Střední průmyslová škola sdělovací techniky
Panská 3
Praha 1

Kombinatorika, pravděpodobnost a statistika

učební text

Jaroslav Reichl

OBSAH

1. KOMBINATORIKA

Kombinatorika se od všech matematických disciplín, s nimiž se studenti seznamují (geometrie, analytická geometrie, diferenciální a integrální počet, …), zásadním způsobem liší: zabývá se pouze konečnými množinami a jejich vlastnostmi.

To znamená, že se zde nesetkáme s „nekonečny“.

Další odlišností je to, že často není možné ověřit si správnost výsledku, k němuž jsme při řešení určité kombinatorické úlohy dospěli. Prakticky totiž není možné vypsat např. všechny možné způsoby losování Sportky, není možné vypsat všechny číselné kombinace hesla na zámku trezoru, … pro jejich velký počet. Řešící je odkázán jen a pouze na svůj vlastní úsudek, který se v kombinatorice podstatně odlišuje od úsudku v ostatních matematických disciplínách.

1.1 Základní kombinatorická pravidla

K řešení velké části kombinatorických úloh vystačíme se dvěma jednoduchými pravidly. Je možné, že tato pravidla již byla při řešení jiných typů úloh používána, aniž bychom věděli, že se jedná právě o kombinatorická pravidla.

Použití těchto pravidel ukážeme na úvodní úloze.

Úvodní úloha: Určete počet všech přirozených dvojciferných čísel, v jejichž dekadickém zápisu se každá číslice vyskytuje nejvýše jednou.

Řešení: Uvedeme dva způsoby řešení, které budou reprezentovat dále definovaná kombinatorická pravidla.

1. způsob řešení: Dvojciferné číslo je zapsáno pomocí dvou cifer vybraných z množiny 0, 1, …, 9. Obecně můžeme libovolnou cifru umístěnou na pozici jednotek kombinovat s libovolnou cifrou umístěnou na pozici desítek daného čísla. V případě, že hledáme dvojciferná čísla, nesmí být na pozici desítek cifra 0. Na pozici desítek tedy můžeme umístit jednu z cifer 1, 2, …, 9 – máme tedy 9 možností výběru. Na pozici jednotek můžeme umístit pouze takovou cifru, která již není na pozici desítek. Na pozici jednotek ale už můžeme umístit nulu. Máme tedy celkem 9 možností výběru cifry, kterou umístíme na pozici jednotek. Vzhledem k tomu, že libovolnou cifru umístěnou na pozici desítek můžeme kombinovat s libovolnou cifrou umístěnou na pozici jednotek, dostáváme celkem 9.9 = 81 možností.

2. způsob řešení: Přirozených dvojciferných čísel je celkem 90 (od 1 do 99 máme 99 čísel a z toho 9 je jednociferných). V devíti z nich se jejich cifry opakují (11, 22, …, 99). Proto počet dvojciferných přirozených čísel, v nichž se jejich cifry neopakují, je 90 – 9 = 81.

Na základě postupů uvedených v této úloze můžeme definovat obecně dvě kombinatorická pravidla.

KOMBINATORICKÉ PRAVIDLO SOUČINU:

POČET VŠECH USPOŘÁDANÝCH k-TIC, JEJICHŽ PRVNÍ ČLEN LZE VYBRAT n_1 ZPŮSOBY, DRUHÝ ČLEN PO VÝBĚRU PRVNÍHO ČLENU n_2 ZPŮSOBY, … AŽ k-TÝ ČLEN PO VÝBĚRU VŠECH PŘEDCHÁZEJÍCÍCH ČLENŮ n_k ZPŮSOBY, JE ROVEN:

$$n_1 \cdot n_2 \cdot \ldots \cdot n_k . \tag{1}$$

KOMBINATORICKÉ PRAVIDLO SOUČTU:

JSOU-LI A_1, A_2, …, A_k KONEČNÉ MNOŽINY, KTERÉ MAJÍ PO ŘADĚ p_1, p_2, …, p_k PRVKŮ, A JSOU-LI KAŽDÉ DVĚ MNOŽINY NAVZÁJEM DISJUNKTNÍ, PAK POČET PRVKŮ MNOŽINY $A_1 \cup A_2 \cup \ldots \cup A_k$ JE ROVEN

$$p_1 + p_2 + \ldots + p_k . \tag{2}$$

V úvodní úloze je první způsob řešení proveden na základě kombinatorického pravidla součinu, druhý způsob řešení na základě kombinatorického pravidla součtu.

1.2 Variace

V kombinatorice se často setkáváme s k-člennými skupinami prvků utvořenými z daných n prvků tak, že v nich záleží na pořadí a žádný z daných prvků se v nich neopakuje.

Např.: počet způsobů rozdělení tří medailí mezi 8 finalistů běhu; počet způsobů rozdělení funkcí starosty, místostarosty, jednatele a pokladníka v 9-ti členném zastupitelstvu; … Vždy se vlastně ptáme na to, kolika způsoby je možné z daného počtu n prvků množiny (atleti, zastupitelé, …) vytvořit uspořádanou k-tici (medailové pozice, funkce v zastupitelstvu, …).

k-ČLENNÁ VARIACE Z n PRVKŮ (RESP. VARIACE k-TÉ TŘÍDY Z n PRVKŮ) JE USPOŘÁDANÁ k-TICE SESTAVENÁ Z TĚCHTO PRVKŮ TAK, ŽE KAŽDÝ SE V NÍ VYSKYTUJE NEJVÝŠE JEDNOU.

Nyní určíme počet všech k-členných variací z n prvků, který se značí symbolem $V(k,n) = V_k(n)$.

Je dáno n vzájemně různých prvků a přirozené číslo $k \leq n$. Pro výběr prvního prvku uspořádané k-tice je n možností, pro výběr druhého prvku po výběru prvního už je ale jen $n-1$ možností, protože na druhém místě již nemůže být prvek, který jsme vybrali na místo první. Po výběru prvních dvou členů je pro výběr třetího členu už jen $n-2$ možností, … Pro výběr k-tého členu po výběru všech předcházejících je právě $n-(k-1)$ možností. Podle kombinatorického pravidla součinu (viz kapitola 1.1) vyjádřeného vztahem (1) tedy pro počet všech uspořádaných k-tic dostáváme: $n\cdot(n-1)\cdot(n-2)\cdot\ldots\cdot(n-(k-1)) = n\cdot(n-1)\cdot(n-2)\cdot\ldots\cdot(n-k+1)$.

VĚTA: POČET $V_k(n)$ VŠECH k-ČLENNÝCH VARIACÍ Z n PRVKŮ JE ROVEN

$$V_k(n) = n\cdot(n-1)\cdot(n-2)\cdot\ldots\cdot(n-k+1). \tag{3}$$

Je zřejmé, že $k \leq n$, neboť nemůžeme např. vybrat uspořádanou trojici ze dvou prvků tak, aby se prvky v dané trojici neopakovaly! Proto bychom také správně měli mluvit o k-členných variacích bez opakování. Většinou se ale dovětek „bez opakování“ nezdůrazňuje. Zdůrazňuje se dovětek „s opakováním“ u variací s opakováním (viz kapitola 1.8).

Vztah pro $V_k(n)$ je možné si pamatovat i takto: jedná se o součin k přirozených čísel, z nichž největší je n a každé další je o jedna menší.

Před řešením každé úlohy z kombinatoriky je nutné si uvědomit, o jaké k-tice se jedná: zda to jsou k-tice, v nichž **závisí na pořadí výběru** jednotlivých prvků, nebo **nezávisí na pořadí výběru** jednotlivých prvků. Pokud na pořadí výběru jednotlivých prvků závisí, jedná se o právě vysvětlené variace. Pokud na pořadí výběru jednotlivých prvků nezávisí, jedná se o kombinace (viz kapitola 1.4).

Budeme-li např. vybírat z celkového počtu 30 žáků jedné třídy čtveřici, která bude mít daný týden službu na tabuli a bude nosit do hodin pomůcky učitelů, bude se jednat o čtyřčlenné kombinace ze 30 prvků. Na pořadí žáků v dané čtveřici nezávisí, neboť žádný člen skupiny nemá žádné privilegované postavení nebo místo.

Budeme-li ze žáků stejné třídy vybírat čtveřici do třídní samosprávy (předseda, místopředseda, pokladník a nástěnkář), bude se jednat o čtyřčlenné variace ze 30 prvků. Na pořadí ve vybrané skupině závisí – každý žák totiž bude mít jinou funkci a s ní spojené i jiné pravomoci a povinnosti.

1.3 Permutace

V kapitole 1.2 byl vysvětlen pojem k-členné variace z n prvkové množiny, tj. uspořádané k-tice, v nichž se každý z daných n prvků vyskytuje nejvýše jednou. Pro přirozená čísla k a n přitom platí: $k \leq n$, neboť pro $k > n$ nelze z daných n prvků utvořit žádnou uspořádanou k-tici, v níž by se žádný prvek neopakoval. Nyní se budeme zabývat případem $k = n$, tj. uspořádanými n-ticemi sestavenými z n prvků. Takové skupiny se nazývají **permutace** (**pořadí**).

PERMUTACE Z n PRVKŮ JE KAŽDÁ n-ČLENNÁ VARIACE SESTAVENÁ Z TĚCHTO PRVKŮ.

Jinými slovy: permutace z n prvků je uspořádaná n-tice sestavená z těchto prvků tak, že každý se v ní vyskytuje právě jednou.

Nyní určíme počet všech permutací z n prvků, který se značí symbolem $P(n)$, poměrně snadno. Stačí si uvědomit, že permutace je zvláštním případem variace, v našem případě jde o n-členné variace z n prvkové množiny. Je tedy možné použít vztah (3) odvozený v kapitole 1.2 a uvědomit si, že platí $k = n$. Dostaneme tak výraz: $P(n) = V_n(n) = n\cdot(n-1)\cdot(n-2)\cdot\ldots\cdot 2\cdot 1$.

Tento výsledek je možné zjednodušit zavedením symbolu $n!$ (*n faktoriál*) pro součin všech přirozených čísel od jedné do n:

PRO KAŽDÉ PŘIROZENÉ ČÍSLO n DEFINUJEME:

$$n! = 1 \cdot 2 \cdot 3 \cdot \ldots \cdot (n-1) \cdot n = \prod_{i=1}^{n} i . \tag{4}$$

Symbol $\prod_{i=1}^{n} i$ je analogický jako symbol $\sum_{i=1}^{n} i$. Rozdíl spočívá v tom, že symbol $\sum_{i=1}^{n} i$ se používá jako zkrácený zápis pro sčítání, zatímco symbol $\prod_{i=1}^{n} i$ jako zkrácený zápis pro násobení.

Nyní můžeme pro počet permutací z *n* prvků vyslovit větu:

<u>VĚTA:</u> POČET $P(n)$ VŠECH PERMUTACÍ Z n PRVKŮ JE

$$P(n) = V_n(n) = n \cdot (n-1) \cdot (n-2) \cdot \ldots \cdot 2 \cdot 1 = n! . \tag{5}$$

Pomocí definičního vztahu faktoriálu (4) je nyní možné upravit i vztah (3) pro počet *k*-členných variací z *n* prvkové množiny, který byl odvozen v kapitole 1.2. Je možné postupně psát:

$$V_k(n) = n \cdot (n-1) \cdot (n-2) \cdot \ldots \cdot (n-k+1) =$$

$$= \frac{n \cdot (n-1) \cdot (n-2) \cdot \ldots \cdot (n-k+1) \cdot (n-k) \cdot (n-k-1) \cdot \ldots \cdot 3 \cdot 2 \cdot 1}{(n-k) \cdot (n-k-1) \cdot \ldots \cdot 3 \cdot 2 \cdot 1} = \frac{n!}{(n-k)!} .$$

Můžeme tedy psát

$$V_k(n) = n \cdot (n-1) \cdot (n-2) \cdot \ldots \cdot (n-k+1) = \frac{n!}{(n-k)!} . \tag{6}$$

Aby bylo možné použít výraz (6) i pro permutace (jakožto zvláštní případ variací), tj. aby výraz (6) platil i pro $k = n$, dodefinujeme

$$0! = 1 . \tag{7}$$

1.4 <u>Kombinace</u>

Až dosud jsme se zabývali skupinami vybranými z daných prvků, ve kterých záleželo na pořadí, tj. uspořádanými *k*-ticemi. Nyní se budeme zajímat o skupiny, v nichž na pořadí záležet nebude.

Takovými skupinami jsou např. počet všech vzájemných utkání v nohejbale, jestliže se turnaje zúčastní 7 mužstev; počet různých možností losování tahu Sportky; …

Při tvoření *k*-tic (nyní již neuspořádaných) budeme opět chtít (stejně jako v kapitolách 1.2 a 1.3), aby každá *k*-tice obsahovala každý z *n* prvků nejvýše jednou, tj. aby se v žádné *k*-tici žádný prvek neopakoval.

Skupiny tohoto typu se nazývají **kombinace**, přesněji *k*-členné kombinace z *n* prvků (resp. kombinace *k*-té třídy z *n* prvků).

k-ČLENNÁ KOMBINACE Z n PRVKŮ (RESP. KOMBINACE k-TÉ TŘÍDY Z n PRVKŮ) JE NEUSPOŘÁDANÁ k-TICE SESTAVENÁ Z TĚCHTO PRVKŮ TAK, ŽE KAŽDÝ SE V NÍ VYSKYTUJE NEJVÝŠE JEDNOU.

Při výpisu všech kombinací *k*-té třídy z dané množiny o *n* prvcích, zjistíme, že se jedná vlastně o všechny *k*-prvkové podmnožiny z množiny *n* prvků.

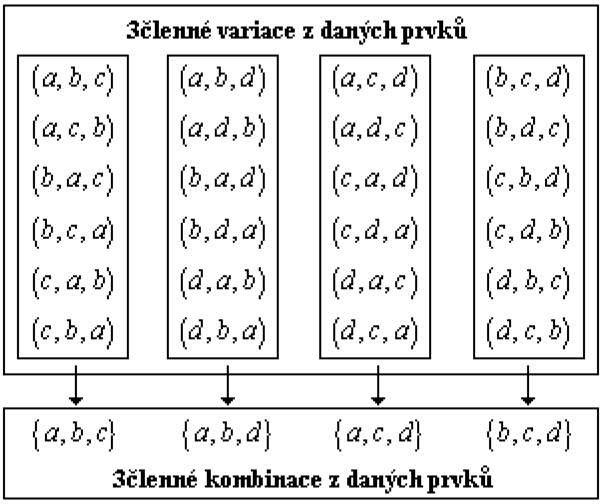
Např. 2členné kombinace z prvků A, B, C, D jsou: $\{A, B\}$, $\{A, C\}$, $\{A, D\}$, $\{B, C\}$, $\{B, D\}$, $\{C, D\}$.

