariance a koeficientu korelace je stejný pro spojité i diskrétní složky náhodného vektoru $\mathbf{X} = (X;Y)$. Proto tyto dvě charakteristiky nebudeme zavádět zvlášť.

Definice 6.21 (Kovariance veličin náhodného vektoru $\mathbf{X} = (X;Y)$) Mějme náhodný vektor $\mathbf{X} = (X;Y)$ a jeho střední hodnoty. Kovarianci náhodných veličin X a Y určíme střední hodnotu součinu jejich odchylek:

$$\operatorname{cov}(X;Y) = E\left[(X - E(X)) \cdot (Y - E(Y))\right] = E(X \cdot Y) - E(X) - E(Y). \tag{6.44}$$

Kovariance vyjadřuje míru vzájemné vazby mezi dvěma náhodnými veličinami X a Y. K výpočtům se pro její jednoduchost používá druhá rovnost z definice.

Uvedeme si ještě významné hodnoty kovariance:

1. $\operatorname{cov}(X;X) = \sigma_x^2$
2. $\operatorname{cov}(Y;Y) = \sigma_y^2$
3. $\operatorname{cov}(X;Y) = \operatorname{cov}(Y;X)$
4. Jestliže jsou náhodné veličiny X a Y nezávislé platí:

$$\operatorname{cov}(X;Y) = 0$$

Poslední charakteristiku, kterou si ukážeme je koeficient korelace. Ten určuje míru lineární závislosti náhodných veličin X a Y.

Definice 6.22 (Koeficient korelace veličin náhodného vektoru $\mathbf{X} = (X;Y)$) Mějme náhodný vektor $\mathbf{X} = (X;Y)$ a rozptyly jeho veličin. Koeficient korelace $\rho(X;Y)$ definujeme vztahem:

$$\rho(X;Y) = \frac{\operatorname{cov}(X;Y)}{\sqrt{\sigma_x^2 \cdot \sigma_y^2}} = \frac{\operatorname{cov}(X;Y)}{\sigma_x \cdot \sigma_y}. \tag{6.45}$$

Uveďme si ještě významné vlastnosti koeficientu korelace:

1. $|\rho(X;Y)| \leq 1$

2. Jestliže pro koeficient korelace platí $\rho(X;Y) = \pm 1$, pak mezi veličinami X a Y existuje lineární závislost:
$$Y = aX + b,$$
kde $a; b \in \mathbb{R}$.
3. Jesltiže pro koeficient korelace platí $\rho(X;Y) = 0$, pak veličiny X a Y jsou nekorelované.

   Poznámka: Vlastnost 3 ovšem neříká, že jsou nezávislé o nezávislosti nás informuje kovariance.

4. Jestliže $\rho(X;Y) > 0$ respektive $\rho(X;Y) < 0$ mluvíme o kladné (přímé) respektive o záporné (nepřímé) korelaci. To znamená obě veličiny současně rostou respektive jedna roste a druhá klesá.
5. Jestliže $\rho(X;Y) \cong \pm 1$ respektive $\rho(X;Y) \cong 0$ veličiny X a Y jsou silně lineární respektive slabě lineární.

■ **Řešený příklad 6.1** Náhodný vektor $\mathbf{X} = (X;Y)$ je dán pravděpodobnostní funkcí (6.2). Určete chybějící hodnotu pravděpodobnostní funkce $p(2;2)$, střední hodnoty a rozptyly veličin X a Y, distribuční funkci náhodného vektoru, $P(X = 3; Y = 2)$, $P(X < 3; Y < 2)$, $P(X > 3; Y > 2)$, $P(1,5 < X \leq 2,5; 2 < Y \leq 4,5)$, podmíněné pravděpodobnostní funkce, podmíněné střední hodnoty a rozptyly. Rozhodněte jestli jsou veličiny X a Y závislé. V případě, že jsou složky vektoru závislé určete kovarianci a koeficient korelace.

Tabulka 6.2: Hodnoty pravděpodobnostní funkce k příkladu (6.1)

$x \setminus y$	0	2	4	6
1	0,05	0,10	0,15	0
2	0,15	?	0	0,05
3	0,20	0	0,05	0,15

Řešení:

Určíme hodnotu pravděpodobnostní funkce $p\,(2;2)$. Z vlastností pravděpodobnostní funkce víme, že platí:

$$\begin{aligned}
1 &= \sum_{i=1}^{3}\sum_{j=0}^{6} p\,(x_i;y_j) \\
1 &= 0,05+0,10+0,15+0+0,15+p\,(2;2)+0+0,05+0,20+0+0,05+0,15 \\
p\,(2;2) &= 0,10
\end{aligned}$$

Tabulku pravděpodobnostní funkce znovu napíšeme (6.3), doplníme o spočítanou hodnotu a rozšíříme ji o vypočtené marginální pravděpodobností funkce a centrální momenty veličin, z kterých posléze získáme střední hodnoty a rozptyly veličin.

Tabulka 6.3: Doplněné hodnoty pravděpodobnostní funkce k příkladu (6.1)

$x \setminus y$	0	2	4	6	$p_X\,(x)$	$x \cdot p_X\,(x)$	$x^2 \cdot p_X\,(x)$
1	0,05	0,10	0,15	0	0,3	0,3	0,3
2	0,15	0,10	0	0,05	0,3	0,6	1,2
3	0,20	0	0,05	0,15	0,4	1,2	3,6
$p_Y\,(y)$	0,4	0,2	0,2	0,2	$\sum$	2,1	5,1
$y \cdot p_Y\,(y)$	0	0,4	0,8	1,2	2,4	-	-
$y^2 \cdot p_Y\,(y)$	0	0,8	3,2	7,2	11,2	-	-

Střední hodnoty a rozptyly náhodných veličin počítáme podle definice (6.13) a (6.15):

$$\begin{aligned}
E\,(X) &= \sum_{i=1}^{3} x_i \cdot p_X\,(x_i) = 1\cdot 0,3+2\cdot 0,3+3\cdot 0,4 = 2,1 \\
E\left(X^2\right) &= \sum_{i=1}^{3} x_i^2 \cdot p_X\,(x_i) = 1^2\cdot 0,3+2^2\cdot 0,3+3^2\cdot 0,4 = 5,1 \\
E\,(Y) &= \sum_{i=0}^{6} y_i \cdot p_Y\,(y_i) = 0\cdot 0,4+2\cdot 0,2+4\cdot 0,2+6\cdot 0,2 = 2,4 \\
E\left(Y^2\right) &= \sum_{i=0}^{6} y_i^2 \cdot p_Y\,(y_i) = 0^2\cdot 0,4+2^2\cdot 0,2+4^2\cdot 0,2+6^2\cdot 0,2 = 11,2
\end{aligned}$$

$$\begin{aligned}
D\left(X\right) &= E\left(X^2\right) - \left[E\left(X\right)\right]^2 = 5,1 - 2,1^2 = 0,69\\
D\left(Y\right) &= E\left(Y^2\right) - \left[E\left(Y\right)\right]^2 = 11,2 - 2,4^2 = 5,44\\
E\left(X;Y\right) &= \sum_{i=1}^{3}\sum_{j=0}^{6} x_i \cdot y_j \cdot p\left(x_i;y_j\right) = 1\cdot 0\cdot 0,05 + 1\cdot 2\cdot 0,10 + 1\cdot 4\cdot 0,15 + 1\cdot 6\cdot 0 +\\
&+2\cdot 0\cdot 0,15 + 2\cdot 2\cdot 0,10 + 2\cdot 4\cdot 0 + 2\cdot 6\cdot 0,05 + 3\cdot 0\cdot 0,20 + 3\cdot 2\cdot 0 +\\
&+3\cdot 4\cdot 0,05 + 3\cdot 6\cdot 0,15 = 5,1
\end{aligned}$$

O závislosti veličiny X a Y rozhodneme podle definice (6.12):

$$\begin{aligned}
p\left(x_i;y_j\right) &= f_X\left(x_i\right)\cdot f_Y\left(y_j\right)\\
p\left(1;0\right) &\neq p_X\left(1\right)\cdot p_Y\left(0\right)
\end{aligned}$$

Našli jsme aspoň jednu dvojici marginálních pravděpodobnostních funkcí, které nesplňují podmínku nezávislosti. Veličiny X a Y jsou závislé, proto má smysl určit kovarianci a koeficient korelace podle definic (6.21) a (6.22).

$$\begin{aligned}
\operatorname{cov}\left(X;Y\right) &= E\left(X;Y\right) - E\left(X\right)\cdot E\left(Y\right) = 5,1 - 2,1\cdot 2,4 = 0,06\\
\rho\left(X;Y\right) &= \frac{\operatorname{cov}\left(X;Y\right)}{\sqrt{D\left(X\right)\cdot D\left(Y\right)}} = \frac{0,06}{\sqrt{0,69\cdot 5,44}} = 0,031
\end{aligned}$$

Veličiny X a Y jsou slabě závislé. Z koeficientu korelace jsme se navíc dozvěděli, že veličiny X a Y jsou slabě přímo lineární.

Distribuční funkci určíme podle definice (6.2) a zapíšeme do tabulky (6.4)

Tabulka 6.4: Distribuční funkce k příkladu (6.1)

$x \setminus y$	$(-\infty;0\rangle$	$(0;2\rangle$	$(2;4\rangle$	$(4;6\rangle$	$(6;\infty)$
$(-\infty;1\rangle$	0	0	0	0	0
$(1;2\rangle$	0	0,05	0,15	0,30	0,30
$(2;3\rangle$	0	0,20	0,40	0,55	0,60
$(3;\infty)$	0	0,40	0,60	0,80	1

Pravděpodobnosti náhodného vektoru určíme podle definice (6.2) distribuční funkce a jejích vlastností:

$$\begin{aligned}
P(X=3;Y=2) &= p(3;2)=0\\
P(X<3;Y<2) &= p(1;0)+p(2;0)=0,05+0,15=0,20\\
P(X>2;Y>2) &= p(3;4)+p(3;6)=0,05+0,15=0,20\\
P(1,5<X\leq 2,5;2<Y\leq 4,5) &= p(2;2)+p(2;4)=0,10+0=0,10
\end{aligned}$$

Podmíněné pravděpodobnostní funkce a následně podmíněné číselné charakteristiky určíme podle definic (6.8), (6.17) a (6.19):

$$p(x_i/y_j)=\frac{p(x_i;y_j)}{p_Y(y_j)}$$

Tabulka 6.5: Podmíněná pravděpodobnostní funkce $p(x/y)$ k příkladu (6.1)

$x \setminus y$	0	2	4	6
1	$\frac{1}{8}$	$\frac{1}{2}$	$\frac{3}{4}$	0
2	$\frac{3}{8}$	$\frac{1}{2}$	0	$\frac{1}{4}$
3	$\frac{1}{2}$	0	$\frac{1}{4}$	$\frac{3}{4}$
$x \cdot p(x/y)$	$\frac{19}{8}$	$\frac{3}{2}$	$\frac{3}{2}$	$\frac{11}{4}$
$x^2 \cdot p(x/y)$	$\frac{49}{8}$	$\frac{5}{2}$	$\frac{3}{2}$	$\frac{31}{4}$

$$p(y_j/x_i)=\frac{p(x_i;y_j)}{p_X(x_i)}$$

Tabulka 6.6: Podmíněná pravděpodobnostní funkce $p(y/x)$ k příkladu (6.1)

$x \setminus y$	0	2	4	6	$y \cdot p(y/x)$	$y^2 \cdot p(y/x)$
1	$\frac{1}{6}$	$\frac{1}{3}$	$\frac{1}{2}$	0	$\frac{8}{3}$	$\frac{28}{3}$
2	$\frac{1}{2}$	$\frac{1}{3}$	0	$\frac{1}{6}$	$\frac{5}{3}$	$\frac{22}{3}$
3	$\frac{1}{2}$	0	$\frac{1}{8}$	$\frac{3}{8}$	$\frac{22}{8}$	$\frac{124}{8}$

Do tabulek (6.5) a (6.6) jsme si připravili i centrální momenty, které využijeme k výpočtům středních hodnot a rozptylů.

$$\begin{aligned}
E(X/y) &= \sum_{i=1}^{3} x_i \cdot p(x_i/y) \\
E\left(X^2/y\right) &= \sum_{i=1}^{3} x_i^2 \cdot p(x_i/y) \\
E(X/0) &= \frac{19}{8} \qquad E(X/2) = \frac{3}{2} \qquad E(X/4) = \frac{3}{2} \qquad E(X/6) = \frac{11}{4} \\
E\left(X^2/0\right) &= \frac{49}{8} \qquad E\left(X^2/2\right) = \frac{5}{2} \qquad E\left(X^2/4\right) = \frac{12}{4} \qquad E\left(X^2/6\right) = \frac{31}{4} \\
E(Y/x) &= \sum_{i=0}^{6} y_i \cdot p(y_i/x) \\
E\left(Y^2/x\right) &= \sum_{i=0}^{6} y_i^2 \cdot p(y_i/x) \\
E(Y/1) &= \frac{8}{3} \qquad E(Y/2) = \frac{5}{3} \qquad E(Y/3) = \frac{8}{3} \\
E\left(Y^2/1\right) &= \frac{28}{3} \qquad E\left(Y^2/2\right) = \frac{22}{3} \qquad E\left(Y^2/3\right) = \frac{124}{8} \\
D(X/y) &= E\left(X^2/y\right) - [E(X/y)]^2 \\
D(X/0) &= \frac{31}{64} \qquad D(X/2) = \frac{1}{4} \qquad D(X/4) = \frac{3}{4} \qquad D(X/6) = \frac{3}{16} \\
D(Y/x) &= E\left(Y^2/x\right) - [E(Y/x)]^2 \\
D(Y/1) &= \frac{20}{9} \qquad D(Y/2) = \frac{41}{9} \qquad D(Y/3) = \frac{127}{16}
\end{aligned}$$

■ **Řešený příklad 6.2** Náhodný vektor **X** $= (X;Y)$ má hustotu pravděpodobnosti:

$$\begin{aligned}
f(x;y) &= K \cdot x^2 (x+y)^2 \qquad \text{kde: } x \in \langle 0;1\rangle\,; \quad y \in \langle 0;1\rangle \\
f(x;y) &= 0 \qquad \text{kde: } x \notin \langle 0;1\rangle\,; \quad y \notin \langle 0;1\rangle
\end{aligned}$$

Určete konstantu K, distribuční funkci, všechny marginální funkce, podmíněné hustoty pravděpodobnosti, podmíněné distribuční funkce a podmíněné marginální charakteristiky a číselné charakteristiky. Rozhodněte, jestli jsou veličiny X a Y nezávislé. Pokud jsou závislé určete kovarianci a koeficient korelace.