Proto je možné *k*-členné kombinace z *n* prvků definovat i takto:

k-ČLENNÁ KOMBINACE Z n PRVKŮ JE k-PRVKOVÁ PODMNOŽINA MNOŽINY TĚMITO n PRVKY URČENÉ.

Při určování počtu všech *k*-členných kombinací z *n* prvků, které se označují $K(k, n)$ resp. $C_k(n)$, vyjdeme z již odvozeného výrazu pro počet všech *k*-členných variací z *n* prvků (viz kapitola 1.2). Všechny *k*-členné variace vytvořené z *n* prvků rozdělíme do skupin tak, aby všechny variace z téže skupiny se lišily pouze pořadím jednotlivých prvků a každé dvě variace z různých skupin se lišily alespoň v jednom prvku. Každá taková skupina obsahuje $k!$ uspořádaných *k*-tic, neboť *k* prvků lze uspořádat právě $k!$ způsoby. Nebude-li záležet na pořadí prvků v dané *k*-tici, pak všech $k!$ uspořádaných *k*-tic „splyne“ do jedné *k*-tice neuspořádané, tj. v jedinou *k*-člennou kombinaci z *n* prvků. To ale znamená, že počet skupin, do nichž jsme původně všechny *k*-členné variace rozdělili, je roven počtu *k*-členných kombinací z daných *n* prvků.

Situaci pro zadané prvky *a*, *b*, *c* a *d* a jejich tříčlenné kombinace schematicky zobrazuje obr. 1.



obr. 1

Vzhledem k tomu, že v každé skupině je $k!$ variací, platí:

$$V_{\mathrm{k}}(n) = k! \cdot C_{\mathrm{k}}(n). \tag{8}$$

Ze vztahu (8) tedy pro počet kombinací *k*-té třídy z *n* prvků plyne vztah $C_{\mathrm{k}}(n) = \dfrac{V_{\mathrm{k}}(n)}{k!}$. Po dosazení ze vztahu (6) postupně dostaneme vztah $C_{\mathrm{k}}(n) = \dfrac{V_{\mathrm{k}}(n)}{k!} = \dfrac{n!}{k! \cdot (n-k)!} = \dfrac{n \cdot (n-1) \cdot (n-2) \cdot \ldots \cdot (n-k+1) \cdot (n-k)!}{k! \cdot (n-k)!}$. Po dalším zjednodušení tedy získáme vztah

$$C_{\mathrm{k}}(n) = \frac{n \cdot (n-1) \cdot (n-2) \cdot \ldots \cdot (n-k+1)}{k!}. \tag{9}$$

Vzhledem k tomu, že v kapitole 1.3 byl dodefinován vztah (7), má zlomek $\dfrac{n!}{k! \cdot (n-k)!}$ smysl nejen pro $n, k \in \mathbb{N} : k \leq n$, ale i pro $k = 0$, dokonce i pro $n = 0$, $k = 0$. Pro tento zlomek, který má v kombinatorice důležitý význam, se používá speciální symbol a název.

PRO VŠECHNA CELÁ NEZÁPORNÁ ČÍSLA n A k, PRO KTERÁ PLATÍ $k \leq n$, PLATÍ

$$\binom{n}{k} = \frac{n!}{k! \cdot (n-k)!}. \tag{10}$$

TENTO SYMBOL SE NAZÝVÁ <u>KOMBINAČNÍ ČÍSLO</u>. ČTE SE „n NAD k“.

Pozor! V kombinačním čísle se nepíše zlomková čára, ačkoliv symbol zlomek připomíná!!!

Vztah (10) je definiční vztah. Pro praktické výpočty je ovšem možné postupovat jednodušeji pomocí „finty“. Dané kombinační číslo můžeme rozepsat také tak, že do jmenovatele zlomku definujícím kombinační číslo rozepíšeme faktoriál čísla *k* (toho čísla, které je v kombinačním čísle „dole“) **<u>VČETNĚ ČÍSLA 1!!!</u>** (Tedy ve jmenovateli bude *k* činitelů!) Do čitatele zlomku pak rozepisujeme faktoriál čísla *n* (toho „horního“ čísla) od *n* do jedné. Přitom vypíšeme pouze tolik čísel, kolik jich je napsáno ve jmenovateli (tedy *k* čísel).

Např. tedy můžeme psát: $\dbinom{12}{4} = \dfrac{12 \cdot 11 \cdot 10 \cdot 9}{4 \cdot 3 \cdot 2 \cdot 1} = \dfrac{11 \cdot 10 \cdot 9}{2} = 99 \cdot 5 = 495$.

Závěr, který jsme odvodili pro počet *k*-členných kombinací z *n* prvků v podobě vztahu (9), lze tedy s využitím definice (10) kombinačního čísla formulovat takto:

<u>VĚTA:</u> POČET $C_{\mathrm{k}}(n)$ VŠECH k-ČLENNÝCH KOMBINACÍ Z n PRVKŮ JE

$$C_{\mathrm{k}}(n) = \binom{n}{k}. \tag{11}$$

1.5 Vlastnosti kombinačních čísel

Kombinační čísla, s nimiž se setkáváme v kombinačních úlohách, mají řadu vlastností.

VĚTA: PRO VŠECHNA PŘIROZENÁ ČÍSLA n PLATÍ:

$$\binom{n}{0}=\binom{n}{n}=1 \tag{12}$$

A

$$\binom{n}{1}=n\,. \tag{13}$$

Důkaz: Důkaz platnosti vztahů (12) a (13) vyplývá přímo z definice (10) kombinačního čísla (viz kapitola 1.4). Vztahy (12) můžeme podle definice (10) psát ve tvaru: $\binom{n}{0}=\frac{n!}{0!\cdot(n-0)!}=\frac{n!}{1\cdot n!}=1$ a $\binom{n}{n}=\frac{n!}{n!\cdot(n-n)!}=\frac{n!}{n!\cdot 0!}=\frac{n!}{n!\cdot 1}=1$, vztah (13) pak ve tvaru $\binom{n}{1}=\frac{n!}{1!\cdot(n-1)!}=\frac{n\cdot(n-1)!}{1\cdot(n-1)!}=n$.

Na vztahy uvedené v této kapitole lze nahlížet i kombinatoricky – např. u vztahů (12) se ptát, kolik existuje n-členných kombinací z n prvků.

Další věta poskytuje návod, jak si v řadě případů při práci s kombinačními čísly zjednodušit výpočty.

VĚTA: PRO VŠECHNA CELÁ NEZÁPORNÁ ČÍSLA n A k TAKOVÁ, ŽE $k \leq n$ PLATÍ:

$$\binom{n}{n-k}=\binom{n}{k}. \tag{14}$$

Důkaz: Důkaz platnosti vztahu (14) vyplývá opět přímo z definice (10) kombinačního čísla. Můžeme totiž psát: $\binom{n}{n-k}=\frac{n!}{(n-k)!\cdot(n-(n-k))!}=\frac{n!}{(n-k)!\cdot k!}=\binom{n}{k}$.

Vztah (14) je vhodné použít v případě, že je nutné vyčíslit např. kombinační číslo $\binom{10}{7}$. Mohli bychom psát $\binom{10}{7}=\frac{10\cdot9\cdot8\cdot7\cdot6\cdot5\cdot4}{7\cdot6\cdot5\cdot4\cdot3\cdot2\cdot1}=\frac{10\cdot9\cdot8}{3\cdot2\cdot1}=120$, ale to je příliš pracné, časově náročné a může být zdroje zbytečných chyb. S využitím vztahu (14) můžeme psát $\binom{10}{7}=\binom{10}{3}=\frac{10\cdot9\cdot8}{3\cdot2\cdot1}=120$.

VĚTA: PRO VŠECHNA CELÁ NEZÁPORNÁ ČÍSLA n A k TAKOVÁ, ŽE $k<n$ PLATÍ:

$$\binom{n}{k}+\binom{n}{k+1}=\binom{n+1}{k+1}. \tag{15}$$

Důkaz: Důkaz platnosti vztahu (15) vyplývá opět přímo z definice (10) kombinačního čísla. Můžeme totiž postupně psát:

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

$$=\frac{n!}{k!\cdot(n-k-1)!}\cdot\left(\frac{1}{n-k}+\frac{1}{k+1}\right)=\frac{n!}{k!\cdot(n-(k+1))!}\cdot\frac{n+1}{(n-k)\cdot(k+1)}=\frac{(n+1)!}{(k+1)!\cdot(n-k)!}=\binom{n+1}{k+1}.$$

.

Předpoklad $k<n$ této věty zaručuje, že je $k+1\leq n$. V případě $k=n$ by totiž kombinační číslo $\binom{n}{k+1}=\binom{n}{n+1}$ nebylo definováno.

Uvedené věty je možné dokázat též kombinatorickou úvahou.

1.6 Pascalův trojúhelník

Vlastnosti kombinačních čísel uvedené v kapitole 1.5 můžeme ilustrovat pomocí schéma zobrazeného na obr. 2, které se nazývá **Pascalův trojúhelník**. Po vyčíslení kombinačních čísel v tomto schématu dostaneme schéma zobrazené na obr. 3.

$n = 0$	$\binom{0}{0}$										
$n = 1$	$\binom{1}{0}$	$\binom{1}{1}$									
$n = 2$	$\binom{2}{0}$	$\binom{2}{1}$	$\binom{2}{2}$								
$n = 3$	$\binom{3}{0}$	$\binom{3}{1}$	$\binom{3}{2}$	$\binom{3}{3}$							
$n = 4$	$\binom{4}{0}$	$\binom{4}{1}$	$\binom{4}{2}$	$\binom{4}{3}$	$\binom{4}{4}$						
$n = 5$	$\binom{5}{0}$	$\binom{5}{1}$	$\binom{5}{2}$	$\binom{5}{3}$	$\binom{5}{4}$	$\binom{5}{5}$					
$n = 6$	$\binom{6}{0}$	$\binom{6}{1}$	$\binom{6}{2}$	$\binom{6}{3}$	$\binom{6}{4}$	$\binom{6}{5}$	$\binom{6}{6}$				
$n = 7$	$\binom{7}{0}$	$\binom{7}{1}$	$\binom{7}{2}$	$\binom{7}{3}$	$\binom{7}{4}$	$\binom{7}{5}$	$\binom{7}{6}$	$\binom{7}{7}$			
$n = 8$	$\binom{8}{0}$	$\binom{8}{1}$	$\binom{8}{2}$	$\binom{8}{3}$	$\binom{8}{4}$	$\binom{8}{5}$	$\binom{8}{6}$	$\binom{8}{7}$	$\binom{8}{8}$		
$n = 9$	$\binom{9}{0}$	$\binom{9}{1}$	$\binom{9}{2}$	$\binom{9}{3}$	$\binom{9}{4}$	$\binom{9}{5}$	$\binom{9}{6}$	$\binom{9}{7}$	$\binom{9}{8}$	$\binom{9}{9}$	
$n = 10$	$\binom{10}{0}$	$\binom{10}{1}$	$\binom{10}{2}$	$\binom{10}{3}$	$\binom{10}{4}$	$\binom{10}{5}$	$\binom{10}{6}$	$\binom{10}{7}$	$\binom{10}{8}$	$\binom{10}{9}$	$\binom{10}{10}$
k:	0	1	2	3	4	5	6	7	8	9	10

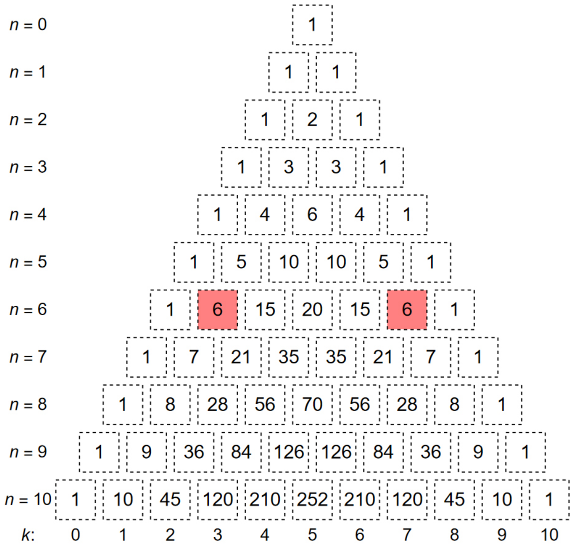
obr. 2

$n = 0$	1										
$n = 1$	1	1									
$n = 2$	1	2	1								
$n = 3$	1	3	3	1							
$n = 4$	1	4	6	4	1						
$n = 5$	1	5	10	10	5	1					
$n = 6$	1	6	15	20	15	6	1				
$n = 7$	1	7	21	35	35	21	7	1			
$n = 8$	1	8	28	56	70	56	28	8	1		
$n = 9$	1	9	36	84	126	126	84	36	9	1	
$n = 10$	1	10	45	120	210	252	210	120	45	10	1
k:	0	1	2	3	4	5	6	7	8	9	10

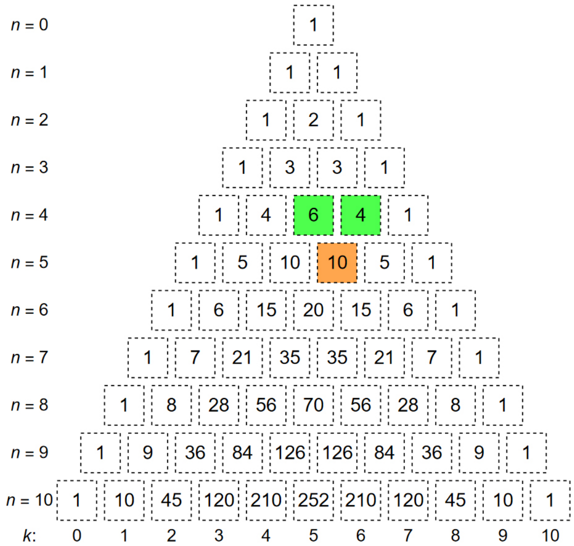
obr. 3

Na obr. 2 a obr. 3 je vidět symetrické rozdělení stejných čísel vzhledem k ose souměrnosti Pascalova trojúhelníku. Tato symetrie souvisí s platností vztahu (14) (viz kapitola 1.5), který je vizualizován na obr. 4.

Navíc jsou tato stejná čísla stejně „vzdálena“ od středu každého řádku.



obr. 4



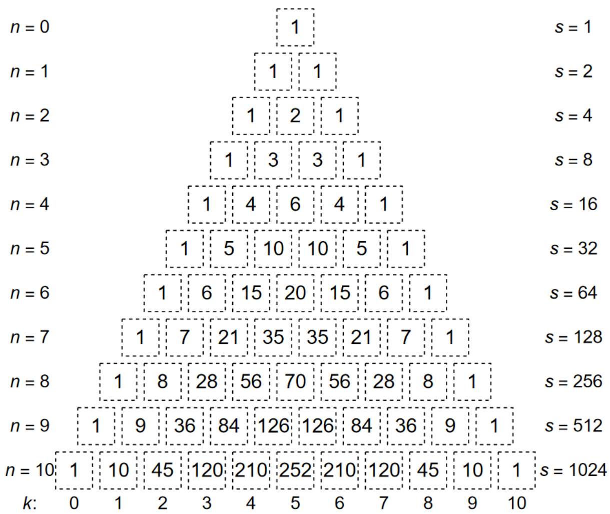
obr. 5

Součet libovolných dvou sousedních čísel v každém řádku Pascalova trojúhelníku je roven číslu, které se nachází v řádku následujícím „pod jejich středem“. Tato vlastnost čísel Pascalova trojúhelníku vyplývá z platnosti vztahu (15) a je vizualizována na obr. 5.

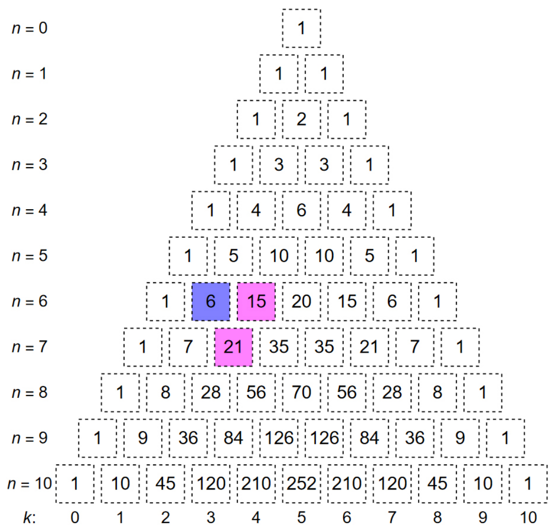
Další zajímavou vlastností, která souvisí s binomickou větou (viz kapitola 1.7) a kterou schematicky zobrazuje obr. 6, je součet kombinačních čísel na daném řádku Pascalova trojúhelníku. Platí vztah

$$\binom{n}{0}+\binom{n}{1}+\binom{n}{2}+\ldots+\binom{n}{n}=2^{n}. \tag{16}$$

Tento vztah je možné dokázat s využitím binomické věty (vztah (18) v kapitole 1.7). S využitím tohoto vztahu můžeme psát: $2^{n}=(1+1)^{n}=1\cdot\binom{n}{0}+1\cdot\binom{n}{1}+1\cdot\binom{n}{2}+\ldots+1\cdot\binom{n}{n}$, což je vztah (16).



obr. 6



obr. 7

Pouze pro čísla ve druhém sloupci Pascalova trojúhelníka platí další vztah:

$$\binom{n}{1}^{2}=\binom{n}{2}+\binom{n+1}{2}, \tag{17}$$

který je vizualizován na obr. 7.

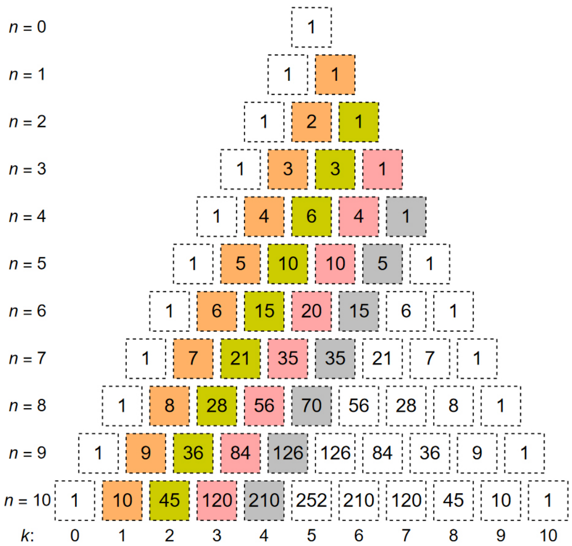
Důkaz vztahu (17) provedeme s využitím definičního vztahu kombinačního čísla (10). Pro levou stranu vztahu (17) můžeme psát $\binom{n}{1}^2 = n^2$. Pro pravou stranu téhož vztahu můžeme postupně psát $\binom{n}{2} + \binom{n+1}{2} = \frac{n \cdot (n-1)}{2} + \frac{(n+1) \cdot n}{2} = \frac{n^2 - n + n^2 + n}{2} = n^2$. Vztah tedy platí.

V Pascalově trojúhelníku lze „číst“ i různé typy čísel. S využitím obr. 8 můžeme tvrdit, že:

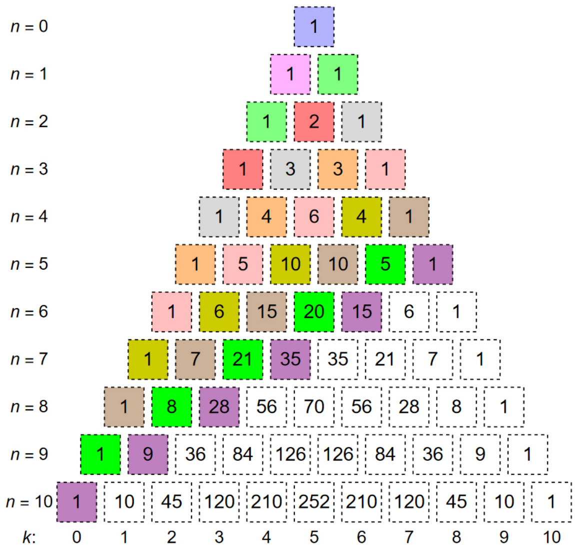
1. ve druhém sloupci trojúhelníku jsou zobrazena přirozená čísla;

Druhý sloupec je sloupec, ve kterém je $k = 1$.

2. ve třetím sloupci jsou tzv. trojúhelníková čísla, tj. taková čísla, že předměty v počtu daném těmito čísly lze v rovině vyskládat do tvaru rovnostranného trojúhelníku;
3. ve čtvrtém sloupci jsou tzv. pyramidální čísla, tj. taková čísla, že předměty v počtu daném těmito čísly lze v prostoru složit to tvaru pravidelného čtyřstěnu (tj. „do pyramidy“);
4. v pátém sloupci jsou tzv. pentatope čísla, tj. taková čísla, že předměty v počtu daném těmito čísly lze ve čtyřrozměrném prostoru složit do tvaru penatatope (čtyřrozměrná analogie pravidelného čtyřstěnu);
5. …



obr. 8



obr. 9

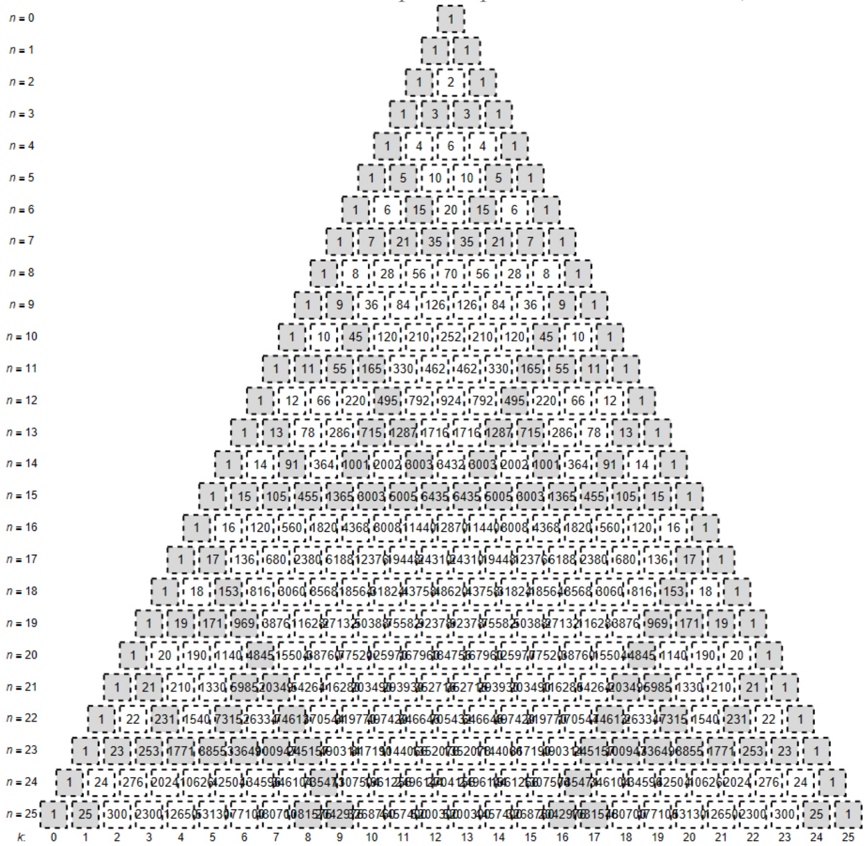
V Pascalově trojúhelníku lze nalézt i členy Fibonacciho posloupnosti (viz vizualizace na obr. 9): 1, 1, 2, 3, 5, 8, 13, 21, 34, 55, 89, …

Jednotlivá kombinační čísla vystupující v Pascalově trojúhelníku určují počet různých cest, kterými se lze z políčka čísla $\binom{0}{0}$ přes políčka v trojúhelníku výše umístěných čísel dostat na políčko daného čísla.

Např. na políčko s číslem $\binom{2}{1} = 2$ se lze dostat z políčka čísla $\binom{0}{0}$ dvěma způsoby: přes levou jedničku (číslo $\binom{1}{0}$) nebo přes pravou jedničku (přes číslo $\binom{1}{1}$). Analogicky to platí pro všechna čísla v trojúhelníku.

Tento způsob hledání počtu cest v Pascalově trojúhelníku odpovídá kombinacím, pomocí kterých byla kombinační čísla vystupující v Pascalově trojúhelníku zavedena.

Poslední vlastnost, kterou z mnoha dalších zmíníme, je vizualizována na obr. 10. Pokud v Pascalově trojúhelníku obarvíme pouze ta pole, ve kterých jsou lichá čísla, zobrazí se vzor typický pro tzv. Sierpińského trojúhelník. Tento fraktálový útvar popsal poprvé roku 1915 polský matematik Wacław Franciszek Sierpiński (1882 – 1969).



obr. 10

1.7 ***Binomická věta***

Srovnáním vzorců pro výpočet n-té mocniny dvojčlenu $a+b$ s Pascalovým trojúhelníkem (viz obr. 2 a obr. 3 v kapitole 1.8) zjistíme, že numerické koeficienty v mnohočlenu, který vznikne umocněním dvojčlenu $a+b$, odpovídají příslušným řádkům v Pascalově trojúhelníku.

Platí totiž:

$$(a+b)^0 = 1$$

$$(a+b)^1 = 1\cdot a + 1\cdot b$$

$$(a+b)^2 = 1\cdot a^2 + 2\cdot ab + 1\cdot b^2\,;$$

$$(a+b)^3 = 1\cdot a^3 + 3\cdot a^2b + 3\cdot ab^2 + 1\cdot b^3\,;\ \ldots$$

Hodnoty vyznačených koeficientů se shodují s čísly ve prvním, ve druhém, ve třetím, ve čtvrtém, … řádku Pascalova trojúhelníku - viz obr. 3.

Tyto úvahy můžeme zobecnit a vyslovit tzv. **binomickou větu**:

<u>BINOMICKÁ VĚTA:</u> PRO VŠECHNA REÁLNÁ ČÍSLA a A b A KAŽDÉ PŘIROZENÉ ČÍSLO n PLATÍ:

$$(a+b)^{n} = \binom{n}{0}a^{n} + \binom{n}{1}a^{n-1}b + \binom{n}{2}a^{n-2}b^{2} + \ldots + \binom{n}{k}a^{n-k}b^{k} + \ldots + \binom{n}{n-1}ab^{n-1} + \binom{n}{n}b^{n}. \tag{18}$$

Zjednodušeně lze psát vztah (18) ve tvaru:

$$(a+b)^{n} == \sum_{k=0}^{n}\binom{n}{k}a^{n-k}b^{k}\,. \tag{19}$$

Místo názvu kombinační číslo se u binomické věty používá termín **binomický koeficient**. Vyjádříme-li výraz $(a+b)^n$ pomocí binomické věty, říkáme, že jsme jej **rozvinuli** podle binomické věty nebo že jsme vytvořili **binomický rozvoj** výrazu $(a+b)^n$.

Důležité pro řadu úloh je si uvědomit, že *k*-tý člen binomického rozvoje má tvar: $\binom{n}{k-1}a^{n-(k-1)}b^{k-1}$.

> Pro řešení řady úloh je nutné si uvědomit, že číslování členů je ve srovnání „se spodním číslem“ kombinačního čísla vystupujícího v daném členu posunuté o jedničku!

<u>Důkaz (binomické věty):</u> Binomickou větu je možné dokázat matematickou indukcí nebo kombinatoricky. Rozvedeme nyní kombinatorický způsob důkazu. Nejprve se podíváme na konkrétní příklad $(a+b)^3 = aaa + aab + aba + baa + abb + bab + bba + bbb$. Na každý z uvedených sčítanců se lze dívat jako na uspořádanou trojici prvků sestavenou z prvků *a* a *b*. Sloučit lze ty sčítance, jimž odpovídají uspořádané trojice s týmž počtem prvků *a* a s týmž počtem prvků *b*. Počet trojic se dvěma prvky *a* a s jedním prvkem *b* je $P'(2,1) = \frac{3!}{2!\cdot 1!} = \binom{3}{1}$, počet trojic s jedním prvkem *a* a s dvěma prvky *b* je $P'(1,2) = \frac{3!}{1!\cdot 2!} = \binom{3}{2}$. Je tedy $(a+b)^3 = a^3 + \binom{3}{1}a^2b + \binom{3}{2}ab^2 + b^3$. (Symbolem P' je označen počet permutací s opakováním, které jsou detailně popsány v kapitole 1.9.)