Řešení:

Nejprve určíme konstantu K. Z vlastností hustoty pravděpodobnosti náhodného vektoru $\mathbf{X} = (X;Y)$ víme, že platí:

$$\int_{-\infty}^{\infty}\int_{-\infty}^{\infty} f(x;y)\,\mathrm{d}x\mathrm{d}y = 1$$

Z této rovnice určíme konstantu K.

$$\begin{aligned}
1 &= \int_{-\infty}^{\infty}\left[\int_{-\infty}^{\infty} K\cdot x^2(x+y)^2\,\mathrm{d}y\right]\mathrm{d}x = \underbrace{\int_{-\infty}^{0}\left[\int_{-\infty}^{0} K\cdot x^2(x+y)^2\,\mathrm{d}y\right]\mathrm{d}x}_{0} + \\
&\quad + \int_{0}^{1}\left[\int_{0}^{1} K\cdot x^2(x+y)^2\,\mathrm{d}y\right]\mathrm{d}x + \underbrace{\int_{1}^{\infty}\left[\int_{1}^{\infty} K\cdot x^2(x+y)^2\,\mathrm{d}y\right]\mathrm{d}x}_{0} = \\
&= K\int_0^1\left(x^2\int_0^1 (x+y)^2\,\mathrm{d}y\right)\mathrm{d}x = K\int_0^1\left(x^2\int_0^1\left(x^2+2xy+y^2\right)\mathrm{d}y\right)\mathrm{d}x = \\
&= K\int_0^1 x^2\left[x^2y+\frac{2xy^2}{2}+\frac{y^3}{3}\right]_0^1\mathrm{d}x = K\int_0^1\left(x^2\left[x^2+x+\frac{1}{3}\right]\right)\mathrm{d}x = \\
&= K\int_0^1\left(x^4+x^3+\frac{x^2}{3}\right)\mathrm{d}x = K\left[\frac{x^5}{5}+\frac{x^4}{4}+\frac{x^3}{9}\right]_0^1 = K\left(\frac{1}{5}+\frac{1}{4}+\frac{1}{9}\right) \\
1 &= K\cdot\frac{36+45+20}{180} \\
K &= \frac{180}{101}
\end{aligned}$$

Distribuční funkci určíme užitím další vlasnosti hustoty pravděpodobnosti:

$$f(x;y) = \frac{\partial^2 F(x;y)}{\partial x\partial y}$$

Po úpravě tohoto vztahu dostaneme:

$$F(x;y) = \int_{-\infty}^{x}\left[\int_{-\infty}^{y} f(s;t)\,\mathrm{d}t\right]\mathrm{d}s = \int_{-\infty}^{x}\left[\int_{-\infty}^{y}\frac{180}{101}s^2(s+t)^2\,\mathrm{d}t\right]\mathrm{d}s =$$

$$
\begin{aligned}
&= \frac{180}{101}\int\limits_{-\infty}^{x} s^2\left[\int\limits_{-\infty}^{y}\left(s^2+2st+t^2\right)\mathrm{d}t\right]\mathrm{d}s = \frac{180}{101}\int\limits_{-\infty}^{x} s^2\left[s^2t+\frac{2st^2}{2}+\frac{t^3}{3}\right]_{\infty}^{y}\mathrm{d}s = \\
&= \frac{180}{101}\int\limits_{-\infty}^{x} s^2\left[s^2y+sy^2+\frac{y^3}{3}-0-0-0\right]\mathrm{d}s = \frac{180}{101}\int\limits_{-\infty}^{x}\left[s^4y+s^3y^2+\frac{s^2y^3}{3}\right]\mathrm{d}s = \\
&= \frac{180}{101}\left[\frac{s^5y}{5}+\frac{s^4y^2}{4}+\frac{s^3y^3}{3\cdot 3}\right]_{-\infty}^{x} = \frac{180}{101}\left(\frac{x^5y}{5}+\frac{x^4y^2}{4}+\frac{x^3y^3}{9}\right) = \\
&= \frac{36x^5y+45x^4y^2+20x^3y^3}{101}
\end{aligned}
$$

Marginální hustoty pravděpodobnosti určíme podle definice (6.7). Při součtu hodnot pravděpodobnostní funkce mimo interval $\langle 0;1\rangle$ k sobě dáváme samé nuly a proto ve výpočtu již tyto intervaly nebudeme rozepisovat.

$$
\begin{aligned}
f_X(x) &= \int\limits_{-\infty}^{\infty} f(x;y)\,\mathrm{d}y = \frac{180}{101}\int\limits_{0}^{1} x^2(x+y)^2\,\mathrm{d}y = \frac{180}{101}x^2\int\limits_{0}^{1}\left(x^2+2xy+y^2\right)\mathrm{d}y = \\
&= \frac{180}{101}x^2\left[x^2y+\frac{2xy^2}{2}+\frac{y^3}{3}\right]_0^1 = \frac{180}{101}x^2\left(x^2+x+\frac{1}{3}\right) = \frac{60}{101}x^2\left(3x^2+3x+1\right) \\
f_Y(y) &= \int\limits_{-\infty}^{\infty} f(x;y)\,\mathrm{d}x = \frac{180}{101}\int\limits_{0}^{1} x^2(x+y)^2\,\mathrm{d}x = \frac{180}{101}\int\limits_{0}^{1}\left(x^4+2x^3y+x^2y^2\right)\mathrm{d}x = \\
&= \frac{180}{101}\left[\frac{x^5}{5}+\frac{2x^4y}{4}+\frac{x^3y^2}{3}\right]_0^1 = \frac{180}{101}\left(\frac{1}{5}+\frac{y}{2}+\frac{y^2}{3}\right) = \frac{6}{101}\left(10y^2+15y+6\right)
\end{aligned}
$$

Stejně jako distribuční funkci určíme i marginální distribuční funkce:

$$
\begin{aligned}
F_X(x) &= \int\limits_{-\infty}^{x} f_T(t)\,\mathrm{d}t = \frac{60}{101}\int\limits_{-\infty}^{x} t^2\left(3t^2+3t+1\right)\mathrm{d}t = \frac{60}{101}\int\limits_{-\infty}^{x}\left(3t^4+3t^3+t^2\right)\mathrm{d}t = \\
&= \frac{60}{101}\left[\frac{3t^5}{5}+\frac{3t^4}{4}+\frac{t^3}{3}\right]_{-\infty}^{x} = \frac{60}{101}\left(\frac{3x^5}{5}+\frac{3x^4}{4}+\frac{x^3}{3}-0-0-0\right) = \\
&= \frac{36x^5+45x^4+20x^3}{101} \\
F_Y(y) &= \int\limits_{-\infty}^{y} f_T(t)\,\mathrm{d}t = \frac{6}{101}\int\limits_{-\infty}^{y}\left(10t^2+15t+6\right)\mathrm{d}t = \frac{6}{101}\left[\frac{10t^3}{3}+\frac{15t^2}{2}+6t\right]_{-\infty}^{y} =
\end{aligned}
$$

$$= \frac{6}{101}\left(\frac{10y^3}{3}+\frac{15y^2}{2}+6y-0-0-0\right)=\frac{20y^3+30y^2+36y}{101}$$

Podmíněné hustoty pravděpodobnosti a distribuční funkce určíme podle definice (6.9) a (6.11).

$$\begin{aligned}
f(x/y) &= \frac{f(x;y)}{f_Y(y)}=\frac{\frac{180}{101}x^2(x+y)^2}{\frac{6}{101}(10y^2+15y+6)}=\frac{30x^2(x+y)^2}{10y^2+15y+6}\\
f(y/x) &= \frac{f(x;y)}{f_X(x)}=\frac{\frac{180}{101}x^2(x+y)^2}{\frac{60}{101}x^2(3x^2+3x+1)}=\frac{3(x+y)^2}{3x^2+3x+1}\\
F(x/y) &= \frac{\int\limits_{-\infty}^{x} f(t;y)\,\mathrm{d}t}{f_Y(y)}=\frac{\frac{180}{101}\int\limits_{-\infty}^{x} t^2(t+y)^2\,\mathrm{d}t}{\frac{6}{101}(10y^2+15y+6)}=\frac{30\int\limits_{-\infty}^{x}(t^4+2t^3y+t^2y^2)\,\mathrm{d}t}{10y^2+15y+6}=\\
&= \frac{30\left[\frac{t^5}{5}+\frac{2t^4y}{4}+\frac{t^3y^2}{3}\right]_{-\infty}^{x}}{10y^2+15y+6}=\frac{\frac{30}{60}(12x^5+30x^4y+20x^3y^2-0-0-0)}{10y^2+15y+6}=\\
&= \frac{6x^5+15x^4y+10x^3y^2}{10y^2+15y+6}\\
F(y/x) &= \frac{\int\limits_{-\infty}^{y} f(x;t)\,\mathrm{d}t}{f_X(x)}=\frac{\frac{180}{101}x^2\int\limits_{-\infty}^{y}(x+t)^2\,\mathrm{d}t}{\frac{60}{101}x^2(3x^2+3x+1)}=\frac{3\int\limits_{-\infty}^{y}(x^2+2xt+t^2)\,\mathrm{d}t}{3x^2+3x+1}=\\
&= \frac{3\left[x^2t+\frac{2xt^2}{2}+\frac{t^3}{3}\right]_{-\infty}^{y}}{3x^2+3x+1}=\frac{3\left(x^2y+xy^2+\frac{y^3}{3}-0-0-0\right)}{3x^2+3x+1}=\frac{3x^2y+3xy^2+y^3}{3x^2+3x+1}
\end{aligned}$$

Střední hodnoty a rozptyly podle definic (6.14) a (6.16).

$$\begin{aligned}
E(X) &= \int\limits_{-\infty}^{\infty} x\cdot f_X(x)\,\mathrm{d}x=\frac{60}{101}\int\limits_{0}^{1} x^3\left(3x^2+3x+1\right)\mathrm{d}x=\\
&= \frac{60}{101}\int\limits_{0}^{1}\left(3x^5+3x^4+x^3\right)\mathrm{d}x=\frac{60}{101}\left[\frac{3x^6}{6}+\frac{3x^5}{5}+\frac{x^4}{4}\right]_0^1=\\
&= \frac{60}{101}\left(\frac{1}{2}+\frac{3}{5}+\frac{1}{4}-0-0-0\right)=\frac{81}{101}\\
E\left(X^2\right) &= \int\limits_{-\infty}^{\infty} x^2\cdot f_X(x)\,\mathrm{d}x=\frac{60}{101}\int\limits_{0}^{1} x^4\left(3x^2+3x+1\right)\mathrm{d}x=
\end{aligned}$$

$$
\begin{aligned}
&= \frac{60}{101}\int\limits_0^1 \left(3x^6+3x^5+x^4\right)\mathrm{d}x = \frac{60}{101}\left[\frac{3x^7}{7}+\frac{3x^6}{6}+\frac{x^5}{5}\right]_0^1 = \\
&= \frac{60}{101}\left(\frac{3}{7}+\frac{1}{2}+\frac{1}{5}-0-0-0\right) = \frac{474}{707} \\
E\left(Y\right) &= \int\limits_{-\infty}^{\infty} y\cdot f_Y\left(y\right)\mathrm{d}y = \frac{6}{101}\int\limits_0^1 y\left(10y^2+15y+6\right)\mathrm{d}y = \\
&= \frac{6}{101}\int\limits_0^1 \left(10y^3+15y^2+6y\right)\mathrm{d}y = \frac{6}{101}\left[\frac{10y^4}{4}+\frac{15y^3}{3}+\frac{6y^2}{2}\right]_0^1 = \\
&= \frac{6}{101}\left(\frac{5}{2}+5+3-0-0-0\right) = \frac{63}{101} \\
E\left(Y^2\right) &= \int\limits_{-\infty}^{\infty} y^2\cdot f_Y\left(y\right)\mathrm{d}y = \frac{6}{101}\int\limits_0^1 y^2\left(10y^2+15y+6\right)\mathrm{d}y = \\
&= \frac{6}{101}\int\limits_0^1 \left(10y^4+15y^3+6y^2\right)\mathrm{d}y = \frac{6}{101}\left[\frac{10y^5}{5}+\frac{15y^4}{4}+\frac{6y^3}{3}\right]_0^1 = \\
&= \frac{6}{101}\left(2+\frac{15}{4}+2-0-0-0\right) = \frac{93}{202} \\
D\left(X\right) &= E\left(X^2\right)-\left[E\left(X\right)\right]^2 = \frac{474}{707}-\left(\frac{81}{101}\right)^2 = 0,0273 \\
D\left(Y\right) &= E\left(Y^2\right)-\left[E\left(Y\right)\right]^2 = \frac{93}{202}-\left(\frac{63}{101}\right)^2 = 0,0713 \\
E\left(X;Y\right) &= \int\limits_{-\infty}^{\infty}\left[\int\limits_{-\infty}^{\infty} xyf\left(x;y\right)\mathrm{d}y\right]\mathrm{d}x = \frac{180}{101}\int\limits_0^1\left[\int\limits_0^1 x^3y\left(x+y\right)^2\mathrm{d}y\right]\mathrm{d}x = \\
&= \frac{180}{101}\int\limits_0^1\left[\int\limits_0^1 \left(x^5y+2x^4y^2+x^3y^2\right)\mathrm{d}y\right]\mathrm{d}x = \\
&= \frac{180}{101}\int\limits_0^1\left[\frac{x^5y^2}{2}+\frac{2x^4y^3}{3}+\frac{x^3y^4}{4}\right]_0^1\mathrm{d}x = \\
&= \frac{180}{101}\int\limits_0^1\left[\frac{x^5}{5}+\frac{2x^4}{3}+\frac{x^3}{4}-0-0-0\right]\mathrm{d}x = \frac{180}{101}\left[\frac{x^6}{12}+\frac{2x^5}{15}+\frac{x^4}{16}\right]_0^1 =
\end{aligned}
$$

$$= \frac{180}{101}\left(\frac{1}{12}+\frac{2}{15}+\frac{1}{16}-0-0-0\right)=\frac{201}{404}$$

Určíme jestli jsou veličiny X a Y nezávislé podle definice (6.12). Určili jsme marginální hustoty pravděpodobností i distribuční funkce. K ověření nezávislosti si vybereme ty, které mají jednodušší tvar.

$$\begin{aligned} f(x;y) &= f_X(x)\cdot f_Y(y) \\ \frac{180}{101}x^2(x+y)^2 &\neq \frac{60}{101}\cdot\frac{6}{101}x^2\left(3x^2+3x+1\right)\left(10y^2+15y+6\right) \end{aligned}$$

Náhodné veličiny X a Y jsou závislé. Můžeme pokračovat v určování kovariance a koeficientu korelace podle definic (6.21) a (6.22). Kdyby byly veličiny nezávislé, nemělo by smysl dále určovat kovarianci a koeficient korelace.

$$\begin{aligned} \text{cov}(X;Y) &= E(X;Y)-E(X)\cdot E(Y)=\frac{201}{404}-\frac{81}{101}\cdot\frac{63}{101}=-0,00272 \\ \rho(X;Y) &= \frac{\text{cov}(X;Y)}{\sqrt{D(X)\cdot D(Y)}}=\frac{-0,00272}{\sqrt{0,0273\cdot 0,0713}}=-0,062 \end{aligned}$$

Veličiny X a Y jsou slabě závislé. Z koeficientu korelace jsme se navíc dozvěděli, že veličiny X a Y jsou slabě nepřímo lineární.