Analogický postup použijeme i v obecném případě $(a+b)^n$. Vynásobením těchto *n* dvojčlenů $a+b$ dostaneme $\binom{n}{1}$ součinů $a^{n-1}b$, neboť existuje $P'(n-1,1) = \binom{n}{1}$ uspořádaných *n*-tic sestavených z $n-1$ prvků *a* a jednoho prvku *b*; dostaneme $\binom{n}{2}$ součinů $a^{n-2}b^2$, neboť existuje $P'(n-2,2) = \binom{n}{1}$ uspořádaných *n*-tic sestavených z $n-2$ prvků *a* a dvou prvků *b*; …

Tento důkaz je i vysvětlením, kde se v binomické větě „objevila“ kombinační čísla.

Koeficienty jednotlivých členů binomického rozvoje jsou tvořeny kombinačními čísly z Pascalova trojúhelníka (viz kapitola 1.5).

1.8 <u>Variace s opakováním</u>

Na rozdíl od kapitol 1.2 až 1.4 nyní upustíme od požadavku, aby se v dané *k*-tici vytvořené z *n* prvků každý prvek opakoval nejvýše jednou. Naopak: připustíme, že se prvek v dané *k*-tici může opakovat až *k*-krát.

> Příkladem variace s opakováním je např. šestimístný PIN tvořený z číslic 0, 1, … , 9, přičemž se každá číslice v daném PINu může opakovat až 6krát.

<u>k-ČLENNÁ VARIACE S OPAKOVÁNÍM</u> Z n PRVKŮ JE USPOŘÁDANÁ k-TICE SESTAVENÁ Z TĚCHTO PRVKŮ TAK, ŽE KAŽDÝ SE V NÍ VYSKYTUJE NEJVÝŠE k-KRÁT.

Pro ilustraci uvedeme dvoučlenné variace s opakováním ze tří prvků *a*, *b* a *c*: (a,a), (a,b), (a,c), (b,a), (b,b), (b,c), (c,a), (c,b), (c,c). Oproti variacím bez opakování se tento výčet liší tím, že jsou zde navíc dvojice (a,a), (b,b) a (c,c). Další rozdíl oproti *k*-členným variacím bez opakování z *n* prvků, které existovaly jen pro $k \le n$, je ten, že *k*-členné variace s opakováním z *n* prvků existují i pro $k > n$.

> Principiálně můžeme mít i dvacetimístný PIN tvořený číslicemi 0, 1, … , 9, které v něm mohou opakovat až 20krát; může tedy existovat např. PIN tvořený dvaceti jedničkami.

Mějme nyní dáno *n* navzájem různých prvků a přirozené číslo *k*. Pro vytváření *k*-členných variací s opakováním z těchto *n* prvků předpokládejme, že každý z uvažovaných prvků je k dispozici v neomezeném počtu exemplářů. Proto je pro výběr každého členu uspořádané *k*-tice k dispozici právě *n* možností, který prvek vybrat na další pozici sestavované *k*-tice, a to nezávisle na tom, které prvky byly už vybrány na předcházející pozice.

> Vzhledem k tomu, že variace s opakováním se typicky používají při vytváření různých kódů, PINů, šifrovacích sekvencí, …, tak požadavek na „neomezené množství počtu exemplářů daného prvku“ je v případě (většinou) alfanumerických znaků určitě splněn.

Pro každý člen je tedy n možností výběru. Podle kombinatorického pravidla součinu (viz kapitola 1.1) je tedy počet těchto uspořádaných k-tic, v nichž se každý prvek může opakovat, roven součinu $\underbrace{n \cdot n \cdot n \cdot \ldots \cdot n}_{k-krát} = n^k$.

<u>VĚTA:</u> POČET $V_k'(n)$ VŠECH k-ČLENNÝCH VARIACÍ S OPAKOVÁNÍM Z n PRVKŮ JE ROVEN:

$$V_k'(n) = n^k . \tag{20}$$

1.9 <u>Permutace s opakováním</u>

V kapitole 1.3 jsme zkoumali permutace z n prvků, tj. uspořádané n-tice, v nichž byl každý z daných n prvků zastoupen právě jednou. Zkoumejme nyní uspořádané skupiny, v nichž je každý z daných n prvků a_1, a_2, a_3, …, a_n zastoupen alespoň jednou, a to v předem určeném počtu: k_1 - krát prvek a_1, k_2 - krát prvek a_2, k_3 - krát prvek a_3, …, k_n - krát prvek a_n .

Tyto skupiny se uplatní tam, kde je potřeba tvořit skupiny navzájem **NEROZLIŠITELNÝCH PŘEDMĚTŮ**: fixy z téže výrobní série lišící se pouze barvou, skupina lidí, mezi nimiž jsou jednovaječná dvojčata, sirky vysypané ze tří krabiček, v nichž se sirky liší pouze barvou hlavičky, …

<u>PERMUTACE S OPAKOVÁNÍM</u> Z n PRVKŮ JE USPOŘÁDANÁ k-TICE SESTAVENÁ Z TĚCHTO PRVKŮ TAK, ŽE KAŽDÝ SE V NÍ VYSKYTUJE ASPOŇ JEDNOU.

Zřejmě platí $k_1 + k_2 + k_3 + \ldots + k_n \geq n$.

V případě, že $k_1 = k_2 = k_3 = \ldots = k_n = 1$, pak $k_1 + k_2 + k_3 + \ldots + k_n = n$, takže se jedná o uspořádané n-tice, v nichž je každý z n prvků zastoupen právě jednou a jedná se tedy o permutace bez opakování. Permutace bez opakování jsou tedy zvláštním případem permutací s opakováním.

Počet všech permutací s opakováním z n prvků, v nichž se jednotlivé prvky opakují k_1 - krát, k_2 - krát, k_3 - krát, …, k_n - krát, se označuje $P'(k_1, k_2, k_3, \ldots, k_n)$ a lze možné je odvodit zobecněním následujícího postupu.

Uvažujme pět různých prvků a_1, a_2, b_1, b_2, b_3, utvořme z nich všech 5! permutací a zkoumejme, kolik z těchto permutací se ztotožní, jestliže u všech pěti prvků odstraníme indexy, tj. položíme $a_1 = a_2 = a$ a $b_1 = b_2 = b_3 = b$. Např. s permutací (b, a, b, b, a) se ztotožní permutace z prvků a_1, a_2, b_1, b_2, b_3 zobrazené v tab. 1:

$(b_1, a_1, b_2, b_3, a_2)$	$(b_2, a_1, b_1, b_3, a_2)$	$(b_3, a_1, b_1, b_2, a_2)$
$(b_1, a_1, b_3, b_2, a_2)$	$(b_2, a_1, b_3, b_1, a_2)$	$(b_3, a_1, b_2, b_1, a_2)$
$(b_1, a_2, b_2, b_3, a_1)$	$(b_2, a_2, b_1, b_3, a_1)$	$(b_3, a_2, b_1, b_2, a_1)$
$(b_1, a_2, b_3, b_2, a_1)$	$(b_2, a_2, b_3, b_1, a_1)$	$(b_3, a_2, b_2, b_1, a_1)$

tab. 1

Pro umístění prvků a_1 a a_2 na místě prvku a je celkem $2!$ možností, pro umístění prvků b_1, b_2 a b_3 na místě prvku b je celkem 3! možností. S permutací (b, a, b, b, a) tedy splyne celkem $2! \cdot 3! = 12$ původních permutací z pěti prvků a_1, a_2, b_1, b_2 a b_3.

Tímto způsobem lze všech 5! permutací z daných pěti prvků a_1, a_2, b_1, b_2 a b_3 rozložit do takových tříd (skupin), že v každé třídě budou právě ty permutace, které se po odstranění indexů u daných prvků ztotožní s jednou vybranou permutací „bez indexů“. A protože každá tato třída odpovídá jediné permutaci s opakováním ze dvou prvků a a tří prvků b a protože má $2! \cdot 3! = 12$ členů, platí pro počet permutací ze dvou prvků, z nichž první se může opakovat 2krát a druhý 3krát, vztah $P'(2, 3) = \frac{5!}{2! \cdot 3!} = \frac{(2+3)!}{2! \cdot 3!}$.

Zobecněním této úvahy lze dojít k následujícímu závěru:

<u>VĚTA:</u> POČET PERMUTACÍ S OPAKOVÁNÍM Z n PRVKŮ, V NICHŽ SE JEDNOTLIVÉ PRVKY OPAKUJÍ k_1- KRÁT, k_2- KRÁT, k_3- KRÁT, …, k_n- KRÁT, JE:

$$P'(k_1, k_2, k_3, ..., k_n) = \frac{(k_1 + k_2 + k_3 + ... + k_n)!}{k_1! \cdot k_2! \cdot k_3! \cdot ... \cdot k_n!}. \tag{21}$$

Vzhledem k tomu, že permutace z n prvků bez opakování jsou zvláštním případem permutací s opakováním z těchto n prvků, mělo by platit: $P'\left(\underbrace{1, 1, ..., 1}_{n-krát}\right) = P(n)$. To platí, neboť $P'\left(\underbrace{1, 1, ..., 1}_{n-krát}\right) = \frac{(1+1+...+1)!}{1! \cdot 1! \cdot ... \cdot 1!} = n! = P(n)$.

Budeme-li uvažovat permutace s opakováním ze dvou prvků, z nichž jeden se opakuje k-krát a druhý $(n-k)$-krát, dostaneme pro počet těchto permutací postupným odvozením ze vztahu (21): $P'(k, n-k) = \frac{(k+(n-k))!}{k! \cdot (n-k)!} = \frac{n!}{k! \cdot (n-k)!} = \binom{n}{k} = C_k(n)$.

<u>VĚTA:</u> POČET PERMUTACÍ S OPAKOVÁNÍM ZE DVOU PRVKŮ OPAKUJÍCÍCH SE k -KRÁT A $(n-k)$-KRÁT JE ROVEN POČTU k-PRVKOVÝCH KOMBINACÍ Z n PRVKŮ A PLATÍ:

$$P'(k, n-k) = \binom{n}{k} = C_k(n). \tag{22}$$

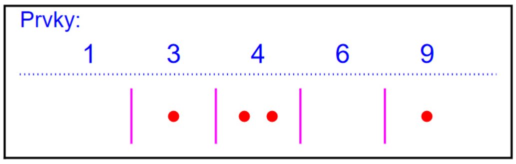
1.10 Kombinace s opakováním

Posledním druhem skupin, kterými se budeme zabývat, jsou skupiny, ve kterých nezávisí na pořadí a jejichž členy se mohou opakovat.

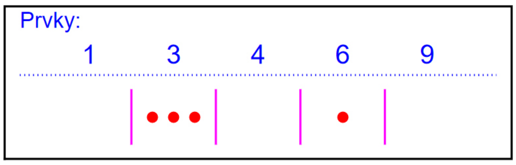
k-<u>ČLENNÁ KOMBINACE S OPAKOVÁNÍM</u> Z n PRVKŮ JE NEUSPOŘÁDANÁ k-TICE SESTAVENÁ Z TĚCHTO PRVKŮ TAK, ŽE KAŽDÝ SE V NÍ MŮŽE VYSKYTOVAT NEJVÝŠE k-KRÁT.

Určíme nyní počet všech k-členných kombinací s opakováním z n prvků, který onačíme $C'_k(n)$. Každou tuto kombinaci „zašifrujeme“ pomocí **uspořádané** skupiny, v níž použijeme $n-1$ svislých příček (čárek), pomocí nichž vznikne n přihrádek, do kterých budeme umísťovat k prvků (teček). Pro každý z daných n prvků zakreslíme do jemu odpovídající přihrádky tolik teček, kolikrát se daný prvek v dané kombinaci vyskytuje; pokud se v dané kombinaci daný prvek nevyskytuje, zůstane jeho přihrádka prázdná.

Uvažujme např. prvky 1, 3, 4, 6 a 9, ze kterých budeme vytvářet čtveřice. Kombinaci (3, 4, 4, 9) odpovídá schéma zobrazené na obr. 11, kombinaci (3, 3, 3, 6) pak odpovídá schéma zobrazené na obr. 12.



obr. 11



obr. 12

Každé k-členné kombinaci s opakováním z n prvků tak odpovídá jediná **uspořádaná** $(k+n-1)$-tice o k tečkách a $n-1$ čárkách (viz právě uvedené konkrétní ukázky).

Na tyto skupiny dvou znaků lze nahlížet jako na permutace s opakováním ze dvou prvků, z nichž jeden se opakuje k-krát a druhý $(n-1)$-krát. Vzhledem k tomu, že se jedná o vzájemně jednoznačné přiřazení, je možné (i s využitím kapitoly 1.9 a vztahů (21) a (22)) postupně psát: $C'_k(n) = P'(k, n-1) = \frac{(k+(n-1))!}{k! \cdot (n-1)!} = \frac{(n+k-1)!}{k! \cdot (n-1)!} = \frac{(n+k-1)!}{k! \cdot ((n+k-1)-k)!} = \binom{n+k-1}{k} = C_k(n+k-1)$.

<u>VĚTA:</u> POČET $C'_k(n)$ VŠECH k-ČLENNÝCH KOMBINACÍ S OPAKOVÁNÍM Z n PRVKŮ JE:

$$C'_k(n) = \binom{n+k-1}{k}. \tag{23}$$

S využitím vztahu (23) je tedy možné tvrdit, že počet $C'_k(n)$ všech k-členných kombinací s opakováním z n prvků je roven počtu všech k-členných kombinací (bez opakování) z $n+k-1$ prvků.

2. PRAVDĚPODOBNOST

2.1 Náhodné pokusy

Z hodin fyziky a chemie známe pokusy, které při přesném dodržení předepsaných počátečních podmínek vedou vždy k témuž, předem očekávanému výsledku. Jinými slovy řečeno: pokus je možné kdykoliv zopakovat s vědomím, že pokud dodržíme předepsané počáteční podmínky, dopadne experiment vždy stejně.