Úlohy:

1. Náhodný vektor **X** $= (X; Y)$ je dán pravděpodobnostní funkcí tabulka (6.7). Určete hodnotu K, střední hodnoty a rozptyly složek náhodného vektoru, kovarianci a koeficient korelace.

Tabulka 6.7: Hodnoty pravděpodobnostní funkce k úloze (1.)

$x \setminus y$	1	3
2	0,15	0,20
3	0,20	0,05
6	0,10	K

2. Náhodný vektor **X** $= (X; Y)$ je dán pravděpodobnostní funkcí tabulka (6.8). Určete distribuční funkci a $P(X = 2; Y = 1)$, $P(X < 3,2; Y < 0,7)$, $P(X > 2,1; Y > 1,1)$ a $P(1 < X \leq 3; 0,5 \leq Y \leq 2,5)$.

Tabulka 6.8: Hodnoty pravděpodobnostní funkce k úloze (2.)

$x \setminus y$	0	1	2
1	0,15	0,05	0,05
2	0,20	0,05	0,15
3	0,10	0	0,10
4	0	0,10	0,05

3. Náhodný vektor **X** $= (X; Y)$ je dán pravděpodobnostní funkcí tabulka (6.9). Určete střední hodnoty a rozptyly složek náhodného vektoru, kovarianci a koeficient korelace.
4. Náhodný vektor **X** $= (X; Y)$ je dán pravděpodobnostní funkcí tabulka (6.10). Určete podmíněné střední hodnoty a rozptyly.
5. Náhodný vektor **X** $= (X; Y)$ je dán pravděpodobnostní funkcí tabulka (6.11). Učete střední hodnoty a rozptyly složek vektoru, kovarianci, koeficient korelace, podmíněnou střední hodnotu veličiny X, když $y = 2$ a podmíněný rozptyl veličiny Y, když $x = 0$.

Tabulka 6.9: Hodnoty pravděpodobnostní funkce k úloze (3.)

$x \setminus y$	3	5	7	9
0	0,02	0,03	0,04	0,01
2	0,10	0,05	0	0,07
4	0,09	0,13	0	0,01
6	0	0,25	0,17	0,03

Tabulka 6.10: Hodnoty pravděpodobnostní funkce k úloze (4.)

$x \setminus y$	-2	-1	0
0	0,23	0,12	0,04
1	0,13	0,05	0,10
2	0,09	0,17	0,07

Tabulka 6.11: Hodnoty pravděpodobnostní funkce k úloze (5.)

$x \setminus y$	0	1	2	3
0	0,008	0,036	0,054	0,027
1	0,060	0,180	0,135	0
2	0,150	0,225	0	0
3	0,125	0	0	0

6. Náhodný vektor $\mathbf{X} = (X;Y)$ je dán pravděpodobnostní funkcí tabulka (6.12). Určete hodnotu K, distribuční funkci, střední hodnoty a rozptyl složek vektoru, kovarianci a koeficient korelace.

Tabulka 6.12: Hodnoty pravděpodobnostní funkce k úloze (6.)

$x \setminus y$	-2	2	6
-2	0,6	0	0
2	0	K	0
6	0	0	0,2

7. Náhodný vektor **X** $= (X;Y)$ má hustotu pravděpodobnosti:

$$\begin{aligned} f(x;y) &= K \cdot (x+y) \qquad 0 \leq x \leq 2 \qquad 0 \leq x \leq 3 \\ f(x;y) &= 0 \qquad \text{jinde.} \end{aligned}$$

   Určete hodnotu konstanty K, distribuční funkci, marginální hustotu pravděpodobnosti, podmíněné hustoty pravděpodobnosti, marginální distribuční funkce a $P(0{,}5 \leq x \leq 1; 1 \leq y \leq 3)$.

8. Náhodný vektor **X** $= (X;Y)$ má distribuční funkci:

$$\begin{aligned} F(x;y) &= \sin x \cdot \sin y \qquad 0 \leq x \leq \frac{\pi}{2} \qquad 0 \leq y \leq \frac{\pi}{2} \\ F(x;y) &= 0 \qquad \text{jinde.} \end{aligned}$$

   Určete hustotu pravděpodobnosti pro $0 \leq x \leq \frac{\pi}{2}$, $0 \leq y \leq \frac{\pi}{2}$, střední hodnoty a rozptyly veličin, kovarianci a koeficient korelace.

9. Náhodný vektor **X** $= (X;Y)$ má hustotu pravděpodobnosti:

$$\begin{aligned} f(x;y) &= \frac{K \cdot x^2}{y^2+1} \qquad 2 \leq x \leq 3 \qquad 0 \leq y \leq 1 \\ f(x;y) &= 0 \qquad \text{jinde.} \end{aligned}$$

   Určete hodnotu konstanty K, marginální hustoty pravděpodobnosti, střední hodnoty a rozptyly veličin, kovarianci a koeficient korelace.

Řešení:

1. $K = 0,30$; $E(X) = 3,85$; $E(Y) = 2,1$; $D(X) = 3,2275$; $D(Y) = 0,99$; $\text{cov}(X;Y) = 0.465$; $\rho(X;Y) = 0.26$
2. 0,05; 0,45; 0,15; 0,30

Tabulka 6.13: Distribuční funkce k úloze (2.)

$x \setminus y$	$(-\infty;0\rangle$	$(0;1\rangle$	$(1;2\rangle$	$(2;\infty)$
$(-\infty;1\rangle$	0	0	0	0
$(1;2\rangle$	0	0,15	0,20	0,25
$(2;3\rangle$	0	0,35	0,45	0,65
$(3;4\rangle$	0	0,45	0,55	0,85
$(4;\infty)$	0	0,45	0,65	1

3. $E(X) = 4,06$, $E(Y) = 5,48$, $D(X) = 4,2764$, $D(Y) = 3,3696$, $\text{cov}(X;Y) = 0,4112$, $\rho(X;Y) = 0,1083$.
4. $E(X/-2) = 0,6889$, $E(X/-1) = 1,1471$, $E(X/0) = 1,1429$, $E(Y/0) = -1,4872$, $E(Y/1) = -1,1071$, $E(Y/2) = -1,0606$, $D(X/-2) = 0,6143$, $D(X/-1) = 0,8313$, $D(X/0) = 0,5034$, $D(Y/0) = 0,4550$, $D(Y/1) = 0,8099$, $D(Y/2) = 0,4812$
5. $E(X) = 1,5$, $E(Y) = 0,9$, $D(X) = 0,75$, $D(Y) = 0,63$, $\text{cov}(X;Y) = -0,45$, $\rho(X;Y) = -0,65465$, $E(X/2) = 0,7143$, $D(Y/0) = 0,72$
6. $K = 0,2$, $E(X) = 3,2$, $E(Y) = 0,4$, $D(X) = 2,56$, $D(Y)\,10,24$, $\text{cov}(X;Y) = 5,12$, $\rho(X;Y) = 1$,

Tabulka 6.14: Distribuční funkce

$x \setminus y$	$(-\infty;-2\rangle$	$(-2;2\rangle$	$(2;6\rangle)$	$(6;\infty)$
$(-\infty;-2\rangle$	0	0	0	0
$(-2;2\rangle$	0	0,6	0,6	0,6
$(2;6\rangle$	0	0,6	0,8	0,8
$(6;\infty)$	0	0,6	0,8	1

7. $K = \frac{1}{15}$, $f_X(x) = \frac{x+y}{2\cdot(y+1)}$, $f_Y(y) = \frac{2}{3}\cdot\frac{x+y}{2x+3}$, $P(0,5 \leq x \leq 1; 1 \leq y \leq 3) = \frac{11}{90}$,

$$F(x;y) = 0 \qquad x \in (-\infty;0) \qquad y \in (-\infty;0)$$

$$\begin{aligned}
&= \frac{x^2+3x}{10} \qquad x \in \langle 0;2\rangle \qquad y \in (3;\infty) \\
&= \frac{y^2+2y}{15} \qquad x \in (2;\infty) \qquad y \in \langle 0;3\rangle \\
&= \frac{x^2y+xy^2}{30} \qquad x \in \langle 0;2\rangle \qquad y \in \langle 0;3\rangle \\
&= 0 \qquad x \in (2;\infty) \qquad y \in (3;\infty)
\end{aligned}$$

8. $f(x;y) = \cos x \cdot \cos y$, $E(X) = \frac{\pi}{2} - 1$, $E(Y) = \frac{\pi}{2} - 1$, $D(X) = \pi - 3$, $D(Y) = \pi - 3$, $\operatorname{cov}(X;Y) = 0$, $\rho(X;Y) = 0$
9. $K = \frac{12}{19}$, $f_X(x) = \frac{3x^2}{19}$, $f_Y(y) = \frac{4}{\pi\cdot(1+y^2)}$, $E(X) = \frac{195}{76}$, $E(Y) = \frac{2\ln 2}{\pi}$, $D(X) = \frac{633}{95} - \left(\frac{195}{76}\right)^2$, $D(Y) = \frac{4}{\pi} - 1 - \frac{4\cdot\ln^2 2}{\pi^2}$, $\operatorname{cov}(X;Y) = 0$, $\rho(X;Y) = 0$

Kapitola 7

Statistika

Získáváním, zpracováním a interpretací hromadných dat se zabývá vědní disciplína **matematická statistika**. Podle toho jakým způsobem budeme s daty pracovat rozdělujeme matematickou statistiku na:

Deskriptivní (popisná) statistika se zabývá efektivním získáváním ukazatelů, které poskytují souhrný obraz zkoumaného jevu.

Idnuktivní statistika se zabývá problémy zobecňování výsledků získaných popisem statistického souboru.

7.1 Základní statistické pojmy

Abychom mohli začít poznávat základy statistiky je potřeba zavést několik pojmů, bez kterých se neobejdeme.

Definice 7.1 (Statistický soubor) Statistický soubor je neprázdná konečná množina objektů se společnou vlastností.

Definice 7.2 (Prvek statistického souboru) Prvkem statistického souboru nazveme každý objekt statistického souboru.

Definice 7.3 (Rozsah statistického souboru) Rozsah statistického souboru je přirozené číslo N, které udává počet prvků statistického souboru.

Definice 7.4 (Znak (argument) statistického souboru) Znak statistického souboru je vlastnost, kteru je označne určitý počet prvků statistického souboru.

Statistické soubory dělíme podle toho kolik mu přiřadíme argumentů (znaků).

7.2 Statistický soubor s jedním argumentem

Mějme statistický soubor, u kterého zkoumáme právě jeden znak (teplota, síla, rychlost, ...). Každý znak nabývá určité míry úspěchu nebo neúspěchu. Podle toho jakou mírou úspěch nebo neúspěch hodnotíme přiřadíme znaku jeho hodnotu. Všechny hodnoty znaku budeme volit tak, abychom je byli schopni seřadit například podle velikosti $x_1 < x_2 < x_3 < \ldots < x_n$. Všimněme si významné podobnosti s náhodnou veličinou. Díky této podobnosti můžeme s hodnotami znaku statistického souboru pracovat jako s hodnotami náhodné veličiny. Tím si značně ulehčíme práci, protože statistický soubor budeme popisovat a charakterizovat podobným způsobem jako náhodnou veličinu.

Začneme tím, že zavedeme variační obor znaku statistického souboru.

Definice 7.5 (Variační obor) Variační obor argumentu X je interval $\langle x_1; x_n\rangle$.

Poznámka: Variační obor je analogií oboru hodnot náhodné veličiny.

Definice 7.6 (Rozpětí statistického souboru) Číslo R vyjadřuje míru rozpětí statistického souboru, pro které platí:

$$R = x_n - x_1. \tag{7.1}$$

Definice 7.7 (Absolutní četnost) Absolutní četnost f_i vyjdřuje míru, se kterou se hodnota x_i argumentu X vyskytuje ve statistickém souboru alespoň jednou.

Definice 7.8 (Statistická řada) Seřadíme-li hodnoty znaku podle velikosti a jeho četnosti řekneme, že statistický soubor tvoří statistickou řadu.

Definice 7.9 (Relativní četnost hodnoty argumentu x_i) Relativní četnost φ_i je hodnota pro níž platí:

$$\varphi_i = \frac{f_i}{N}, \tag{7.2}$$

kde f_i je absolutní četnost hodnoty x_i a N je rozsah statistického souboru.

Relativní četnost určuje, jak velkou část statistického souboru zabírá právě četnost f_i jedná se o analogii pravděpodobnosti hodnoty x_i náhodné veličiny.

Ještě nám chybí funkční zobrazení, které nás bude informovat o rozložení absolutní četnosti v celém statistickém souboru. Můžeme si všimnout, že právě relativní četnost se řídí výběrem znaku x_i. Tím se ukazuje, že relativní četnost je funkcí hodnoty znaku x_i. Funkce $\varphi_i = \varphi(x_i)$ je analogií pravděpodobnostní funkce diskrétní náhodné veličiny a platí pro ni:

1. $0 \leq \varphi_i \leq 1$
2. $\sum\limits_{i=1}^{n} = 1$

Funkci $\varphi_i = \varphi(x_i)$ nazveme empirická pravděpodobnostní funkce statistické řády.

Jestliže se nám povedlo určit pravděpodobností funkci statistické řady budeme chtít určit i distribuční funkci této statistické řady. Bude jí funkce $\Phi : \Phi(x)$, kterou nazveme empirická distributivní funkce funkce náhodné řady a její hodnoty zjistíme z relativní kumulativní četnosti.

Definice 7.10 (Relativní kumulativní četnost) Relativní kumulativní četnost Φi je číslo, pro které platí:

$$\Phi_i = \frac{\sum\limits_{k=1}^{i} f_k}{n}. \tag{7.3}$$

Pro empirickou distributivní funkci pak platí následující vlastnosti:

Vlastnosti 7.1

1. *Jestliže $x < x_i$, pak:*
$$\Phi(x_i) = 0$$
2. *Jestliže $x > x_i$, pak:*
$$\Phi(x_i) = 1$$

Díky analogii distributivní a pravděpodobnostní funkce s funkcemi náhodné veličiny je můžeme snadno graficky zobrazit.

7.2.1 Charakteristiky statistického znaku s jedním argumentem

Vyjdeme z podobnosti statistického souboru a náhodné veličiny a určíme stejné charakteristiky, které rozdělíme znovu do skupin.

Charakteristiky polohy

Definice 7.11 (Empirická střední hodnota) Empirická střední hodnota je hodnota, která splňuje rovnici:

$$\overline{x} = \frac{1}{n} \cdot \sum_{i=1}^{n} f_i \cdot x_i, \tag{7.4}$$

kde f_i je absolutní četnost zkoumaného znaku.

Empirická střední hodnota vyjadřuje stejně jako střední hodnota náhodné veličiny očekávanou hodnotu znaku.

Další charakteristiky polohy jsou modus, medián a p-kvantily. Modus vyjadřuje hodnotu, která má nejčastější zaspoutení ve statistickém souboru (hodnotu s největší četností). Medián určuje prostřední hodnotu souboru a p-kvantily dělí statistický soubor v poměru $p : (1 - p)$.

Definice 7.12 (Modus) Modus $\mathrm{Mod}\,(X)$ znaku statistického souboru je hodnota s největší absolutní četností $f_i = \max$..

Definice 7.13 (Medián) Medián $\mathrm{Me}\,(X)$ znaku statistického souboru je hodnota, která jej dělí na dva stejné podsoubory o stejném rozsahu souboru. Je-li počet prvků lichý $n = 2k + 1$ pak:

$$\mathrm{Me}\,(X) = x_k \tag{7.5}$$

Je-li počet prvků sudý $n = 2k$ pak:

$$\mathrm{Me}\,(X) = \frac{x_k + x_{k+1}}{2}. \tag{7.6}$$

Definice 7.14 (Empirický p-kvantil) Empirický p-kvantil je hodnota statistického znaku x_p, pro kterou platí, že $100p$ procent prvků je nejvýše rovno hodnotě x_p.

Významné p-kvantily

- Empirický kvartil - rozděluje statistický soubor na čtyři stejně velké podsoubory se stejným rozsahem.

- Empirický decil - rozděluje statistický soubor na deset stejně velkých podsouborů se stejným rozsahem.
- Empirický percentil - rozděluje statistický soubor na sto stejně velkých podsouborů se stejným rozsahem.

Charakteristiky variability

Při určování charakteristik variability také využijeme podobnosti s náhodnou veličinou. Podle náhodné veličiny určíme rozptyl, směrodatnou odchylku, průměrnou odchylku a variační koeficient. Význam těchto charakteristi však zůstává nezměněn i u statistického souboru.

Definice 7.15 (Empirický rozptyl) Empirický rozptyl znaku X od empirické střední hodnoty $\overline{x}$ je dán vztahem:

$$s_x^2 = \frac{1}{n} \cdot \sum_{i=1}^{n} f_i \cdot (x_i - \overline{x})^2 \tag{7.7}$$

Definice 7.16 (Empirická směrodatná odchylka) Empirická směrodatná odchylka znaku X od empirické střední hodnoty $\overline{x}$ je dána vztahem:

$$s_x = \sqrt{\frac{1}{n} \cdot \sum_{i=1}^{n} f_i \cdot (x_i - \overline{x})^2}. \tag{7.8}$$

Definice 7.17 (Empirická průměrná odchylka) Průměrná odchylka $\overline{\delta}$ od empirické střední hodnoty x je dána vztahem:

$$\overline{\delta} = \frac{1}{n} \cdot \sum_{i=1}^{n} f_i \left| x_i - \overline{x} \right|. \tag{7.9}$$

Empirická průměrná odchylka vyjadřuje průměrnou hodnotu všech absolutních odchylek od emprické střední hodnoty $\overline{x}$.

Definice 7.18 (Empirický variační koeficient) Empirický variační koeficient ν je dán vztahem:

$$\nu = \frac{s_x}{\overline{x}} = \frac{\sqrt{\frac{1}{n} \sum_{i=1}^{n} f_i \cdot (x_i - \overline{x})^2}}{\frac{1}{n} \cdot \sum\limits_{i=1}^{n} f_i \cdot x_i}. \tag{7.10}$$

Empirický variační koeficient vyjadřuje míru odchýlení směrodatné odchylky od střední hodnoty. Často se vyjadřuje v procentech:

$$\nu_p = \nu \cdot 100$$

Odvodíme i momenty statistického souboru, které dále využijeme k určení dalších charakteristik statistického souboru.

Definice 7.19 (Počáteční empirický moment k-tého řádu) Počáteční empirický moment k-tého řádu m_k je dán vztahem:

$$m_k = \frac{1}{n} \cdot \sum_{i=1}^{n} f_i \cdot x_i^k. \tag{7.11}$$

Definice 7.20 (Centrální empirický moment k-tého řádu) Centrální empirický moment k-tého řádu n_k je dán vztahem:

$$n_k = \frac{1}{n} \cdot \sum_{i=1}^{n} f_i \cdot (x_i - \overline{x})^k. \tag{7.12}$$

Definice 7.21 (Normovaný empirický moment k-tého řádu) Normovaný empirický moment k-tého řádu η je dán vztahem:

$$\eta = \frac{n_k}{s_x}. \tag{7.13}$$

Normované momenty použijeme při určení koeficientu šikmosti a špičatosti.

Definice 7.22 (Empirický koeficient šikmosti) Empirický koeficient šikmosti γ_1 je dán vztahem:

$$\gamma_1 = \eta_3 = \frac{n_3}{s^3} = \frac{\frac{1}{n} \cdot \sum\limits_{i=1}^{n} f_i \cdot (x_i - \overline{x})^3}{s^3}. \tag{7.14}$$

Stejně jako u náhodné veličiny získáváme informaci o míře nesymetrie statistického souboru.

Definice 7.23 (Empirický koeficient špičatosti) Empirický koeficient špičatosti γ_2 je dán vztahem:

$$\gamma_2 = \eta_4 - 3 = \frac{n_4}{s^4} - 3 = \frac{\frac{1}{n} \cdot \sum\limits_{i=1}^{n} f_i \, (x_i - \overline{x})^4}{s^4} - 3. \tag{7.15}$$

Připomeňme si způsob výpočtu koeficientů šikmosti a špičatosti z katipitoly náhodná veličina. Tam jsme spočítali tyto koeficienty pomocí počátečních momentů. Díky analogii můžeme použít tyto vztahy i na statitistický soubor.

$$\begin{aligned}\gamma_1 &= \frac{n_3}{s^3} = \frac{m_3 - 3m_1m_2 + m_1^3}{s^3}\\ \gamma_2 &= \frac{n_4}{s^4} - 3 = \frac{m_4 - 4m_1m_3 + m_1^2m_2 - 3m_1^4}{s^4} - 3\end{aligned}$$

K dalším charakteristikám charakteristiky patří kvartilové odchylky, které nejsou tolik ovlivněny extrémními hodnotami statistického souboru.

Kvartilová odchylka $Q = x_{75} - x_{25}$

Relativní kvartilová odchylka $Q_{\text{rel}} = \frac{x_{75}-x_{25}}{x_{75}+x_{25}}$

Kvartilová šikmost $\gamma_0 = \frac{x_{75}+x_{25}-2x_{50}}{x_{75}-x_{25}}$

■ **Řešený příklad 7.1** Sledovaný statistický znak nabyl těchto hotnot:
60, 80, 80, 100, 100, 100, 100, 120, 120, 150, 150, 160, 180, 200, 200, 200, 200, 200, 220, 250, 250, 250, 280, 300, 300, 300, 300, 350, 350, 360, 380, 400, 400, 400, 400, 420, 450, 500, 500, 500.
Určete rozptyl, modus, medián, kvartily, empirickou střední hodnotu, empirický rozptyl, empirický koeficient šikmosti, empirický koeficient špičatosti, kvartilovou odchylku, relativní kvartilovou odchylku a kvartilovou šikmost.
Řešení:

Zadaná data si sepíšeme do tabulky (7.1) s absolutní četností, relativní četností, kumulativní četností a relativní kumulativní četností.

Rozsah statistického souboru je:

$$R = x_{19} - x_1 = 500 - 60 = 440$$

Modus, medián a kvartily určíme podle definic (7.12), (7.13) a (7.14).

$$\begin{aligned}Mod\,(X) &= f_i = \text{max.} = 200\\ Me\,(X) &= \frac{x_{10} + x_{11}}{2} = 265\end{aligned}$$

Tabulka 7.1: Tabulka četností k příkladu (7.1)

x_i	60	80	100	120	150	160	180	200	220
f_i	1	2	4	2	2	1	1	5	1
φ_i	$\frac{1}{40}$	$\frac{2}{20}$	$\frac{1}{10}$	$\frac{1}{20}$	$\frac{1}{20}$	$\frac{1}{40}$	$\frac{1}{40}$	$\frac{1}{8}$	$\frac{1}{40}$
F_i	1	3	7	9	11	12	13	18	19
Φ_i	$\frac{1}{30}$	$\frac{3}{40}$	$\frac{7}{40}$	$\frac{9}{40}$	$\frac{11}{40}$	$\frac{12}{40}$	$\frac{13}{40}$	$\frac{18}{40}$	$\frac{19}{40}$

250	280	300	350	360	380	400	420	450	500
3	1	4	2	1	1	4	1	1	3
$\frac{3}{40}$	$\frac{1}{40}$	$\frac{1}{10}$	$\frac{1}{20}$	$\frac{1}{40}$	$\frac{1}{40}$	$\frac{1}{10}$	$\frac{1}{40}$	$\frac{1}{40}$	$\frac{3}{40}$
22	23	27	29	30	31	35	36	37	40
$\frac{22}{40}$	$\frac{23}{40}$	$\frac{27}{40}$	$\frac{29}{40}$	$\frac{30}{40}$	$\frac{31}{40}$	$\frac{35}{40}$	$\frac{36}{40}$	$\frac{37}{40}$	$\frac{40}{40}$

$$\begin{aligned} z_{0,25} &= 40 \cdot 0,25 + 0,5 = 10,5 \Rightarrow x_{0,25} = 150 \\ z_{0,5} &= 40 \cdot 0,5 + 0,5 = 20,5 \Rightarrow x_{0,5} = 220 \\ z_{0,75} &= 40 \cdot 0,75 + 0,5 = 30,5 \Rightarrow x_{0,75} = 380 \end{aligned}$$

Empirickou střední hodnotu a rozptyl určíme podle definic (7.11) a (7.15).

$$\begin{aligned} \overline{x} &= \frac{\sum\limits_{i=1}^{n} x_i f_i}{N} \\ \overline{x} &= \frac{10360}{40} = 259 \\ s_x^2 &= \frac{\sum\limits_{i=1}^{n} f_i \cdot (x_i - \overline{x})^2}{N} \\ s_x^2 &= 16679 \end{aligned}$$

Empirický koeficient šikmosti a špičatosti určíme podle definic (7.22) a (7.23).

$$\begin{aligned} \gamma_1 &= \frac{n_3}{s^3} = 0,266911 \\ \gamma_2 &= \frac{n_4}{s^4} - 3 = -1,06104 \end{aligned}$$

Kvartilovou odchylku, relativní kvartilovou dochylku a kvartilovou šikmost určíme:

$$\begin{aligned} Q &= x_{0,75} - x_{0,25} \\ Q_r &= \frac{Q}{x_{0,75} + x_{0,25}} = 0,43 \\ \gamma_0 &= \frac{x_{0,75} + x_{0,25} - 2 \cdot x_{0,5}}{Q} = 0,39 \end{aligned}$$

7.3 Statistický soubor se dvěma argumenty

Stejně jako můžeme vytvořit náhodný pokus, při němž získáme dvě náhodné veličiny, z kterých jsme vytvořili náhodný vektor. Můžeme získat i statistický soubor u něhož budeme vyhodnocovat dva znaky. Z analogie znaku statistického souboru a náhodné veličiny pak usuzujeme, že vlastnosti a charakteristiky náhodného vektoru budou předlohou pro určení vlastností a charakteristik statistického souboru se dvěma argumetny X a Y, jejichž hodnoty také dokážeme seřadit podle velikosti $x_1 < x_2 < x_3 < \ldots < x_n$ a $y_1 < y_2 < y_3 < \ldots < y_m$. Celkový počet všech dvojic argumetnů statistického souboru pak označíme N.

Definice 7.24 (Absolutní četnost $f_{i;j}$) Absolutní četnost vyjadřuje počet výskytů uspořádané dvojice $[x_i; y_j]$ ve statistickém souboru.

Absolutní četnost obvykle zapisujeme do podobné tabulky jako pravděpodobnostní funkci náhodného vektoru $\mathbf{X} = (X; Y)$, kterou nazýváme plošná nebo také kontingenční tabulka (7.2).