Mezi experimenty, o kterých byla zmínka, patří např. volně puštěný kamínek padající k zemi, zapálený papír, který v přítomnosti vzduchu shoří, …

Ve vědě, výzkumu i v praktické činnosti se ale setkáváme většinou s pokusy, které i při dodržení předepsaných podmínek vedou k různým výsledkům – výsledky pokusu se mohou od jednoho provedení pokusu k jinému měnit. Takové pokusy nazýváme **náhodné pokusy**. Jejich výsledky závisí nejen na předepsaných počátečních podmínkách, ale také na náhodných vlivech.

Při změření výšky, ze které bude volně puštěn kamínek na podložku v různých experimentech, a doby dopadu na podložku vyjde po výpočtu z rovnic popisujících volný pád téměř stejná velikost tíhového zrychlení. Rozdíly mezi velikostmi tíhového zrychlení vypočtené z různých měřeních budou způsobeny jak nepřesností měření výšky a doby dopadu, tak náhodnými vlivy (třes ruky uvolňující kamínek, lokální změna teploty vedoucí k lokální změně hustoty vzduchu, a tím i ke změně velikosti odporové síly brzdící pohyb kamínku, …).

Klasickými náhodnými pokusy jsou losování loterie, tahy sportky, házení hrací kostkou, házení mincí, míchání karet, … Pro vědecký výzkum a další praktické uplatnění jsou důležitější pokusy, při nichž se na pokusných zvířatech testuje účinek nového léku, a pokusy, jejichž cílem je přesné stanovení:

1. fyzikální konstanty (velikost rychlosti světla ve vakuu, gravitační konstanta, …);
2. lokální charakteristiky (velikost místního tíhového zrychlení, hustota vzduchu, …);
3. parametrů daného přístroje (ohnisková vzdálenost čočky, elektromotorické napětí zdroje, …);
4. …

Příklady, na nichž bude následující teorie vysvětlována, budeme volit z oblasti klasických pokusů (losování loterie, házení kostek, …), neboť jednotlivé zákonitosti je možné zde sledovat lépe, názorněji, a tedy i jednodušeji.

Nebude tedy nutné zpočátku sledovat i fyzikální, chemické či biologické aspekty dané problematiky.

Budeme předpokládat, že u každého náhodného pokusu jsme schopni předem vyjmenovat všechny jeho možné výsledky, a to tak, že platí:

1. <u>výsledky se navzájem vylučují</u> – nastal-li jeden výsledek, nemohl nastat druhý;

Padne-li při házení kostkou jednička nemůže zároveň při stejném hodu padnout i šestka.

2. <u>jeden z výsledků nastane vždy</u> – nemůže nastat žádný jiný výsledek než jeden z možných výsledků.

Při házení šestistěnnou kostkou padne vždy 1 nebo 2 nebo 3 nebo 4 nebo 5 nebo 6, ale nic jiného.

Množinu všech takto stanovených výsledků pokusu nazveme **množinou všech možných výsledků pokusu**. Značí se většinou Ω a její prvek pak ω.

Ve středoškolské matematice (a jejích aplikacích) je množina Ω vždy konečná.

Při stanovení množiny všech možných výsledků pokusu máme určitou libovůli v závislosti na tom, podle jakého hlediska chceme výsledky pokusu rozlišovat, jak detailně (jemně) chceme rozlišovat, …

Ze třídy, v níž je 30 žáků, mají být losem určeni 4 žáci, kteří se podrobí testu na přítomnost alkoholu v krvi. Pokud chceme určit jen to, kteří žáci se k testu dostaví, jsou výsledky losování dány čtyřčlennými kombinacemi ze 30 prvků a jejich počet je $\binom{30}{4}$. Pokud ale chceme určit losem i to, v jakém pořadí budou k testu přistupovat, jsou výsledky losování dány čtyřčlennými variacemi ze 30 prvků a jejich počet je $30 \cdot 29 \cdot 28 \cdot 27$.

Rozhodnutí, zda bude nebo nebude v tomto případě záviset na pořadí žáků, musíme učinit na začátku celého experimentu, a pak experiment podle toho vyhodnocovat.

2.2 Jevy

Podmnožiny množiny všech možných výsledků nazýváme **jevy**.

Za jev tedy považujeme: padnutí jedničky při hodu kostkou; padnutí sudého čísla při hodu kostkou; padnutí jedničky nebo čísla většího než čtyři při hodu kostkou; vylosování losi s číslem 259638 při tahu loterie; první kartou, kterou obdrží první hráč hry ze zamíchaného balíčku karet, bude žaludské eso; student u zkoušky si vylosuje otázku, na kterou není připraven; …

Jevy popisujeme zpravidla nějakou vlastností, která je společná prvkům této podmnožiny. Jevy se označují velkými písmeny latinské abecedy (A, B, C, …). V daném pokusu lze rozeznávat tolik jevů, kolik je podmnožin množiny všech možných výsledků Ω. Mezi jevy se počítají i dva zvláštní případy:

1. nemožný jev – prázdná množina, která je podmnožinu množiny všech možných výsledků Ω;

Při házení šestistěnnou kostkou padne sedmička, …

2. jistý jev – celá množina Ω, která je též podmnožinou množiny všech možných výsledků Ω.

Při házení symetrickou mincí padne panna nebo orel; při házení šestistěnnou kostkou padne jedno z čísel 1, 2, 3, 4, 5 nebo 6; …

Jestliže má množina Ω celkem *m* prvků, pak existuje 2^m různých jevů, z nichž je předmětem našeho zájmu obvykle jen několik.

Jevy jsou velmi podobné podmnožinám, přesto se pro vztahy a operace s jevy v počtu pravděpodobnosti vžilo poněkud odlišné názvosloví:

1. je-li $\omega \in \mathrm{A}$, říkáme, že výsledek ω je **příznivý** jevu A;

Jev A: na kostce padne sudé číslo a $\omega = 2$ (tj. padla dvojka).

2. je-li $\mathrm{A} \subset \mathrm{B}$, říkáme, že jev A je **podjevem** jevu B;

A – na kostce padne dvojka, B – na kostce padne číslo menší než čtyři. Nastane-li podjev A, nastává současně i jev B.

3. jev $\mathrm{A} \cup \mathrm{B}$, který nastává právě tehdy, nastane-li aspoň jeden z jevů A a B, se nazývá **sjednocením** jevů A a B;
4. jev $\mathrm{A} \cap \mathrm{B}$, který nastává právě tehdy, nastanou-li oba jevy A a B současně, se nazývá **průnikem** jevů A a B;
5. je-li $\mathrm{A} \cap \mathrm{B} = \emptyset$, říkáme, že jevy A a B se **navzájem vylučují**;

Např. A – na kostce padne sudé číslo, B – na kostce padne trojka.

6. jev A', který nastává právě tehdy, když jev A nenastává, nazýváme jevem **opačným** k jevu A.

Např. jev A – při hodu mincí padne panna, pak jev opačný jevu A je jev A': při hodu mincí padne orel.

Ilustrační příklad je uveden v kapitole 2.2.1.

2.2.1 Ilustrační příklad na vysvětlení základních vlastností jevů

Uvažujme hru, která spočívá v házení dvěma kostkami (např. bílou a černou) a v níž výhra nastává, padne-li alespoň na jedné kostce šestka. Hod kostkami je náhodným pokusem s těmito možnými výsledky: $\Omega = \{(1,1);(1,2);(1,3);...;(1,6);(2,1);(2,2);(2,3);...;(6,6)\}$, kde první číslo ve dvojici udává počet ok na bílé kostce a druhé číslo počet ok na černé kostce.

Výhra v této hře (tj. alespoň jedna šestka) je jev V: $\mathrm{V} = \{(1,6);(2,6);...;(5,6);(6,1);(6,2);...;(6,6)\}$. Šestka na bílé kostce je jev B, pro který platí: $\mathrm{B} = \{(6,1);(6,2);...;(6,6)\}$, šestka na černé kostce je jev C, pro který platí: $\mathrm{C} = \{(1,6);(2,6);...;(5,6);(6,6)\}$. Současné nastání obou jevů B a C je průnik $\mathrm{B} \cap \mathrm{C} = \{(6,6)\}$. Nastání alespoň jednoho z jevů B a C je sjednocení $\mathrm{B} \cup \mathrm{C} = \mathrm{V}$. Jev B i jev C jsou podjevy jevu V, tedy: $B \subset V$ a $C \subset V$. Jev V' (opačný jev k jevu V) znamená, že ani na jedné kostce nepadla šestka. Jevy V a V' se vzájemně vylučují, tj. $V \cap V' = \emptyset$.

2.3 Pravděpodobnosti

Uvažujme nějaký pokus s množinou možných výsledků Ω. Pokus provedeme celkem *n* krát a pro každý možný výsledek pokusu ω zaznamenáme, kolik pokusů (z *n* provedených) skončilo právě tímto výsledkem. Toto číslo $n(\omega)$ se nazývá **četnost** výsledku ω, podíl $\frac{n(\omega)}{n}$ se nazývá **relativní četnost** výsledku ω (v *n* pokusech). Četnosti $n(\omega)$ jsou celá nezáporná čísla, jejichž součet je roven *n*; relativní četnosti $\frac{n(\omega)}{n}$ jsou nezáporné zlomky, jejichž součet je roven 1.

Při losování pomocí mince, při házení kostkou, … chceme spravedlivě určit výherce, další postup hry, sázky, … Přesvědčení o spravedlnosti tohoto postupu vychází ze dvou poznatků:

1. z úvahy, že není důvod, proč by padnutí jedné strany mince, jednoho čísla na hrací kostce, … mělo být pravděpodobnější než druhé;
2. ze zkušenosti, že relativní četnosti všech výsledků (padnutí obou stran mince, padnutí jednoho z šesti čísel na hrací kostce, …) jsou skutečně přibližně stejné, pokud provedeme velký počet hodů.

Tuto skutečnost zobrazuje i tab. 2, v níž jsou uvedeny četnosti a relativní četnosti při deseti (četnost $n_{10}(\omega)$), sto (četnost $n_{100}(\omega)$) a tisíci (četnost $n_{1000}(\omega)$) hodech standardní šestistěnnou kostkou. Je vidět, že při 1000 hodech jsou četnosti (a tedy i relativní četnosti) rozloženy rovnoměrněji než při deseti hodech kostkou.

Hod	$n_{10}(\omega)$	$\frac{n_{10}(\omega)}{10}$	$n_{100}(\omega)$	$\frac{n_{100}(\omega)}{100}$	$n_{1000}(\omega)$	$\frac{n_{1000}(\omega)}{1000}$
1	4	0,4	11	0,11	163	0,163
2	1	0,1	15	0,15	181	0,181
3	3	0,3	13	0,13	164	0,164
4	1	0,1	22	0,22	186	0,186
5	1	0,1	18	0,18	154	0,154
6	0	0	21	0,21	152	0,152

tab. 2

Na základě předcházejících úvah můžeme vyslovit definici pravděpodobnosti:

MÁ-LI NÁHODNÝ POKUS *m* MOŽNÝCH VÝSLEDKŮ A JSOU-LI TYTO VÝSLEDKY STEJNĚ PRAVDĚPODOBNÉ, PAK O KAŽDÉM Z NICH ŘEKNEME, ŽE MÁ <u>PRAVDĚPODOBNOST</u>

$$p = \frac{1}{m}. \tag{24}$$

Výsledky hodu mincemi, hodu kostkou, … jsou stejně pravděpodobné, jedná-li se o ideální minci, kostku, …, tj. těleso geometricky pravidelné a stejnorodé. V případě, že by byla mince či kostka nějakým způsobem „upravena“ (např. dovnitř do blízkosti jedné stěny by bylo přidáno závažíčko), bylo by padnutí jednotlivých stěn mince či kostky nesymetrické. Ovšem znalost toho, že mince či kostka je „upravená“, nestačí k určení pravděpodobnosti. Můžeme jen říci, že pravděpodobnost daného hodu bude větší či menší než u příslušné mince či kostky neupravené. Pokud je třeba stanovit tuto pravděpodobnost „upravené“ mince či kostky, je možný pouze tento postup: Na základě provedeného velkého počtu pokusů zjistit relativní četnost daného hodu a tuto relativní četnost pak přijmout za přibližnou hodnotu pravděpodobnosti daného hodu.

PŘESNOU HODNOTU PRAVDĚPODOBNOSTI PAK POVAŽUJEME ZA NEZNÁMOU KONSTANTU, LEŽÍCÍ NĚKDE BLÍZKO ZJIŠTĚNÉ RELATIVNÍ ČETNOSTI.

Přesně tímto způsobem probíhají veškerá vědecká měření vedoucí k určování hodnot fyzikálních konstant či sledovaných veličin.

Za hodnotu proměřované fyzikální konstanty se vezme hodnota ležící dostatečně blízko k hodnotě, která byla naměřena s nejvyšší četností.

Vzhledem k tomu, že pravděpodobnost tedy velmi úzce souvisí s relativní četností, je přirozené, že i pravděpodobnosti mají podobné vlastnosti.

<u>VĚTA:</u> MĚJME NÁHODNÝ POKUS S MNOŽINOU MOŽNÝCH VÝSLEDKŮ Ω. PRAVDĚPODOBNOSTI $p(\omega)$ TĚCHTO VÝSLEDKŮ JSOU NEZÁPORNÁ ČÍSLA, JEJICHŽ SOUČET JE ROVEN JEDNÉ.

2.4 Pravděpodobnosti jevů

<u>Ilustrační příklad:</u> Je náhodně zvolená rodina se třemi dětmi. Pokud máme zájem jen o pohlaví nejstaršího, prostředního a nejmladšího dítěte, pak máme tyto výsledky pokusu: (c,c,c), (c,c,d), (c,d,c), (c,d,d), (d,c,c), (d,c,d), (d,d,c), (d,d,d). V jednotlivých závorkách znamená c chlapce, d dívku a první písmeno značí nejstarší, druhé prostřední a třetí nejmladší dítě. Nebudeme-li uvažovat, že se ve skutečnosti rodí o něco více chlapců než děvčat, budou mít obě pohlaví stejnou pravděpodobnost narození, a tedy i všechny možné výsledky uvažovaného pokusu mají stejnou pravděpodobnost: $\frac{1}{8}$. Jaká je pravděpodobnost, že se v uvažované rodině narodí dva chlapci a jedno děvče?

<u>Řešení:</u> Pravděpodobnost toho, že se v rodině narodí dva chlapci a jedno děvče, je pak $\frac{3}{8}$, neboť tomuto jevu odpovídají tři možné výsledky: (c,c,d), (c,d,c), (d,c,c).

<u>PRAVDĚPODOBNOST JEVU</u> A, KTERÁ SE ZNAČÍ $P(\mathrm{A})$, SE DEFINUJE JAKO SOUČET PRAVDĚPODOBNOSTÍ VÝSLEDKŮ PŘÍZNIVÝCH JEVU A. PLATÍ TEDY:

$$P(\mathrm{A}) = \sum_{\omega \in \mathrm{A}} p(\omega). \tag{25}$$

Pokud má pokus celkem n stejně pravděpodobných výsledků, je pravděpodobnost jevu A rovna:

$$P(\mathrm{A}) = \frac{m(\mathrm{A})}{n}, \tag{26}$$

kde $m(\mathrm{A})$ je počet výsledků příznivých jevu A.

Vztah (26) lze vyjádřit slovně takto: $P(\mathrm{A}) = \dfrac{\textit{počet příznivých výsledků}}{\textit{počet všech možných výsledků}}$.

Z definice pravděpodobnosti vyplývají i její vlastnosti:

1. pravděpodobnost nemožného jevu je rovna nule: $P(\emptyset) = 0$;
2. pravděpodobnost jistého jevu je rovna jedné: $P(\Omega) = 1$;
3. pro pravděpodobnost libovolného jevu A platí: $P(\mathrm{A}) \in \langle 0; 1 \rangle$.

Na základě poslední uvedené vlastnosti pravděpodobnosti je zřejmé, že v řadě běžných situacích je možné pravděpodobnost vyjádřit i v procentech.

2.5 Sčítání pravděpodobností

Nechť se jevy A a B navzájem vylučují, tj. $\mathrm{A} \cap \mathrm{B} = \emptyset$. Pravděpodobnost jejich sjednocení je rovna:

$$P(\mathrm{A} \cup \mathrm{B}) = \sum_{\omega \in \mathrm{A} \cup \mathrm{B}} p(\omega) = \sum_{\omega \in \mathrm{A}} p(\omega) + \sum_{\omega \in \mathrm{B}} p(\omega) = P(\mathrm{A}) + P(\mathrm{B}). \tag{27}$$

Tedy:

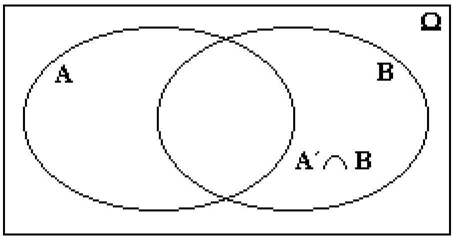
<u>VĚTA:</u> PRAVDĚPODOBNOST SJEDNOCENÍ DVOU NAVZÁJEM SE VYLUČUJÍCÍCH JEVŮ JE ROVNA SOUČTU JEJICH PRAVDĚPODOBNOSTÍ.

Tuto větu lze zobecnit na r navzájem se vylučujících jevů:

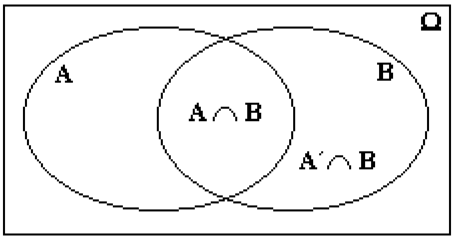
<u>**V:**</u> Jsou-li $\mathrm{A}_1, \mathrm{A}_2, ..., \mathrm{A}_r$ navzájem se vylučující jevy, tj. $\mathrm{A}_\mathrm{i} \cap \mathrm{A}_\mathrm{j} = \emptyset$ pro $i, j = 1, 2, ..., r$ a $i \neq j$, potom:

$$P(\mathrm{A}_1 \cup \mathrm{A}_2 \cup ... \cup \mathrm{A}_\mathrm{r}) = P(\mathrm{A}_1) + P(\mathrm{A}_2) + ... + P(\mathrm{A}_\mathrm{r}). \tag{28}$$

Nechť se jevy A a B nyní vzájemně nevylučují, tj. $\mathrm{A} \cap \mathrm{B} \neq \emptyset$. Jev $\mathrm{A} \cup \mathrm{B}$ je možné vyjádřit jako sjednocení vylučujících se jevů A a $\mathrm{A}' \cap \mathrm{B}$ (viz obr. 13) a jev B lze vyjádřit jako sjednocení vylučujících se jevů $\mathrm{A} \cap \mathrm{B}$ a $\mathrm{A}' \cap \mathrm{B}$ (viz obr. 14).



obr. 13



obr. 14

Podle vztahu (27) a na základě právě provedených úvah o náhradním vyjádření jevů je možné psát: $P(\mathrm{A}\cup\mathrm{B})=P(\mathrm{A})+P(\mathrm{A}'\cap\mathrm{B})$ a $P(\mathrm{B})=P(\mathrm{A}\cap\mathrm{B})+P(\mathrm{A}'\cap\mathrm{B})$. Odtud je již možné vyjádřit hledanou pravděpodobnost $P(\mathrm{A}\cup\mathrm{B})$ dvou navzájem se nevylučujících jevů ve tvaru:

$$P(\mathrm{A}\cup\mathrm{B})=P(\mathrm{A})+P(\mathrm{B})-P(\mathrm{A}\cap\mathrm{B}). \tag{29}$$

Mnemotechnicky je možné si vztah (29) pamatovat tak, že je analogický vztahu pro výpočet počtu prvků sjednocení dvou množin, které mají nenulový průnik.

Vztah (29) platí i pro dva navzájem se vylučující jevy A a B, neboť pro tyto jevy je $P(\mathrm{A}\cap\mathrm{B})=0$, a tedy vztah (29) přejde na vztah (27).

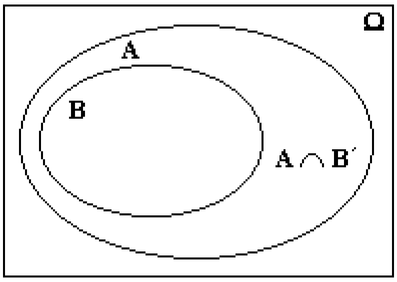
Pro libovolný jev A platí, že jevy A a A' se navzájem vylučují a že $\mathrm{A}\cup\mathrm{A}'=\Omega$. Platí tedy: $P(\mathrm{A})+P(\mathrm{A}')=P(\Omega)=1$. Odtud pro pravděpodobnost jevu opačného k jevu A dostáváme:

$$P(\mathrm{A}')=1-P(\mathrm{A}). \tag{30}$$

Vztah $\mathrm{A}\cup\mathrm{A}'=\Omega$ je dobré si představit na konkrétním příkladu. Pokud při házení šestistěnnou kostkou bude jev A „padne trojka“, pak jev opačný k jevu A je „nepadne trojka“. Při hodu tedy určitě nastane právě jeden z popsaných jevů – buď jev A nebo jev k němu opačný. A jejich sjednocením je celá množina všech možných výsledků pokusu (tj. že padne jedno z čísle 1, 2, 3, 4, 5 nebo 6).

Je-li B podjevem jevu A, tj. $\mathrm{B}\subset\mathrm{A}$, potom je možné jev A vyjádřit jako sjednocení vzájemně se vylučujících jevů B a $\mathrm{A}\cap\mathrm{B}'$ (viz obr. 15), takže platí: $P(\mathrm{A})=P(\mathrm{B})+P(\mathrm{A}\cap\mathrm{B}')$. Odtud je možné vyjádřit:

$$P(\mathrm{A}\cap\mathrm{B}')=P(\mathrm{A})-P(\mathrm{B}). \tag{31}$$



obr. 15

Právě uvedenou úvahu shrneme do věty popisující pravděpodobnosti jevu B, který je podjevem jevu A.

JE-LI JEV B PODJEVEM JEVU A, TJ. POKUD PLATÍ $\mathrm{B}\subset\mathrm{A}$, POTOM:

PLATÍ VZTAH (31), TJ. $P(\mathrm{A}\cap\mathrm{B}')=P(\mathrm{A})-P(\mathrm{B})$;

PRO PRAVDĚPODOBNOSTI JEVŮ A A B PLATÍ:

$$P(\mathrm{B})\leq P(\mathrm{A}). \tag{32}$$

Platnost vztahu (32) vyplývá ze vztahu (31), který definuje pravděpodobnost $P(\mathrm{A}\cap\mathrm{B}')$, která je (jako jakákoliv jiná rozumně definovaná pravděpodobnost) nezáporná. Proto je nezáporná i pravá strana výrazu (31), odkud již plyne platnost vztahu (32).

2.6 Nezávislé jevy (násobení pravděpodobností)

V úvodu kapitoly 2.4 byl popsán příklad s rodinami, které mají tři děti. Uvedli jsme, že existuje celkem osm různých možností, co se pořadí narození jednotlivých dětí týče. Každý takový jev má pravděpodobnost $P=\dfrac{1}{8}$.

<u>Ilustrační příklad:</u> Jaká je pravděpodobnost, že v rodině bude nejstarší chlapec a prostřední děvče?

<u>Řešení:</u> Jevu A, který značí „nejstarší dítě je chlapec", přísluší pravděpodobnost $P(\mathrm{A}) = \frac{4}{8} = \frac{1}{2}$, jevu B, který značí „prostřední dítě je děvče", přísluší pravděpodobnost $P(\mathrm{B}) = \frac{4}{8} = \frac{1}{2}$. Pro jev $\mathrm{A} \cap \mathrm{B}$ („nejstarší je chlapec a prostřední je děvče") dostáváme $P(\mathrm{A} \cap \mathrm{B}) = \frac{2}{8} = \frac{1}{4}$. Zároveň ale vidíme, že platí vztah: $P(\mathrm{A} \cap \mathrm{B}) = P(\mathrm{A}) \cdot P(\mathrm{B})$.

Jednotlivé pravděpodobnosti uvedené v tomto příkladě byly určeny pomocí vztahu (26) a dat uvedených v úvodním rozboru úlohy v úvodu kapitoly 2.4.

Pokud bychom stejným způsobem chtěli vyšetřovat pravděpodobnost jevu C, který značí „všechny tři děti jsou chlapci", dostáváme $P(\mathrm{C}) = \frac{1}{8}$. V tomto případě, ale $P(\mathrm{A} \cap \mathrm{C}) \neq P(\mathrm{A}) \cdot P(\mathrm{C})$. Důvod je ten, že jevy A a B lze považovat za nezávislé (pohlaví prvního dítěte nemá vliv na pohlaví druhého dítěte), zatímco jevy A a C nezávislé nejsou (v tomto případě totiž jev C závisí na jevu A).

Nezávislé jevy můžeme tedy definovat takto:

ŘEKNEME, ŽE <u>JEVY</u> A A B <u>JSOU NEZÁVISLÉ</u>, JESTLIŽE PLATÍ:

$$P(\mathrm{A} \cap \mathrm{B}) = P(\mathrm{A}) \cdot P(\mathrm{B}). \tag{33}$$

Obecně jevy A, B, C, …, Z se nazývají nezávislé, jestliže pravděpodobnosti průniku libovolných dvou, tří, čtyř, … z nich jsou rovny součinům jejich pravděpodobností. Jsou-li jevy A, B, C, …, Z nezávislé, jsou nezávislé i každé dva z nich.

<u>VĚTA:</u> JSOU-LI A A B NEZÁVISLÉ JEVY, POTOM JSOU TAKÉ NEZÁVISLÉ DVOJICE JEVŮ: A A B', A' A B A TAKÉ A' A B'.

<u>Důkaz:</u> Stačí si uvědomit, že (s využitím obr. 14) platí: $\mathrm{A} \cap \mathrm{B}' = \mathrm{A} \cap (\mathrm{A} \cap \mathrm{B})'$. Odtud vyplývá $P(\mathrm{A} \cap \mathrm{B}') = P\left(\mathrm{A} \cap (\mathrm{A} \cap \mathrm{B})'\right)$. Vzhledem k tomu, že $(\mathrm{A} \cap \mathrm{B}) \subset \mathrm{A}$ je možné použít vztah (31) psát: $P\left(\mathrm{A} \cap (\mathrm{A} \cap \mathrm{B})'\right) = P(\mathrm{A}) - P(\mathrm{A} \cap \mathrm{B})$. Uvědomíme-li si, že jevy A a B jsou nezávislé, je možné psát ve shodě se vztahem (33): $P(\mathrm{A} \cap \mathrm{B}) = P(\mathrm{A}) \cdot P(\mathrm{B})$.

Nyní je možné celý rozpis důkazu shrnout:

$$P(\mathrm{A} \cap \mathrm{B}') = P\left(\mathrm{A} \cap (\mathrm{A} \cap \mathrm{B})'\right) = P(\mathrm{A}) - P(\mathrm{A} \cap \mathrm{B}) = P(\mathrm{A}) - P(\mathrm{A}) \cdot P(\mathrm{B}) = P(\mathrm{A}) \cdot (1 - P(\mathrm{B})) =$$

$= P(\mathrm{A}) \cdot P(\mathrm{B}')$. Z toho tedy vyplývá, že jevy A a B' jsou nezávislé.

Analogicky je možné dokázat i zbývající dva případy uvedené ve větě.

2.7 <u>*Nezávislé pokusy*</u>

Uvažujme pokus, který se skládá ze dvou dílčích pokusů (např. hod dvěma kostkami). Nechť ω_1 značí libovolný výsledek prvního pokusu a $p_1(\omega_1)$ jeho pravděpodobnost; ω_2 nechť značí libovolný výsledek druhého pokusu a $p_2(\omega_2)$ jeho pravděpodobnost. Za možné výsledky uvažovaného „sdruženého" pokusu budeme považovat všechny dvojice (ω_1, ω_2), které budou mít pravděpodobnosti $p(\omega_1, \omega_2)$.

Ze všech pokusů jsou důležité **nezávislé pokusy**.