Tabulka 7.2: Plošná (kontingenční) tabulka

$x \setminus y$	y_1	y_2	$\ldots$	y_m
x_1	$f_{1;1}$	$f_{1;2}$	$\ldots$	$f_{1;m}$
x_2	$f_{2;1}$	$f_{2;2}$	$\ldots$	$f_{2;m}$
$\vdots$	$\vdots$	$\vdots$	$\ddots$	$\vdots$
x_n	$f_{n;1}$	$f_{n;2}$	$\ldots$	$f_{n;m}$

Definice 7.25 (Relativní četnost $\varphi_{i;j}$) Hodnotu $\varphi_{i;j}$, která odpovídá rovnici:

$$\varphi_{i;j} = \frac{f_{i;j}}{N}, \tag{7.16}$$

nazýváme relativní četnost statistického souboru.

Relativní četnost můžeme chápat jako analogii k pravděpodobnostní funkci náhodného vektoru. V takovém případě určíme také její marginální četnosti.

Definice 7.26 (Marginální četnosti hodnot x_i a y_j) Marginální četnost hodnoty x_i je dána vztahem:

$$M_X = \sum_{j=1}^{m} f_{i;j}. \tag{7.17}$$

Marginální četnost hodnoty y_j je dána vztahem:

$$N_Y = \sum_{i=1}^{n} f_{i;j}. \tag{7.18}$$

Marginální četnosti vyjadřují rozložení četnosti hodnot x_i nebo y_j bez omezení druhou hodnotou. Z definice navíc přímo vyplívá vlastnost:

Vlastnosti 7.2

1. *Pro všechny hodnoty:*

$$\sum_{i=1}^{n}\sum_{j=1}^{m} f_{i;j} = \sum_{j=1}^{m} N_j = \sum_{i=1}^{n} M_i = N \tag{7.19}$$

7.3.1 Charakteristiky statistického souboru se dvěma argumenty

Znovu vyjdeme z podobnosti statistického souboru a náhodného vektoru a proto určíme i podobné charakteristiky. Při určování nejpoužívanějších charakteristik bychom měli vycházet z momentových charakteristik, které jsou jak už jsme řekli analogií momentů náhodného vektoru a proto si dovolíme přeskočit jejich odvození a rovnou si uvedeme konkrétní příklady.

Definice 7.27 (Empirická střední hodnota argumentů) Empirická střední hodnota argumentu X je dána vztahem:

$$\overline{x} = \frac{1}{N}\sum_{i=1}^{n} M_i \cdot x_i \tag{7.20}$$

Empirická střední hodnota argumentu Y je dána vztahem:

$$\overline{y} = \frac{1}{N}\sum_{j=1}^{m} N_j \cdot y_j \tag{7.21}$$

Definice 7.28 (Empirická disperze (rozptyl) argumentu) Empirický rozptyl argumentu X je dán vztahem:

$$D(X) = s_x^2 = \frac{1}{N}\sum_{i=1}^{n} M_i (x_i - \overline{x})^2 = \frac{1}{N}\sum_{i=1}^{n} M_i x_i^2 - \overline{x}^2 \tag{7.22}$$

Empirický rozptyl argumentu Y je dán vztahem:

$$D(Y) = s_y^2 = \frac{1}{N}\sum_{j=1}^{m} N_j (y_j - \overline{y})^2 = \frac{1}{N}\sum_{j=1}^{m} N_j y_j^2 - \overline{y}^2 \tag{7.23}$$

Definice 7.29 (Empirická kovariance) Empirická kovariance argumentů X a Y je dána vztahem:

$$\begin{aligned} \mathrm{cov}_{x;y} &= \frac{1}{N}\sum_{i=1}^{n}\sum_{j=1}^{m} f_{i;j}(x_i - \overline{x})(y_j - \overline{y}) = && (7.24)\\ &= \frac{1}{N}\sum_{i=1}^{n}\sum_{j=1}^{m} f_{i;j} x_i y_j - \overline{x}\cdot\overline{y}. && (7.25)\end{aligned}$$

Definice 7.30 (Empirický koeficient korelace) Empirický koeficient korelace je dán vztahem:

$$\rho_{x;y} = \frac{\mathrm{cov}_{x;y}}{\sqrt{D(X)\cdot D(Y)}} = \frac{\mathrm{cov}_{x;y}}{s_x \cdot s_y}, \tag{7.26}$$

kde s_x a s_y je jsou emprické směrodatné odchylky.

Poznámka: Pokud je každá hodnota dvojice $[x_i; y_j]$ ve statistickém souboru zastoupena právě jednou, pak pro každou absolutní četnost $f_{i;j}$ platí:

$$f_{i;j} = 1$$

V takovém případě můžeme výpočet koeficientu korelace zjednodušit:

$$\rho_{x;y} = \frac{N\cdot\sum\limits_{i=1}^{n}\sum\limits_{j=1}^{m} x_i y_j - \sum\limits_{i=1}^{n} x_i \cdot \sum\limits_{j=1}^{m} y_j}{\sqrt{\left[N\cdot\sum\limits_{i=1}^{n} x_i^2 - \left(\sum\limits_{i=1}^{n} x_i\right)^2\right]\cdot\left[N\cdot\sum\limits_{j=1}^{m} y_j^2 - \left(\sum\limits_{j=1}^{m} y_j\right)^2\right]}}. \tag{7.27}$$

Empirický korelační koeficient charakterizuje míru lineární závislosti argumentů statistického souboru. Jeho vlastnosti jsou stejné jako u náhodného vektoru.

U statistických souborů navíc určujeme regresní přímky.

Definice 7.31 (Regresní přímka argumentu X a Y) Regresní přímka argumentu X vzhledem k argumentu Y je dána rovnicí:

$$x - \overline{x} = \frac{\mathrm{cov}_{x;y}}{s_y^2}\left(y - \overline{y}\right). \tag{7.28}$$

Regresní přímka argumentu Y vzhledem k argumentu X je dána rovnicí:

$$y - \overline{y} = \frac{\mathrm{cov}_{x;y}}{s_x^2}\left(x - \overline{x}\right). \tag{7.29}$$

Regresní přímky pomáhají v odhadu nově naměřených hodnot. Z rovnic regresních přímek si můžeme všimnout, že obě procházejí jedním bodem $[\overline{x}; \overline{y}]$, zároveň pokud je koeficient korelace 1 nebo -1, tak jsou argumenty X a Y lineárně závislé a regresní přímky jsou totožné. Jestliže koeficient korelace nabývá jiné hodnoty, jsou regresní přímky různoběžné a můžeme určit jejich odchylku α.

$$\mathrm{tg}\alpha = \frac{1 - \rho_{x;y}^2}{|\rho_{x;y}|} \cdot \frac{s_x \cdot s_y}{s_x^2 + s_y^2}. \tag{7.30}$$

Ukazuje se, že ne vždy je vhodné k regresi volit přímku. O tom jak vhodně volit regresní funkci se více dozvíme v kapitole regresní a korelační analýza.

■ **Řešený příklad 7.2** Při zjišťování hustoty tělesa ve tvaru kvádru bylo potřeba změřit všechny jeho míry (délka, šířka, hloubka). Určete plošnou tabulku četností statistického souboru, kde argument X je délka a Y šířka. Určete empirické střední hodnoty a rozptyly argumentů. Změřená data byla zaznamenána jako soubor uspořádaných trojic [délka; šířka; hloubka].

$[14,6; 4,98; 9,8]$, $[14,6; 4,99; 10,1]$, $[14,8; 4,98; 10]$, $[14,8; 4,98; 9,9]$, $[14,8; 4,98; 9,9]$, $[14,8; 4,98; 10,1]$, $[14,8; 5; 9,9]$, $[14,8; 5; 10]$, $[15; 4,99; 10,2]$, $[15; 5; 9,9]$, $[15; 5; 9,9]$, $[15; 5; 9,9]$, $[15; 5; 9,9]$, $[15; 5; 10]$, $[15; 5; 10]$, $[15; 5; 10,1]$, $[15; 5; 10,2]$, $[15; 5,01; 10]$, $[15; 5,01; 10]$, $[15; 5,01; 10,1]$, $[15; 5,01; 10,2]$, $[15,1; 5; 9,8]$, $[15,1; 5; 10,1]$, $[15,1; 5,01; 9,8]$, $[15,1; 5,01; 9,8]$, $[15,1; 5,01; 9,8]$, $[15,1; 5,01; 9,8]$, $[15,1; 5,01; 9,9]$, $[15,1; 5,01; 9,9]$, $[15,1; 5,01; 9,9]$, $[15,1; 5,01; 10]$, $[15,1; 5,01; 10]$, $[15,1; 5,01; 10]$, $[15,1; 5,01; 10,1]$, $[15,1; 5,02; 10,1]$, $[15,1; 5,02; 10,2]$, $[15,2; 5,01; 9,8]$, $[15,2; 5,01; 10]$, $[15,2; 5,02; 9,9]$, $[15,2; 5,02; 9,9]$, $[15,2; 5,02; 10]$, $[15,2; 5,02; 10]$, $[15,2; 5,02; 10,1]$, $[15,2; 5,02; 10,2]$, $[15,2; 5,02; 10,2]$, $[15,2; 5,03; 10,1]$, $[15,3; 5,02; 10]$, $[15,3; 5,03; 10]$, $[15,3; 5,03; 10,1]$, $[15,3; 5,03; 10,2]$.

Řešení:

Nejprve vytvoříme plošnou tabulku četností a doplníme ji o empirické marginální četnosti a součiny $x_i M_i$, $x_i^2 M_i$, $y_j N_j$, $y_j^2 N_j$, které posléze použijeme k určení empirických středních hodnot a rozptylů.

Tabulka 7.3: Rozšířená plošná tabulka četností k příkladu (7.2)

$x \setminus y$	4,98	4,99	5	5,01	5,02	5,03	M_i	$x_i M_i$	$x_i^2 M_i$
14,6	1	1	0	0	0	0	2	29,2	462,32
14,8	1	3	2	0	0	0	6	88,8	1314,24
15	0	1	8	4	0	0	13	195	2925
15,1	0	0	2	11	2	0	15	226,5	3420,15
15,2	0	0	0	2	7	1	10	152	2310,4
15,3	0	0	0	0	1	3	4	61,2	936,36
N_j	2	5	12	17	10	4	$\sum$ 50	-	-
$y_j N_j$	9,96	24,95	60	85,17	50,2	20,12	-	-	-
$y_j^2 N_j$	49,6	124,5	300	426,7	252	101,2	-	-	-

Potom co jsme určili plošnou tabulku četností (7.3) můžeme určit podle definic (7.27) a (7.28) empirické střední hodnoty a rozptyly.

$$
\begin{aligned}
\overline{x} &= \frac{\sum\limits_{i=1}^{n} x_i M_i}{N} = 15,054 \\
\overline{y} &= \frac{\sum\limits_{j=1}^{m} y_j N_j}{N} = 5,008 \\
D(X) &= \frac{\sum\limits_{i=1}^{n} x_i^2 M_i}{N} - \overline{x}^2 = 0,0265 \\
D(Y) &= \frac{\sum\limits_{j=1}^{m} y_j^2 N_j}{N} - \overline{y}^2 = 0,0001
\end{aligned}
$$

■ **Řešený příklad 7.3** Určete empirické střední hodnoty, rozptyly, kovarianci, koeficient korelace, rovnice regresních přímek a odchylku těchto přímek statistického souboru, který je dán plošnou tabulkou (7.4).

Tabulka 7.4: Plošná tabulka hodnot k příkladu (7.3)

$x \setminus y$	0,12	0,13	0,14	0,15	0,16	0,17	0,18
14,6	0	2	7	9	5	3	1
14,8	1	0	3	13	10	3	0
15	2	5	16	26	13	9	7
15,1	0	4	10	21	17	4	6
15,2	3	0	12	20	0	15	2
15,3	0	1	2	5	1	1	3

Řešení:

Nejprve si kontingenční tabulku rozšíříme o empirické marginální četnosti a počáteční momenty x_iM_i, $x_i^2M_i$, y_jN_j, $y_j^2N_j$, pomocí kterých určíme empirické střední hodnoty a rozptyly. Tabulka (7.5).

Tabulka 7.5: Doplněná plošná tabulka k příkladu (7.3)

$x \setminus y$	0,12	0,13	0,14	0,15	0,16	0,17	0,18	M_i	x_iM_i	$x_i^2M_i$
14,6	0	2	7	9	5	3	1	27	394,2	5755
14,8	1	0	3	13	10	3	0	30	444	6571
15	2	5	16	26	13	9	7	78	1170	17550
15,1	0	4	10	21	17	4	6	62	936,2	14137
15,2	3	0	12	20	0	15	2	52	790,4	12014
15,3	0	1	2	5	1	1	2	13	198,9	3043
N_j	6	12	50	94	46	35	19	$N = 262$	-	-
y_jN_j	0,72	1,56	7	14,1	7,36	5,95	3,42	-	-	-
$y_j^2N_j$	0,09	0,20	0,98	2,12	1,18	1,01	0,62	-	-	-

Poté určíme podle definice (7.27) empirické střední hodnoty a podle definice (7.28) empirické rozptyly.

$$\overline{x} = \frac{\sum\limits_{i=1}^{n} x_iM_i}{N} = 15,0141$$

$$\begin{aligned}\overline{y} &= \frac{\sum\limits_{j=1}^{m} y_j N_j}{N} = 0,1531 \\ s_x^2 &= \frac{\sum\limits_{i=1}^{n} x_i^2 M_i}{N} - \overline{x}^2 = 0,0356 \\ x_y^2 &= \frac{\sum\limits_{j=1}^{m} y_j^2 N_j}{N} - \overline{y}^2 = 0,0002\end{aligned}$$

Empirickou kovarianci určíme podle definice (7.29).

$$\text{cov}_{x;y} = \frac{\sum\limits_{i=1}^{n}\sum\limits_{j=1}^{m} f_{i;j} x_i y_j}{N} - \overline{x}\cdot\overline{y} = 0,000136$$

Empirický koeficient korelace určíme podle definice (7.30).

$$\rho_{x;y} = \frac{\text{cov}_{x;y}}{\sqrt{s_x^2 \cdot s_y^2}} = 0,052899$$

Z hodnoty kovariance jsme se dozvěděli, že argumenty X a Y jsou na sobě slabě závislé a z koeficientu korelace jsme zjistili, že jsou slabě přímo lineární.

Regresní přímky určíme podle definice (7.31).

$$\begin{aligned}x - \overline{x} &= \frac{\text{cov}_{x;y}}{s_y^2}(y - \overline{y}) \\ x - 15,0141 &= 0,735(y - 0,1531) \\ y - \overline{y} &= \frac{\text{cov}_{x;y}}{s_x^2}(x - \overline{x}) \\ y - 0,1531 &= 0,004\,(x - 15,0141)\end{aligned}$$

Vzhledem k tomu, že argumenty X a Y nejsou lineárně závislé a regresní přímky jsou různé určíme i jejich odchylku podle rovnice (7.30).

$$\begin{aligned}\text{tg}\alpha &= \frac{1-\rho_{x;y}^2}{|\rho_{x;y}|}\cdot\frac{s_x \cdot s_y}{s_x^2 + s_y^2} \\ \text{tg}\alpha &= 1,3501 \\ \alpha &= 53^\circ 28'\end{aligned}$$

Úlohy:

1. Určete rozsah, modus, medián, kvartily, empirickou střední hodnotu, empirický rozptyl, empirický koeficient šikmosti a špičatosti u statistického souboru daného tabulkou (7.6).

Tabulka 7.6: Tabulka k úloze (1.)

x_i	0	1	2	3	4
f_i	7	44	51	30	12

2. Určete empirickou střední hodnotu, empirický rozptyl, empirický koeficient šikmosti a špičatosti statistického souboru daného tabulkou (7.7).

Tabulka 7.7: Tabulka k úloze (2.)

x_i	390	410	430	450	470	490	510	530	550	570
f_i	7	10	14	22	25	12	3	3	2	2

3. Při měření IQ byla zjištěna tato data:
   71, 71, 78, 82, 82, 91, 92, 92, 97, 102, 102, 102, 103, 105, 105, 109, 111, 111, 111, 112, 112, 114, 114, 114, 116, 121, 122, 122, 126, 131, 137.
   Určete modus, medián, kvartyly, empirickou střední hodnotu, empirický rozptyl, empirickou směrodatnou odchylku, empirický
   koeficient šikmosti a empirický koeficient špičatosti.

4. Vypočtěte charakteristiky statistického souboru se dvěma argumenty. Tabulka (7.8). (Příklad zpracujte v tabulkovém editoru).

5. Vypočtěte číselné charakteristiky statistického souboru se dvěma argumenty, který je zadán tabulkou (7.9).

6. U studentů závěrečného ročníku byly zaznamenány výsledky státních zkoušek z fyziky, matematiky a z obhajoby bakalářské práce. Výsledky jsou uspořádány do trojic, pro které platí: první pozice = fyzika, druhá pozice = matematika, třetí pozice obhajoba. Vytvořte statistický soubor se dvěma argumenty, kde X je známka zkoušky z fyziky a Y výsledek zkoušky z matematiky. Určete jeho charakteristiky.
   111, 111, 112, 112, 113, 122, 122, 121, 122, 123, 124, 122, 121, 131, 132, 143, 212, 212,

Tabulka 7.8: Tabulka k úloze (4.)

$x \setminus y$	10	20	30	40	50	60	70	80	90	100
1	19	5	0	0	0	0	0	0	0	0
2	23	116	11	0	0	0	0	0	0	0
3	1	41	98	9	0	0	0	0	0	0
4	0	4	32	65	41	13	2	0	0	0
5	0	0	5	21	57	20	7	1	0	0
6	0	0	0	0	1	4	21	46	3	0
7	0	0	0	0	0	1	2	11	13	1
8	0	0	0	0	0	0	0	1	3	2

Tabulka 7.9: Tabulka k úloze (5.)

x	1,1	1,5	2	2,6	3,3	4,1	5	6,1	7,3	8,6	10	11,5	13,1	14,8
y	8	14	12	16	6	10	12	8	18	16	20	18	12	10

212, 213, 212, 212, 221, 224, 223, 222, 222, 222, 223, 222, 231, 233, 232, 232, 231, 231, 232, 233, 234, 232, 231, 233, 232, 234, 233, 233, 233, 233, 232, 232, 241, 242, 314, 312, 311, 313, 313, 313, 313, 322, 323, 333, 332, 332, 334, 333, 333, 333, 332, 334, 321, 324, 323, 322, 323, 323, 323, 323, 324, 323, 334, 333, 333, 333, 332, 333, 334, 333, 343, 343, 331, 332, 334, 333, 333, 333, 333, 333, 332, 333, 334, 332, 332, 333, 332, 331, 332, 333, 333, 333, 342, 343, 344, 343, 343, 343, 424, 434, 443, 432, 431, 432, 433, 442, 443, 443, 443, 443, 443, 442, 444, 444, 444, 444, 444.

7. Vytvořte statistický soubor se dvěma agumenty, kde X je známka zkoušky z fyziky a Y výsledek obhajoby bakalářské práce. Statistický soubor vytvořte z dat získaných v příkladu (6.).

Řešení:

1. $n = 114$, $R = 4$, $Mod\,(X) = 2$, $Me\,(X) = 2$, $x_{0,25} = 1$, $x_{0,5} = 2$, $x_{0,75} = 3$, $\overline{x} = 1,97$, $s_x^2 = 1,0409$, $\gamma_1 = 0,25172$, $\gamma_2 = -0,55425$.
2. $\overline{x} = 457,4$, $s_x^2 = 1493,24$, $\gamma_1 = 0,5179$, $\gamma_2 = 0,5517$
3. $Mod\,(X) = \{102; 111; 114\}$, $Me\,(X) = 122$, $x_{0,25} = 92$, $x_{0,5} = 109$, $x_{0,75} = 114$, $\overline{x} =$, $s_x^2 = 267,38$, $s_x = 16,35$, $\gamma_1 = -0,4008$, $\gamma_2 = -0,3558$
4. $\overline{x} = 3,78$, $\overline{y} = 40,19$, $s_x^2 = 2,4157$, $s_y^2 = 466,1084$, $s_x = 1,55$, $s_y = 21,59$, $\mathrm{cov}_{x;y} = 30,8975$, $\rho_{x;y} = 0,9208$
5. $\overline{x} = 12,86$, $\overline{y} = 6,5$, $s_x^2 = 16,9796$, $s_y^2 = 18,77$, $s_x = 4,12$, $s_y = 4,3$, $\mathrm{cov}_{x;y} = 0$, $\rho_{x;y} = 0$
6. $\overline{x} = 2,64$, $\overline{y} = 2,69$, $s_x^2 = 0,75$, $s_y^2 = 0,822$, $\mathrm{cov}_{x;y} = 0,354$, $\rho_{x;y} = 0,45$, regresní přímky: $y = 0,472x + 1,1445$, $x = 0,43y + 1,48$, $\alpha = 41°30'$
7. $\overline{x} = 2,64$, $\overline{y} = 2,607$, $s_x^2 = 0,75$, $s_y^2 = 0,787$, $\mathrm{cov}_{x;y} = 0,295$, $\rho_{x;y} = 0,384$, regresní přímky: $y = 0,393x + 1,571$, $x = 0,374 + 1,661$, $\alpha = 48°$

Kapitola 8

Regresní a korelační analýza

Regresní a korelační analýza slouží k tomu, abychom náhodně změřeným datům přiřadů určitou přibližnou funkční závislost těchto dat, ke které se více, či méně blíží. To znamená, že hledáme funkci, která nám pomůže lépe zpracovat naměřená data a odhadnout jakých hodnot budou nabývat další data.

Konkrétně pak platí:

Regresní analýza je metoda, kterou na základě již změřených dat, závislosti veličin, určujeme jaký by měl být výsledek dalšího měření.

Korelační analýza zkoumá vzájemnou závislost statistických údajů. Konkrétně korelační analýza určuje jak moc dobře jsme určili regresní funkci.

8.1 Regresní analýza

8.1.1 Obecné odvození regresní analýzy

Nyní si ukážeme jak postupujeme při regresní analýze. Tento postup posléze rozvedeme do speciálních případů regresní analýzy.

Pro zjednodušení úvah budeme dále pracovat jen se statistickým souborem dvojic naměřených dat ve tvaru $[x_i; f_i]$, kde $i = 1; 2; \ldots; n$. Hodnoty f_i ze všech dvojic $[x_i; f_i]$ budeme považovat za funkci nezávislé proměnné x_i:

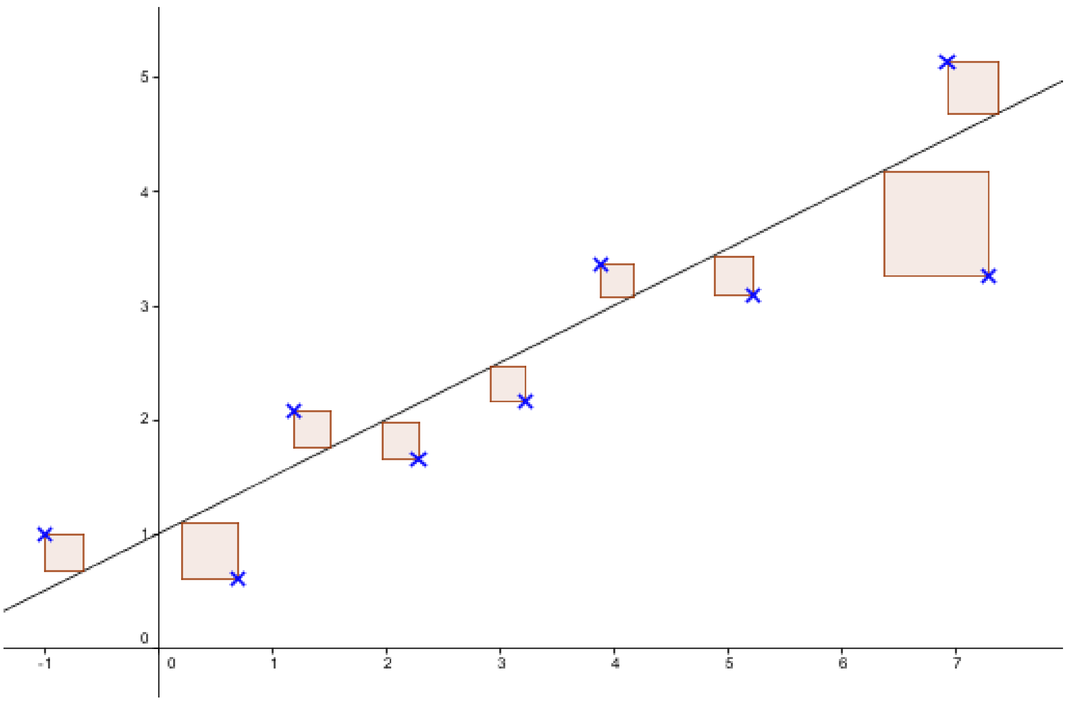
$$f_i = \varphi(x_i) + k, \tag{8.1}$$

kde funkce $\varphi(x_i)$ je jednoznačně určená proměnnou x a nazveme ji determinovaná složka. Reálná konstanta k je pak náhodná složka, kterou můžeme určit třeba jako chybu měření veličiny f_i.

Nyní se konečně dostáváme k regresní analýze, která se zabývá hledáním funkční závislosti $f_i = \varphi(x_i)$. Pokud je to možné změřený statistický soubor nejprve graficky znázorníme a podle toho se rozhodneme jaký typ regresní funkce zvolíme (lineární nebo nelineární regrese). Na základě zvolené regrese je už potřeba pouze určit koeficienty hledané regresní funkce. K určení těchto koeficientů nejčastěji používáme metodu nejmenších čtverců.

Metoda nejmenších čtverců

Název metody souvisí s tím, že hledáme koeficienty regresní funkce tak, aby odchylky změřených dat od funkčních hodnot regresní funkce byly co nejmešní, kde každá odchylka ρ_i^2 je druhá mocnina, která vyjadřuje obsah čtverce.



Obrázek 8.1: Metoda nejmenších čtverců

Pro nejvhodnější regresní funkci tedy platí vztah:

$$R(c_0; c_1; c_2; \ldots; c_m) = \sum_{i=1}^{n} \rho_i^2 = \sum_{i=1}^{n} (f_i - \varphi(x_i))^2 = \text{min.}, \tag{8.2}$$

kde funkce $R : R(c_0; c_1; c_2; \ldots; c_m)$ je funkce složená z odchylek a nazveme ji kriteriální funkce.

Všimněme si, že pokud je funkce R složená z odchylek (druhých mocnin) musí být nezáporná a tím pádem taková funkce musí mít minimum. Pro určení minima je potřeba nejprve nalézt stacionární bod, pro který platí:

$$\frac{\partial R}{\partial c_j} = 0, \tag{8.3}$$

kde $j = 0; 1; 2; \ldots; m$.

Při hledání stacionárních bodů získáme řešitelnou soustavu rovnic, jejímž řešením získáme koeficienty c_j regresní funkce $\varphi(x_i)$.

Jak jsme již uvedli, podle odhadu regresní funkce budeme dělit regresní analýzu do dvou skupin.

8.1.2 Lineární regrese

Ve většině případů statistického souboru se snažíme volit regresní funkci tak, aby byla lineární kombinací elementárních funkcí a platilo:

$$\varphi(x_i) = c_0\Phi_0(x_i) + c_1\Phi_1(x_i) + c_2\Phi_2(x_i) + \ldots + c_m\Phi_m(x_i) = \sum_{j=0}^{m} c_j \cdot \Phi_j(x_i),$$

kde pro koeficinety i a j platí $m < n$.

Dosazením zvolené regresní funkce do kriteriální funkce dostaneme:

$$R(c_0; c_1; c_2; \ldots; c_m) = \sum_{i=1}^{n} \left(\rho_i^2 = \sum_{i=1}^{n} f_i - \sum_{j=0}^{m} c_j \cdot \Phi_j(x_i) \right)^2 = \text{min.}$$

Najdeme stacionární body funkce R:

$$\frac{\partial R}{\partial c_k} = 0,$$

kde $k = 0; 1; 2; \ldots; m$.