ŘEKNEME, ŽE <u>DÍLČÍ POKUSY JSOU NEZÁVISLÉ</u>, JESTLIŽE PRO VŠECHNY MOŽNÉ VÝSLEDKY (ω_1, ω_2) PLATÍ:

$$p(\omega_1, \omega_2) = p_1(\omega_1) \cdot p_2(\omega_2). \tag{34}$$

<u>VĚTA:</u> JSOU-LI DÍLČÍ POKUSY NEZÁVISLÉ A JE-LI JEV A URČEN JEN VÝSLEDKEM PRVNÍHO DÍLČÍHO POKUSU A JEV B JEN VÝSLEDKEM DRUHÉHO POKUSU, PAK JEVY A A B JSOU NEZÁVISLÉ.

Chápeme-li nezávislost v běžném, nematematickém smyslu slova, pak je zpravidla jasné, které dva pokusy jsou nezávislé a které ne. Matematická definice nezávislosti (dvou) pokusů, která byla právě uvedena, je volena tak, aby co nejlépe odpovídala tomuto běžnému chápání.

Příklady nezávislých pokusů: hod bílou a černou hrací kostkou, hod mincí, tažení prvního a druhého čísla ve sportce, …

V případě pokusu, který se skládá z *n* dílčích pokusů, řekneme, že dílčí pokusy jsou nezávislé, jestliže pro každý možný výsledek pokusu $(\omega_1, \omega_2, ..., \omega_n)$ platí:

$$p(\omega_1, \omega_2, ..., \omega_n) = p_1(\omega_1) \cdot p_2(\omega_2) \cdot ... \cdot p_n(\omega_n), \qquad (35)$$

kde symboly mají obdobný význam jako v případě dvou pokusů. I tvrzení o nezávislosti jevů určených různými dílčími pokusy zůstává platné i pro případ *n* dílčích pokusů.

2.8 Binomické rozdělení (Bernoulliho schéma)

Uvažujme nyní *n* nezávislých pokusů, z nichž každý má tytéž dva možné výsledky; jejich pravděpodobnosti nechť jsou ve všech pokusech tytéž: *p* a *q*. Přitom ovšem platí $p + q = 1$. Možné výsledky sdruženého pokusu jsou tedy všechny *n*-tice symbolů odpovídajících dvěma uvažovaným výsledkům pokusu. Pravděpodobnosti těchto *n*-tic jsou tedy součiny *n* činitelů *p* nebo *q* podle typu výsledku sdruženého pokusu.

Např. při házení symetrickou mincí je pravděpodobnost, že padne orel (O) i panna (P) stejná – tj. $p = q = 0,5$. Při pěti hodech může nastat např. tento výsledek: POOOP; pravděpodobnost tohoto pokusu by tedy byla $0,5^5$.

Při házení symetrickou šestistěnnou kostkou je pravděpodobnost, že padne číslo 6 rovna $\frac{1}{6}$, pravděpodobnost, že nepadne číslo 6 je rovna $\frac{5}{6}$. Výsledek experimentu pěti hodů je tento: 12634 a jeho pravděpodobnost je $\frac{1}{6} \cdot \left(\frac{5}{6}\right)^4 = 0,08$.

Speciálně tak dostaneme:

1. pravděpodobnost, že všechny pokusy skončí výsledkem, který má pravděpodobnost *p*, je p^n;
2. pravděpodobnost, že všechny pokusy skončí výsledkem, který má pravděpodobnost *q*, je q^n.

Uvažujme nyní takový výsledek, že prvních *k* pokusů skončilo výsledkem, jemuž odpovídá pravděpodobnost *p*, a zbývajících $n - k$ pokusů skončilo výsledkem, jemuž odpovídá pravděpodobnost *q*. Takový výsledek má pravděpodobnost $p^k \cdot q^{n-k}$. Stejnou pravděpodobnost mají i výsledky, v nichž:

1. posledních *k* pokusů má výsledek, jemuž odpovídá pravděpodobnost *p*, a předcházejících $n - k$ pokusů má výsledek, jemuž odpovídá pravděpodobnost *q*;
2. určitých *k* pokusů má výsledek, jemuž odpovídá pravděpodobnost *p*, a ostatních $n - k$ pokusů má výsledek, jemuž odpovídá pravděpodobnost *q*.

Uvažujme nyní takový jev A_k, který znamená, že právě *k* pokusů je těch s pravděpodobností *p*, a tedy zbývajících $n - k$ pokusů je těch s pravděpodobností *q*. Přitom je jedno, o kterých *k* pokusů půjde. Tomuto jevu jsou příznivé ty *n*-tice, které mají pravděpodobnost $p^k \cdot q^{n-k}$. Těchto *n*-tic je tolik, kolika způsoby lze vybrat *k* pokusů z celkového počtu pokusů *n*, tedy $\binom{n}{k}$.

Při házení symetrickou kostkou se zaměřením na padnutí čísla 6 jsou tyto výsledky stejně pravděpodobné: 12645, 16245, 62541, …

Můžeme tedy formulovat závěr:

MĚJME n NEZÁVISLÝCH POKUSŮ, Z NICHŽ KAŽDÝ SKONČÍ BUĎ ZDAREM S PRAVDĚPODOBNOSTÍ p NEBO NEZDAREM S PRAVDĚPODOBNOSTÍ q. PRAVDĚPODOBNOST JEVU A_k, ŽE PRÁVĚ k POKUSŮ SKONČÍ ZDAREM, JE:

$$P(A_k)=\binom{n}{k}\cdot p^k\cdot q^{n-k} \tag{36}$$

PRO $k=0,1,2,\dots,n$.

S využitím binomické věty (viz kapitola 1.7) a vztahu mezi pravděpodobnosti p a q můžeme psát:

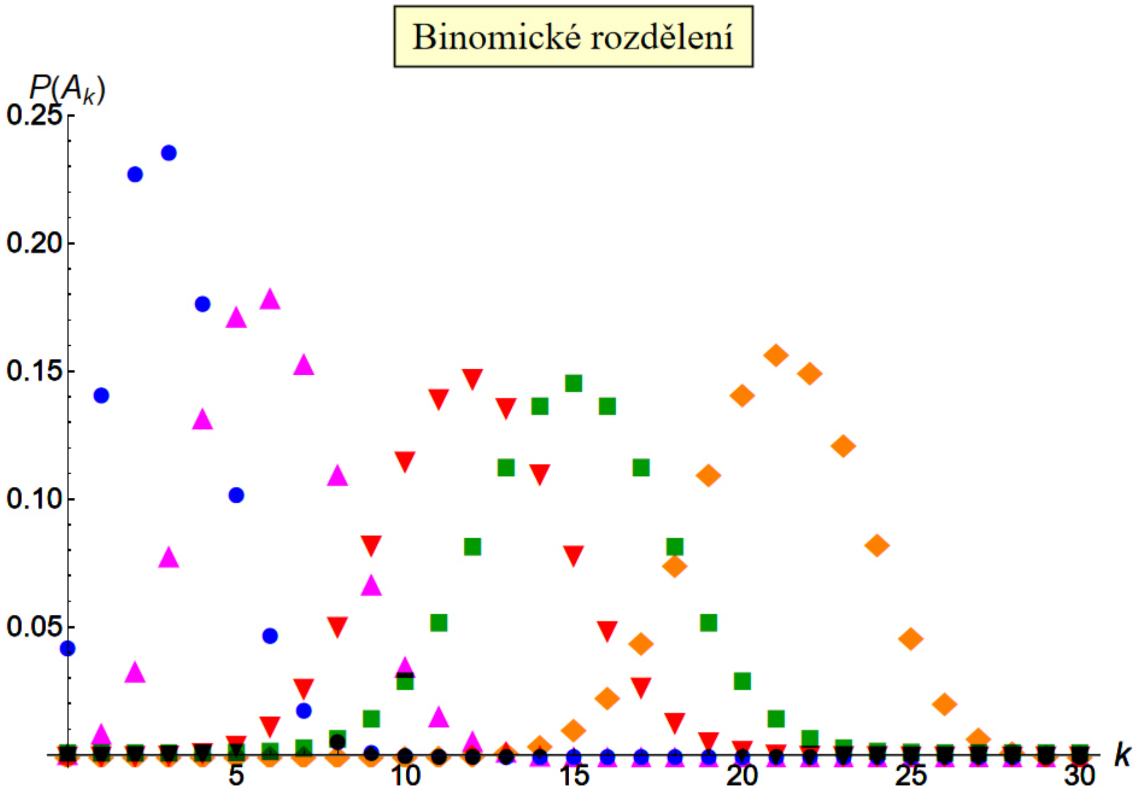
$(p+q)^n=\binom{n}{0}\cdot q^n+\binom{n}{1}\cdot p\cdot q^{n-1}+\binom{n}{2}\cdot p^2\cdot q^{n-2}+\dots+\binom{n}{n}\cdot p^n$. Vzhledem k tomu, že $p+q=1$, je i $(p+q)^n=1$.

Stejný závěr dostaneme i z pravděpodobností úvahy: Jevy „právě k pokusů bude zdařilých" se pro $k=0,1,2,\dots,n$ navzájem vylučují a jeden z nich nastává vždy. Jejich sjednocení je proto jistý jev, a proto platí $\binom{n}{0}\cdot q^n+\binom{n}{1}\cdot p\cdot q^{n-1}+\binom{n}{2}\cdot p^2\cdot q^{n-2}+\dots+\binom{n}{n}\cdot p^n=1$.

Pravděpodobnosti $\binom{n}{k}\cdot p^k\cdot q^{n-k}$ pro $k=0,1,2,\dots,n$ tvoří tzv. **binomické rozdělení** (**Bernoulliho schéma**). Tímto problémem se zabýval švýcarský matematik Jacob Bernoulli (1655 – 1705).

Binomické rozdělení je pro 30 nezávislých pokusů zobrazeno na obr. 16 pro pravděpodobnost:

1. $p=0{,}1$ – viz ●;
2. $p=0{,}2$ – viz ▲;
3. $p=0{,}4$ – viz ▼;
4. $p=0{,}5$ – viz ■;
5. $p=0{,}7$ – viz ◆.



obr. 16

Z grafu zobrazeného na obr. 16 je patrné, že pro pravděpodobnost uvažovaného jevu A_k lze vyčíst:

1. její maximum se s rostoucí pravděpodobností p posouvá k vyššímu k;
2. její maximum nabývá velkých hodnot pro malé (resp. velké) hodnoty pravděpodobnosti p;

3. pro většinu hodnot k je relativně malá.

2.9 Podmíněná pravděpodobnost

Uvažujme příklad, kdy házíme dvěma čtyřstěnnými hracími kostkami, tj. kostkami, na nichž mohou padnout čísla 1, 2, 3 nebo 4. Jaká je pravděpodobnost, že součet čísel na obou kostkách bude roven šesti? Jaká pravděpodobnost, že součet čísel na obou kostkách bude roven šesti za předpokladu, že na že jedné z kostek padlo číslo menší než tři?

Všechny možnosti, jak mohou obě kostky padnout jsou zobrazeny v tab. 3 (první číslo uspořádané dvojice je číslo, které padlo na první kostce, druhé číslo je to, které padlo na druhé kostce):

$$\begin{aligned}&\{[1;1];[1;2];[1;3];[1;4];\\&[2;1];[2;2];[2;3];[2;4];\\&[3;1];[3;2];[3;3];[3;4];\\&[4;1];[4;2];[4;3];[4;4]\}\end{aligned}$$

tab. 3

Jev A je jev „součet čísel na kostkách je roven šesti“, jev B je jev „na jedné z kostek padlo číslo menší než 3“. Pravděpodobnost, že nastane jev A, určíme z příznivých dvojic: $\mathrm{A}=\{[2;4];[3;3];[4;2]\}$. Tedy $P(\mathrm{A})=\frac{3}{16}$.

Jevu B vyhovují uspořádané dvojice zobrazené v tab. 4, v nichž alespoň jedno číslo je menší než tři:

$$\begin{aligned}\mathrm{B}=\{&[1;1];[1;2];[1;3];[1;4];\\&[2;1];[2;2];[2;3];[2;4];\\&[3;1];[3;2];\\&[4;1];[4;2]\}\end{aligned}$$

tab. 4

Z dvojic zobrazených v tab. 4 nás zajímají ty, které mají součet rovný šesti – a to jsou dvojice $\mathrm{A}|\mathrm{B}=\{[2;4];[4;2]\}$. Hledaná pravděpodobnost tedy je: $P(\mathrm{A}|\mathrm{B})=\frac{2}{12}=\frac{1}{6}$.

Na výpočet pravděpodobnosti $P(\mathrm{A}|\mathrm{B})$ lze tedy nahlížet tak, že jí počítáme „standardním vztahem“ (tj. *počet příznivých výsledků pokusu dělený počtem všech výsledků*), jen celkový počet výsledků vybíráme z již zredukované původní množiny všech možných výsledků.

Uvažujme nyní obecně takový pokus, který má možné výsledky ω a jejichž pravděpodobnost je $p(\omega)$. Představme si nyní situaci, že se během pokusu dozvíme, že nastal jev B. To pro další vyšetřování pravděpodobností znamená, že musíme původní pravděpodobnosti $p(\omega)$ nahradit „novými“ pravděpodobnostmi $p(\omega|\mathrm{B})$, které zohlední fakt, že máme informaci o tom, že nastal jev B. Tyto pravděpodobnosti $p(\omega|\mathrm{B})$ se nazývají **podmíněné pravděpodobnosti za podmínky B**.

Pojem podmíněná pravděpodobnost je nutné správně chápat.

Je-li jev H „náhodně vybraný člověk je fanoušek hokeje“, pak $P(\mathrm{H})$ je nepodmíněná (tj. „normální“) pravděpodobnost.

Pokud budeme zkoumat, zda náhodně vybraný muž (jev M) je fanoušek hokeje (jev H), pak se ptáme na podmíněnou pravděpodobnost $P(\mathrm{H}|\mathrm{M})$. Při jejím experimentálním hledání je tedy nutné splnit podmínku, že oslovený respondent je muž, a až poté se ptát, zda fandí hokeji. Lze očekávat, že $P(\mathrm{H}|\mathrm{M})>P(\mathrm{H})$, tj. že mezi muži bude více fanoušků hokeje než mezi všemi lidmi (např. ve městě).