Při hledání stacionárních bodů získáme řešitelnou soustavu rovnic:

$$\sum_{j=0}^{m} \left[c_j \cdot \sum_{i=1}^{n} \left(\Phi_j\left(x_i\right) \cdot \Phi_k\left(x_i\right) \right) \right] = \sum_{i=1}^{n} \left(f_i \cdot \Phi_k\left(x_i\right) \right), \tag{8.4}$$

kde $k = 0; 1; 2; \ldots; m$.

Z této soustavy rovnic získáme koeficienty lineární regresní funkce.

Řekli jsme, že lineární regresní funkci volíme tak, aby se skládala z elementárních funkcí. Některé z nich si ukážeme.

Lineární regresní funkce: pro statistický soubor $[x_i; f_i]$, kde $i = 1; 2; \ldots; n$.

$$\varphi\left(x\right) = c_0 + c_1 x \tag{8.5}$$

Soustava řešitelných rovnic má tvar:

$$c_0 \cdot n + c_1 \cdot \sum_{i=1}^{n} x_i = \sum_{i=1}^{n} f_i \tag{8.6}$$

$$c_0 \cdot \sum_{i=1}^{n} x_i + c_1 \cdot \sum_{i=1}^{n} x_i^2 = \sum_{i=1}^{n} \left(x_i \cdot f_i\right) \tag{8.7}$$

Kvadratická regresní funkce: pro statistický soubor $[x_i; f_i]$, kde $i = 1; 2; \ldots; n$.

$$\varphi\left(x\right) = c_0 + c_1 x + c_2 x^2 \tag{8.8}$$

Soustava řešitelných rovnic má tvar:

$$c_0 \cdot n + c_1 \cdot \sum_{i=1}^{n} x_i + c_2 \cdot \sum_{i=1}^{n} x_i^2 = \sum_{i=1}^{n} f_i \tag{8.9}$$

$$c_0 \cdot \sum_{i=1}^{n} x_i + c_1 \cdot \sum_{i=1}^{n} x_i^2 + c_2 \cdot \sum_{i=1}^{n} x_i^3 = \sum_{i=1}^{n} \left(x_i \cdot f_i\right) \tag{8.10}$$

$$c_0 \cdot \sum_{i=1}^{n} x_i^2 + c_1 \cdot \sum_{i=1}^{n} x_i^3 + c_2 \cdot \sum_{i=1}^{n} x_i^4 = \sum_{i=1}^{n} \left(x_i^2 \cdot f_i\right) \tag{8.11}$$

Hyperbolická regresní funkce: pro statistický soubor $[x_i; f_i]$, kde $i = 1; 2; \ldots; n$.

$$\varphi(x) = c_0 + \frac{c_1}{x} \tag{8.12}$$

Soustava řešitelných rovnic má tvar:

$$c_0 \cdot n + c_1 \cdot \sum_{i=1}^{n} \frac{1}{x_1} = \sum_{i=1}^{n} f_i \tag{8.13}$$

$$c_0 \cdot \sum_{i=1}^{n} \frac{1}{x_i} + c_1 \cdot \sum_{i=1}^{n} \frac{1}{x_i^2} = \sum_{i=1}^{n} \frac{f_i}{x_i} \tag{8.14}$$

Logaritmická regresní funkce pro statistický soubor $[x_i; f_i]$, kde $i = 1; 2; \ldots; n$.

$$\varphi(x) = c_0 + c_1 \cdot \log x \tag{8.15}$$

Soustava řešitelných rovnic má tvar:

$$c_0 \cdot n + c_1 \cdot \sum_{i=1}^{n} \log x_i = \sum_{i=1}^{n} f_i \tag{8.16}$$

$$c_0 \cdot \sum_{i=1}^{n} \log x_i + c_1 \cdot \sum_{i=1}^{n} (\log x_i)^2 = \sum_{i=1}^{n} (f_i \cdot \log x_i) \tag{8.17}$$

Vlastnosti 8.1 *V případě, že je k lineární regresi potřeba užít jinou polynomickou funkci než lineární nebo kvadratickou je řešitelná soustava rovnic pouze analogií právě lineární a kvadratické regresní funkce.*

8.1.3 Nelineární regrese

V případě, že všechny výše zmíněné lineární regrese selžou použijeme nelineární regresní funkci. O tom jak poznat jestli lineární regrese odpovídá statistickému souboru nebo ne, se dozvíme v části Korelační analýza. Vraťme se k nelineární regresi. Hledání koeficientů nelineární regrese je analogií hledání koeficientů lineární regrese. Při řešení získané soustavy rovnic nastane problém, protože tato soustava není obecně lineární a tím pádem je taková soustava hůře řešitelná. Proto se u nelineární regrese snažíme o její transformaci na lineární regresi. K transformaci využíváme nahrazení nelineární funkce linerání funkcí. Lépe tento výklad pochopíme ukázkou řešeného příkladu (8.4).

8.2 Korelační analýza

Neoddělitelnou součástí regresní analýzy je korelační analýza. Korelační analýza nás informuje o tom jak moc statistická data přiléhají k regresní funkci. Nejčastěji používaným měřítkem je index korelace, pro který platí:

$$I_{f;x} = \frac{s_{\varphi(x)}}{s_f} = \sqrt{\frac{\sum\limits_{i=1}^{n}\left(\varphi\left(x_i\right) - \overline{\varphi\left(x\right)}\right)^2}{\sum\limits_{i=1}^{n}\left(f_i - \overline{f}\right)^2}} = \sqrt{\frac{N \cdot \sum\limits_{i=1}^{n}\varphi^2\left(x_i\right) - \left(\sum\limits_{i=1}^{n}\varphi\left(x\right)\right)^2}{N \cdot \sum\limits_{i=1}^{n} f_i^2 - \left(\sum\limits_{i=1}^{n} f_i\right)^2}}, \tag{8.18}$$

kde s_f a $s_{\varphi(x)}$ jsou empirické směrodatné odchylky statistického souboru.

Index korelace může nabývat nejvýše hodnot $I_{x;y} \in \langle 0; 1\rangle$. Blíží-li se index korelace k hodnotě 1, vybrali jsme regresní funkci dobře (body statistického souboru dobře přiléhají k regresní funkci). V opačném případě se regresní funkce naprosto nehodí k dalšímu zpracování dat.

Použijeme-li logaritmickou regresní funkci počítáme koeficient korelace ne pomocí rovnice 8.18, ale takto:

$$I_{f;x} = 1 - \frac{s_v^2}{s_{\varphi(x)}^2}, \tag{8.19}$$

kde $s_f^2 = s_v^2 + s_{\varphi(x)}$.

Nakonec pokud najdeme pomocí transformace nelineární regresní funkci vyvstane otázka, z kterých dat budeme koeficient korelace počítat. Koeficient korelace v takovém případě můžeme počítat z transformované lineární regresní funkce.

■ **Řešený příklad 8.1** Při ověřování Ohmova zákona byla naměřena data uvedená v tabulce (8.1). Vystihněte závislost veličiny U[V] a I[A] vhodnou regresní funkcí, určete

Tabulka 8.1: Naměřená data k příkladu (8.1)

U	2	3	4	5	6	7	8
I	0,131	0,187	0,217	0,263	0,270	0,287	0,312
9	10	11	12	13	14	15	16
0,342	0,407	0,433	0,457	0,482	0,503	0,524	0,539

ji a určete index korelace.

Řešení:

Ze znalosti Ohmova zákona a voltampérové charakteristiky budeme předpokládat, že závislost veličin je lineární a platí:

$$I = c_0 + c_1 \cdot U$$

Nejprve vypočteme součty pro sestavení soustavy řešitelných rovnic lineární regresní funkce (8.6) a (8.7). Tyto rovnice sestavíme a vypočteme hodnoty koeficientů c_0 a c_1.

$$\begin{aligned}
n &= 15 \\
\sum_{i=1}^{15} U_i &= 135 \\
\sum_{i=1}^{15} U_i^2 &= 1495 \\
\sum_{i=1}^{15} I_i &= 5,353 \\
\sum_{i=1}^{15} U_i \cdot I_i &= 56,302
\end{aligned}$$

Sestavíme soustavu rovnic:

$$\begin{aligned}
c_0 \cdot n + c_1 \cdot \sum_{i=1}^{n} x_i &= \sum_{i=1}^{n} f_i \\
c_0 \cdot \sum_{i=1}^{n} x_i + c_1 \cdot \sum_{i=1}^{n} x_i^2 &= \sum_{i=1}^{n} f_i \cdot x_i
\end{aligned}$$

$$\begin{aligned}
15c_0 + 135c_1 &= 5,353 \\
135c_0 + 1495c_1 &= 56,302
\end{aligned}$$

Řešením této soustavy rovnic získáme koeficienty $c_0 = 0,0291$ a $c_1 = 0,0946$ lineární regresní funkce.

$$I_v = 0,0291 + 0,0946U$$

Regresní funkci, kterou jsme určili je ještě potřeba prověřit, jak dobře tato funkce vystihuje průběh statistického souboru. K tomu využijeme korelační index, který spočteme podle rovnice (8.18). Nejdříve je potřeba určit empirické směrodatné odchylky s_f

a s_φ. Směrodatnou odchylku s_f získáme z rozptylu veličiny I ze zadání a směrodatnou odchylku s_φ získáme z rozptylu vypočtené veličiny I_v, kterou jsme získali z lineární regresní funkce. K tomu si připravíme tabulku (8.2).

Tabulka 8.2: Tabulka vypočtených hodnot I_{v_i}

U	2	3	4	5	6	7	8
I_v	0,153	0,182	0,211	0,240	0,269	0,298	0,327
9	10	11	12	13	14	15	16
0,357	0,386	0,415	0,444	0,473	0,502	0,531	0,560

Určíme rozptyly:

$$\begin{aligned} s_I^2 &= n \cdot \sum_{i=1}^{n} I_i^2 - \left(\sum_{i=1}^{n} I_i\right)^2 = 3,585 \\ s_{I_v}^2 &= n \cdot \sum_{i=1}^{n} I_{v_i}^2 - \left(\sum_{i=1}^{n} I_{v_i}\right)^2 = 3,557 \end{aligned}$$

Po určení rozptylů je potřeba jen dosadit do vztahu pro index korelace:

$$I_{I_v;I} = \frac{s_{I_v}}{s_I} = \sqrt{\frac{3,557}{3,585}} = 0,996$$

Podle indexu korelace $I_{y;x}$, který se blíží k hodnotě 1, zvolená lineární regresní funkce $I_v = 0,0946 + 0,0291U$ hodnoty statistického souboru velice dobře vystihuje.

■ **Řešený příklad 8.2** Při bržděném volném pádu parašutisty byla každou sekundu zaznamenána jeho poloha. Výsledky měření jsou zadány v tabulce (8.3). Vystihněte

Tabulka 8.3: Naměřená data k příkladu (8.2)

t	0	1	2	3	4	5	6	7
y	1	0,996	0,989	0,974	0,948	0,910	0,861	0,800
-	8	9	10	11	12	13	14	15
-	0,728	0,643	0,545	0,431	0,308	0,171	0,022	0

závislost veličin y a t vhodnou regresní funkcí, určete ji a určete index korelace.

Řešení:

I když se jedná o bržděný volný pád, budeme předpokládat, že závislost veličin y a t je kvadratická platí:

$$y = c_0 + c_1 t + c_2 t^2$$

Stejně jako v předchozím příkladu si připravíme důležité součty k sestavení soustavy řešitelných rovnic kvadratické regresní funkce (8.9) (8.10) (8.11). Tyto rovnice sestavíme a vypočteme hodnoty koeficientů c_0, c_1 a c_2.

$$\begin{aligned}
n &= 16 \\
\sum_{i=1}^{16} t_i &= 120 \\
\sum_{i=1}^{16} t_i^2 &= 1240 \\
\sum_{i=1}^{16} t_i^3 &= 14400 \\
\sum_{i=1}^{16} t_i^4 &= 178312 \\
\sum_{i=1}^{16} y_i &= 10,326 \\
\sum_{i=1}^{16} y_i \cdot t_i &= 53,033 \\
\sum_{i=1}^{16} y_i \cdot t_i^2 &= 404,721
\end{aligned}$$

Sestavíme soustavu rovnic:

$$\begin{aligned}
c_0 \cdot n + c_1 \cdot \sum_{i=1}^{n} x_i + c_2 \cdot \sum_{i=1}^{n} x_i^2 &= \sum_{i=1}^{n} f_i \\
c_0 \cdot \sum_{i=1}^{n} x_i + c_1 \cdot \sum_{i=1}^{n} x_i^2 + c_2 \cdot \sum_{i=1}^{n} x_i^3 &= \sum_{i=1}^{n} x_i \cdot f_i \\
c_0 \cdot \sum_{i=1}^{n} x_i^2 + c_1 \cdot \sum_{i=1}^{n} x_i^3 + c_2 \cdot \sum_{i=1}^{n} x_i^4 &= \sum_{i=1}^{n} x_i^2 \cdot f_i
\end{aligned}$$

$$\begin{aligned}
16c_0 + 120c_1 + 1240c_2 &= 10,326 \\
120c_0 + 1240c_1 + 14400c_2 &= 53,033 \\
1240c_0 + 14400c_1 + 178312c_2 &= 404,721
\end{aligned}$$

Řešením této soustavy rovnic získáme koeficienty $c_0 = 1,003$, $c_1 = 0,0053$ a $c_2 = -0,0051$ kvadratické regresní funkce:

$$y_v = 1,003 + 0,0053t - 0,0051t^2$$

Regresní funkci, kterou jsme určili je ještě potřeba prověřit, jak dobře tato funkce vystihuje průběh statistického souboru. Stejně jako v předchozím příkladu spočteme index korelace.

$$\begin{aligned}
s_y^2 &= 30,618 \\
s_{y_v}^2 &= 29,948
\end{aligned}$$

Nyní dosadíme do indexu korelace:

$$I_{y_v;y} = \sqrt{\frac{29,948}{30,618}} = 0,978$$

Podle indexu korelace $I_{y_v;y}$, který se blíží hodnotě 1, zvolená kvadratická regresní funkce $y_v = 1,003 + 0,0053t - 0,0051t^2$ hodnoty statistického souboru velice dobře vystihuje.

■ **Řešený příklad 8.3** Při určování hodnoty odporu neznámého vodiče jsme zjistili jeho závislost na průřezu, která je dána tabulkou (8.4). Vystihněte závislost veličiny $R[\Omega]$ a

Tabulka 8.4: Naměřená data k příkladu (8.3)

S	0,01	0,05	0,10	0,15	0,20	0,25	0,30
R	23,00	4,35	2,52	1,79	1,04	0,91	0,75
0,35	0,40	0,45	0,50	0,55	0,60	0,65	0,70
0,63	0,51	0,48	0,43	0,40	0,35	0,34	0,30

$S[\mathrm{m}^2]$ vhodnou regresní funkcí, určete ji a určete index korelace.

Řešení:

Předpokládáme, že závislost odporu vodiče na průřezu je nepřímá. Proto vybereme hyperbolickou regresní funkci:

$$R = c_0 + \frac{c_1}{S}$$

Nejprve vypočteme součty pro sestavení soustavy řešitelných rovnic hyperbolické regresní funkce (8.13) a (8.14). Tyto rovnice sestavíme a vypočteme hodnoty koeficientů c_0 a c_1.

$$\begin{aligned} n &= 15 \\ \sum_{i=1}^{15} \frac{1}{S_i} &= 165,031 \\ \sum_{i=1}^{15} \frac{1}{S_i^2} &= 10630,4 \\ \sum_{i=1}^{15} f_i &= 37,8 \\ \sum_{i=1}^{15} \frac{f_i}{S_i} &= 2442,737 \end{aligned}$$

Sestavíme soustavu rovnic:

$$\begin{aligned} c_0 \cdot n + c_1 \cdot \sum_{i=1}^{n} \frac{1}{x_i} &= \sum_{i=1}^{n} f_i \\ c_0 \cdot \sum_{i=1}^{n} \frac{1}{x_i} + c_1 \cdot \sum_{i=1}^{n} \frac{1}{x_i^2} &= \sum_{i=1}^{n} \frac{f_i}{x_i} \\ 15c_0 + 165,031c_1 &= 37,8 \\ 165,031c_0 + 10630,4c_1 &= 2442,737 \end{aligned}$$

Řešením této soustavy rovnic získáme koeficienty $c_0 = -0,0098$ a $c_1 = 0,2299$ hyperbolické regresní funkce:

$$R_v = -0,0098 + \frac{0,2299}{S}$$

Regresní funkci, kterou jsme určili je ještě potřeba prověřit, jak dobře tato funkce vystihuje průběh statistického souboru. Stejně jako v předchozích příkladech spočteme index korelace.

$$\begin{aligned} s_R^2 &= 6993,804 \\ s_{R_v}^2 &= 6988,218 \end{aligned}$$

Nyní dosadíme do indexu korelace:

$$I_{R_v;R} = \sqrt{\frac{6988,218}{6993,804}} = 0,999$$

Podle indexu korelace $I_{R_v;R}$, který se blíží hodnotě 1, zvolená hyperbolická regresní funkce $R_v = -0,0098 + \frac{0,2299}{S}$ hodnoty statistického souboru velice dobře vystihuje.

■ **Řešený příklad 8.4** Při měření atmosférického tlaku byla zaznamenána data do tabulky (8.5).

Tabulka 8.5: Naměřená data k příladu (8.4)

h	0,01	0,05	0,10	0,15	0,20	0,25	0,3
p	101,213	58,842	45,502	17,653	15,172	9	3,721
-	0,4	0,5	0,6	0,7	0,8	0,9	1
-	2,236	0,520	0,275	0,046	0,051	0,010	0,003

Vystihněte závislost veličin p[kPa] a h[km] vhodnou regresní funkcí, určete ji a určete index korelace.

Řešení:

Ze znalosti závislosti atmosférického tlaku na změně výšky budeme předpokládat, že i závislost naměřených veličin je exponenciální. Proto odhadneme regresní funkci:

$$p = c_0 \cdot e^{-c_1 \cdot h}$$

Jak vidíme nejedná se o funkci, jejíž postup určení koeficientů jsme popsali v části lineární regrese. Předpokládáme, že se jedná o nelineární regresi. Budeme se snažit tuto zvolenou regresní funkci převést na jednu z lineárních regresních funkcí.

$$\begin{aligned} p &= c_0 \cdot e^{-c_1 \cdot h} \\ \ln p &= \ln\left(c_0 \cdot e^{-c_1 \cdot h} -\right) \\ \ln p &= \ln c_0 + \ln\left(e^{c_1 \cdot h}\right) \\ \ln p &= \ln c_0 - c_1 \cdot h \end{aligned}$$

Zavedeme substituci veličiny $\ln p = x$ a pro zjednodušení výpočtů navíc substituujeme $\ln c_0 = a$ a $-c_1 = b$.

$$x = a + b \cdot h$$

Tato regresní funkce už lineární je a její řešení jsme si už jednou ukázali v řešeném příkladu (8.1). Stejným způsobem vypočteme potřebné součty a potom sestavíme soustavu řešitelných rovnic.

$$\begin{aligned} n &= 14 \\ \sum_{i=1}^{14} h_i &= 5,96 \\ \sum_{i=1}^{14} h_i^2 &= 3,9376 \\ \sum_{i=1}^{14} x_i &= \sum_{i=1}^{14} \ln p_i = 4,001835 \\ \sum_{i=1}^{14} x_i \cdot h_i &= \sum_{i=1}^{14} h_i \cdot \ln p_i = -12,7199 \end{aligned}$$

Sestavíme soustavu rovnic:

$$\begin{aligned} 14a + 5,96b &= 4,001835 \\ 5,96a + 3,9376b &= -12,7199 \end{aligned}$$

Řešením této soustavy rovnic získáme koeficienty a a b, z kterých potom můžeme určit koeficienty c_0 a c_1.

$$\begin{aligned} a &= 4,6707 \\ b &= -10,3000 \\ \ln c_0 &= 4,6707 \\ c_0 &= 106,7729 \\ c_1 &= -b \\ c_1 &= 10,3000 \end{aligned}$$

Regresní funkce funkce má tvar:

$$p_v = 106,7729 \cdot e^{-10,3 \cdot h}$$

Index korelace můžeme vypočítat z transformované veličiny X nebo z veličiny p. Vzhledem k tomu, že jsme závislost veličiny p určili konkrétně index korelace spočteme pomocí této veličiny. Spočítáme rozptyly a pak dosadíme do indexu korelace. Ještě před tím, ale vypočteme hodnoty atmosférického tlaku p_v, které zapíšeme do tabulky (8.6).

Tabulka 8.6: Tabulka vypočtených hodnot p_{v_i}

h	0,01	0,05	0,10	0,15	0,20	0,25	0,30
p_v	96,323	63,797	38,119	22,776	13,609	8,131	4,858
-	0,40	0,50	0,60	0,70	0,80	0,90	1,00
-	1,735	0,619	0,221	0,079	0,0282	0,010	0,004

$$\begin{aligned} s_p^2 &= 165224,571 \\ s_{p_v}^2 &= 155721,358 \\ I_{p;h} &= \sqrt{\frac{155721,358}{165224,571}} = 0,9708 \end{aligned}$$

Podle indexu korelace $I_{p;h}$, který se blíží k hodnotě 1, zvolená nelineární regresní funkce $p_v = 106,7729 \cdot e^{-10,3 \cdot h}$ hodnoty statistického souboru velice dobře vystihuje.

Úlohy:

1. Při měření závislosti jistých dvou veličin X a Y. Výsledky měřeření jsou v tabulce (8.7). Rozhodněte, jestli závislost veličin X a Y lépe vystihuje kvadratická nebo

Tabulka 8.7: Tabulka naměřených hodnot k úloze (1.)

x	1	1	2	4	6
y	0	1	4	5	5

hyperbolická regresní funkce.

2. Při určování počtu vyzářených částic v daném čase byla naměřena data, která jsme zaznamenali do tabulky (8.8). Rozhodněte jestli lineární regresní funkce kvalitně

Tabulka 8.8: Tabulka naměřených hodnot k úloze (2.)

t	1	2	5	7	8	9
n	80	302	34	193	250	42

vystihuje naměřená data.

3. Při snižování tlaku v uzavřené nábodě (vývěva) byla sledována závislost počtu částic (počet částic je zaznamenám v miliardách) na čase v sekundách. Naměřená data jsou v tabulce (8.9). Sestavte pomocí naměřených dat lineární regresní funkci

Tabulka 8.9: Tabulka naměřených hodnot k úloze (3.)

t	5	12	20	26	29	38	65	126
n	19	17	18	17	17	15	14	7

a pomocí této funkce odhadněte za jak dlouho bude teoreticky v nádobě 0 částic.

4. Během laboratorního měření zrychlení automobilu jsme měřili okamžitou rychlost v km $\cdot$ s^{-1} a čas v sekundách. V ideálním případě by měla rychlost splňovat tuto závislost na čase:

$$v(t) = c_0 + c_1 t^2$$

Najděte koeficienty c_0 a c_1 regresní funkce a rozhodněte, jak dobře tato funkce vystihuje naměřená data. Zjistěte za jak dlouho dosáhne automobil rychlosti 1 km $\cdot$ s^{-1}. Naměřená data jsme zaznamenali do tabulky (8.10).

Tabulka 8.10: Tabulka naměřených hodnot k úloze (4.)

t	2,4	3,5	5	6,9	10
v	0,0141	0,0281	0,0562	0,1125	0,2250

5. Charakterizujte těsnost zvolené závislosti ve tvaru $y = c_0 + c_1 \cdot \log x$ mezi proměnými x a y. Vypočítejte index korelace. Naměřená data jsou v tabulce (8.11).

Tabulka 8.11: Tabulka naměřených hodnot k úloze (5.)

x	1	1	3	3	5	6	7	7
y	70	104	162	210	200	250	240	260

6. Zjišťovalo se, zda u souboru chlapců je závislost v počtu provedených shybů a kliků. Naměřená data jsou v tabulce (8.12). Určete jestli závislost mezi počtem

Tabulka 8.12: Tabulka naměřených hodnot k úloze (6.)

n	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15
s	1	3	2	0	5	6	1	4	3	5	6	2	1	1	8
k	10	15	15	0	40	25	7	31	30	35	41	10	14	9	64

shybů a kliků lépe vystihuje lineární nebo kvadratická regresní funkce.

7. Při měření veličiny X a Y byly zjištěny údaje tabulka (8.13). Určete koeficienty c_0 a

Tabulka 8.13: Tabulka naměřených hodnot k úloze (7.)

x	6,4	8,3	9,6	11,0	13,6	15,1
y	1,2	1,6	1,4	2,1	2,2	1,9
-	17,1	19,5	20,5	21,7	24,6	25,7
-	2,9	3,1	4,4	6,1	5,7	7,2

c_1 exponenciální regresní funkce ve tvaru $y = c_0 \cdot e^{c_1 \cdot x}$ a index korelace.

1. $y = -2,15 + 2,942x - 0,2914x^2$, $I = 0,99$, $y = 6,060 - \frac{5,565}{x}$, $I = 0,985$, lépe vystihuje závislost kvadratická regresní funkce
2. $n = 172,4 - 4,169t$, lineární regresní funkce nevystije kvalitně naměřená data.
3. $n = 19,3 - 0,095t$, $t_0 = 203,6\,[\mathrm{s}]$
4. $v = 0,00141 + 0,0022506t^2$, $I = 0,9996$, $t = 21$
5. $y = 88,32 + 191,54 \cdot \log x$, $I = 0,96$
6. $y = 6,6939 + 1,6463x$, $I = 0,927577$, $y = 3,7354 + 4,8667x + 0,243x^2$, $I = 0,93943$
7. $y = 0,655 \cdot e^{0,09034 \cdot x}$, $I = 0,96$

Literatura

[1] EADIE W. T., DRIJARD D, etc.: *Statistical methods in experimental physics*, Horth-Holland Publishing Company, Amsterdam, 1971. ISBN 0 7204 0239 5

[2] OTIPKA P., ŠMAJSTRLA V.: *Pravděpodobnost a statistika*, Vysoká škola báňská - Technická univerzita, Ostrava, 2006. ISBN 80-248-1194-4

[3] PAVELKA L., DOLEŽALOVÁ J.: *Pravděpodobnost a statistika*, Vysoká škola báňská - Technická univerzita, Ostrava, 1999. ISBN 80-7078-976-X

[4] NAGY I.: *Pravděpodobnost a matematická statistika cvičení*, Vydavatelství ČVUT, Praha, 2002. ISBN 80-01-02454-7

[5] JARUŠKOVÁ D.: *Pravděpodobnost a matematická statistika 12*, Vydavatelství ČVUT, Praha, 2006. ISBN 80-01-03427-5

[6] JARUŠKOVÁ D., HÁLA M.: *Pravděpodobnost a matematická statistika 12 - příklady*, Vydavatelství ČVUT, Praha, 1998. ISBN 80-01-01827-X

[7] FABIAN F., KLUIBER Z.: *Fyzika a pravděpodobnost*, Nakladatelství ARSCI, Praha, 2005, ISBN 80-86078-52-3

[8] PLOCKI A., TLUSTÝ P.: *Pravděpodobnost a statistika pro začátečníky a mírně pokročilé*, Prometheus, spol. s r. o., Praha, 2007. ISBN 978-80-7196-330-1

[9] ŠEDIVÝ P.: *Teplotní závislosti fyzikálních veličin. Studijní text pro soutěžící FO a ostatní zájemce o fyziku*, MAFY, Hradec Králové, 2001. ISBN 80-86148-53-X

[10] PETÁKOVÁ J.: *Matematika - příprava k maturitě a k přijímacím zkouškám na vysoké školy*, Prometheus, spol. s r. o., Praha, 1998. ISBN 80-7196-099-3

[11] DUPAČ V., HUŠKOVÁ M.: *Pravděpodobnost a matematická statistika*, Karolinum, Praha, 1999. ISBN 80-246-0009-9

[12] PROCHÁZKOVÁ E.: *Úvod do fyzikálního praktika*, Pedagogická fakulta, Jihočeská univerzita, České Budějovice, 1991.

[13] VYBÍRAL B.: *Zpracování dat fyzikálních měření. Studijní text pro soutěžící FO, studující fyziku na UHK a ostatní zájemce o fyziku*, MAFY, Hradec Králové, 2002. ISBN 80-86148-54-8