Přitom ke každé podmíněné pravděpodobnosti existuje **inverzní podmíněná pravděpodobnost**, v níž jsou obě podmínky „prohozeny“.

Ve výše uvedeném příkladu s hokejem bychom se ptali, zda náhodně vybraný fanoušek hokeje je muž. Tato podmíněná pravděpodobnost $P\left(\mathrm{M}|\mathrm{H}\right)$ bude patrně velmi blízká jedné (mezi fanoušky hokeje budou patrně převažovat muži).

Zřejmě pak platí $p\left(\omega|\mathrm{B}\right)=0$ pro $\omega\notin\mathrm{B}$, protože tyto výsledky už nemohou nastat. Pravděpodobnosti výsledků $\omega\in\mathrm{B}$ pak musíme úměrně zvětšit, aby jejich součet byl opět roven jedné. Musí tedy platit:

$$p\left(\omega|\mathrm{B}\right)=\frac{p(\omega)}{P(\mathrm{B})} \tag{37}$$

pro $\omega\in\mathrm{B}$. (V tomto případě uvažujeme jev B s nenulovou pravděpodobností.)

Zlomek (37) je skutečně větší než pravděpodobnost $p(\omega)$, protože $P(\mathrm{B})\in(0;1)$.

Podmíněná pravděpodobnost libovolného jevu A za podmínky B je pak definována jako součet pravděpodobností výsledků příznivých jevu A, tedy:

$$P\left(\mathrm{A}|\mathrm{B}\right)=\sum_{\omega\in\mathrm{A}}p\left(\omega|\mathrm{B}\right)=\sum_{\omega\in\mathrm{A}}\frac{p(\omega)}{P(\mathrm{B})}=\frac{P(\mathrm{A}\cap\mathrm{B})}{P(\mathrm{B})}\,. \tag{38}$$

2.9.1 Pokus se stejně pravděpodobnými výsledky

Pokud se jedná o pokus se stejně pravděpodobnými výsledky, je $P(\mathrm{A}\cap\mathrm{B})=\dfrac{m(\mathrm{A}\cap\mathrm{B})}{m}$ a $P(\mathrm{B})=\dfrac{m(\mathrm{B})}{m}$. Potom platí (po dosazení do vztahu (38)): $P\left(\mathrm{A}|\mathrm{B}\right)=\dfrac{m(\mathrm{A}\cap\mathrm{B})}{m(\mathrm{B})}$, kde $m(\mathrm{X})$ označuje počet výsledků příznivých jevu X.

2.9.2 Nezávislé jevy

Jsou-li jevy A a B navzájem nezávislé, pak $P\left(\mathrm{A}|\mathrm{B}\right)=\dfrac{P(\mathrm{A})\cdot P(\mathrm{B})}{P(\mathrm{B})}=P(\mathrm{A})$ a $P\left(\mathrm{B}|\mathrm{A}\right)=\dfrac{P(\mathrm{A})\cdot P(\mathrm{B})}{P(\mathrm{A})}=P(\mathrm{B})$. Tedy nastání jevu B neovlivní pravděpodobnost jevu A a naopak.

2.9.3 Závislé jevy

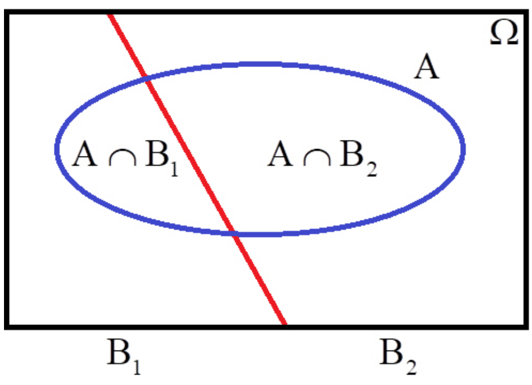
Pro jevy A a B, které nejsou nezávislé, můžeme podmínku pro podmíněnou pravděpodobnost psát ve tvaru:

$$P(\mathrm{A}\cap\mathrm{B})=P(\mathrm{B})\cdot P\left(\mathrm{A}|\mathrm{B}\right)=P(\mathrm{A})\cdot P\left(\mathrm{B}|\mathrm{A}\right). \tag{39}$$

Vztah (39) se někdy nazývá **vzorec pro násobení pravděpodobností**.

2.9.4 Navzájem se vylučující jevy

Uvažujme dva jevy B_1 a B_2, které se navzájem vylučují, přičemž jeden z nich nastává určitě, tj. platí: $\mathrm{B}_1\cap\mathrm{B}_2=\emptyset$ a současně $\mathrm{B}_1\cup\mathrm{B}_2=\Omega$.



obr. 17

Potom pro libovolný jev A z množiny Ω (ve shodě se schematickým znázorněním na obr. 17) platí: $\mathrm{A}=\mathrm{A}\cap\mathrm{B}_1\cup\mathrm{A}\cap\mathrm{B}_2$. Jevy $\mathrm{A}\cap\mathrm{B}_1$ a $\mathrm{A}\cap\mathrm{B}_2$ se opět vylučují, a proto můžeme psát:

$P(\mathrm{A}) = P(\mathrm{A} \cap \mathrm{B}_1) + P(\mathrm{A} \cap \mathrm{B}_2)$. Vyjádříme-li nyní pravděpodobnosti obou průniků, dostaneme vztah:

$$P(\mathrm{A}) = P(\mathrm{B}_1) \cdot P(\mathrm{A}|\mathrm{B}_1) + P(\mathrm{B}_2) \cdot P(\mathrm{A}|\mathrm{B}_2). \qquad (40)$$

Vztah (40) je tzv. **vzorec pro celkovou pravděpodobnost**.

3. STATISTIKA

Statistika zkoumá jevy na dostatečně rozsáhlém souboru případů a hledá ty vlastnosti jevů, které se projeví až v souboru případů a ne na jednom případě.

> Typickým příkladem je průměr známek ve škole z daného předmětu, průměrná volební účast ve volbách, procentuální zastoupení homosexuálně orientovaných jedinců v populaci, …

3.1 Statistický soubor, statistické jednotky, znak

Základním pojmem je **statistický soubor** a jeho prvky, které se nazývají **statistické jednotky**. Tento soubor vyšetřujeme z hlediska zvoleného **znaku** (nebo více znaků). Hodnota znaku musí být vždy jednoznačně stanovena. Existují přitom dva druhy znaků:

1. kvantitativní znak – jeho hodnota je určena číselnou hodnotou (výška postavy v cm, známka z matematiky, počet sourozenců, příjem uvedený v Kč, …);
2. kvalitativní znak – jeho hodnoty jsou dány kvalitou (povolání dané osoby, barva očí, pohlaví, rodinný stav, …), nejjednodušší kvalitativní znaky jsou přitom ty, které jsou dány určitým jevem a jeho opakem (muž – žena, prospěl – neprospěl, voják – nevoják, …); tyto znaky se nazývají **alternativní znaky**.

> Příkladem statistického souboru je databáze žáků školy, ve které je o každém žákovi uložena řada informací – jméno, příjmení, datum narození, místo narození, adresa bydliště, číslo ISIC karty, známky na pololetních vysvědčeních, uvolnění z předmětů (např. z TV), … Jednotlivými statistickými jednotkami jsou pak jednotliví žáci a jejich „vlastnosti". Znakem, z hlediska kterého statistický soubor vyšetřujeme, může být „narozen v květnu", „bydlí v Praze", „muž", „průměr známek na všech vysvědčeních z matematiky je menší než 1,5", …

3.2 Rozdělení četností a jeho grafické znázornění

3.2.1 Zavedení pojmů a jejich vysvětlení

Při statistickém šetření se zpravidla vyšetřuje více znaků, které nás zajímají jak každý zvlášť, tak i ve vzájemném vztahu.

> Při předvolebním průzkumu zajímá politické strany, jak bude kdo volit (jeden znak – „strana, kterou volím"). Je ale zajímavé také vědět, jak volí lidé z venkova a z větších měst („strana, kterou volím" a „bydliště"), jak volí lidé různého vzdělání („strana, kterou volím" a „nejvyšší dosažené vzdělání") nebo náboženství („strana, kterou volím" a „vyznání"), …

Pro další výklad se omezíme jen na šetření, ve kterých nás bude zajímat jen jeden znak. Výsledkem šetření tedy je seznam jednotek s udáním hodnoty znaku u každé z nich. Jsou-li jednotky v seznamu očíslovány 1, 2, …, n ($n \in \mathbb{N}$), pak jim odpovídající hodnoty znaku x označíme symboly $x_1, x_2, ..., x_n$. Často může znak nabývat jen určitého počtu r různých hodnot; tyto hodnoty znaku označíme symboly $x_1^*, x_2^*, ..., x_r^*$.

> Ve statistickém souboru známek z matematiky na výročním vysvědčení 30 žáků jedné třídy nebude 30 různých hodnot ($n = 30$) ale jen pět ($r = 5$) – tj. jedna z pěti známek používaných ke klasifikaci.

Pro každou možnou hodnotu x_j^* zjistíme, kolikrát se vyskytla mezi znaky $x_1, x_2, ..., x_n$. Takto zjištěný počet n_j se nazývá **četnost** hodnoty x_j^*. Součet četností všech možných hodnot znaku se rovná počtu všech jednotek souboru, tj. platí:

$$\sum_{j=1}^{r} n_j = n . \tag{41}$$

Relativní četnost je pak definována vztahem

$$\nu_j = \frac{n_j}{n} \tag{42}$$

a udává, jaká část souboru má hodnotu znaku x_j^*. Na základě vztahů (41) a (42) je zřejmé, že platí:

$$\sum_{j=1}^{r} \nu_j = 1 . \tag{43}$$

Relativní četnost je možné vyjadřovat také v procentech; součet relativních četností je pak roven 100 %.

Rozdělení četností daného znaku statistického souboru můžeme graficky znázornit různými způsoby:

1. tabulkou rozdělení četností – zpravidla dvouřádková (resp. dvousloupcová) tabulka, v jejímž prvním řádku (resp. sloupci) jsou uvedeny hodnoty x_j^* znaku a ve druhém řádku (resp. sloupci) je uvedena četnost nebo relativní četnost (udaná desetinným číslem nebo v procentech);
2. spojnicovým diagramem (polygonem četností) – spojení bodů, jejichž první souřadnice je hodnota x_j^* znaku a druhá souřadnice je odpovídající četnost;
3. sloupkovým diagram (histogram) – používá se zejména tehdy, jsou-li hodnoty znaku sdruženy do intervalů; tyto intervaly pak tvoří základny sloupků a odpovídající četnosti udávají výšky sloupků;

Do intervalů lze sdružit např. tělesnou výšku postavy udanou v centimetrech (vytvoříme intervaly 150 - 154, 155 - 159, 160 - 164, …), počty obyvatel ve městech (0 - 5000, 5001 - 10000, 10001 - 15000, …) a podobně.

4. kruhovým diagram – různým hodnotám znaku odpovídají kruhové výseče, jejichž plošné obsahy (a tedy středové úhly) jsou úměrné četnostem (v tomto grafu se velmi často používají relativní četnosti vyjádřené v procentech).

Histogram se používá většinou pro znázornění rozdělení četností kvantitativních znaků a kruhový diagram se používá většinou pro znázornění rozdělení četností kvalitativních znaků.

3.2.2 Konkrétní ukázka

Máme k dispozici seznam známek z matematiky na pololetním vysvědčení v jedné třídě:

$$1, 1, 1, 2, 2, 3, 2, 4, 3, 3, 4, 5, 3, 4, 5, 1, 2, 2, 1, 2, 3, 4, 3, 3, 3. \tag{44}$$

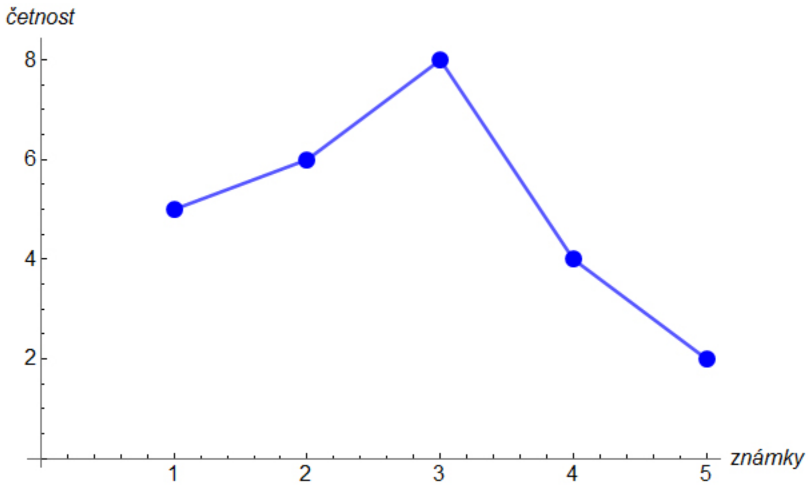
Pro některé statistické charakteristiky (uvedené v kapitole 3.3) je nutné řadu známek (44) seřadit vzestupně. Dostaneme tak řadu

$$1, 1, 1, 1, 1, 2, 2, 2, 2, 2, 2, 3, 3, 3, 3, 3, 3, 3, 3, 4, 4, 4, 4, 5, 5. \tag{45}$$

Tabulka rozdělení četností je zobrazena na obr. 18. Na základě ní je pak sestrojen polygon četností zobrazený na obr. 19. Jednotlivé body tohoto grafu odpovídají jednotlivým známkám a jejich četnostem.

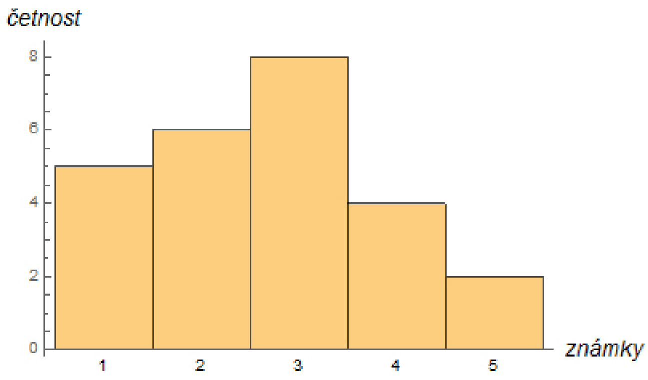
Známka	Četnost
1	5
2	6
3	8
4	4
5	2

obr. 18



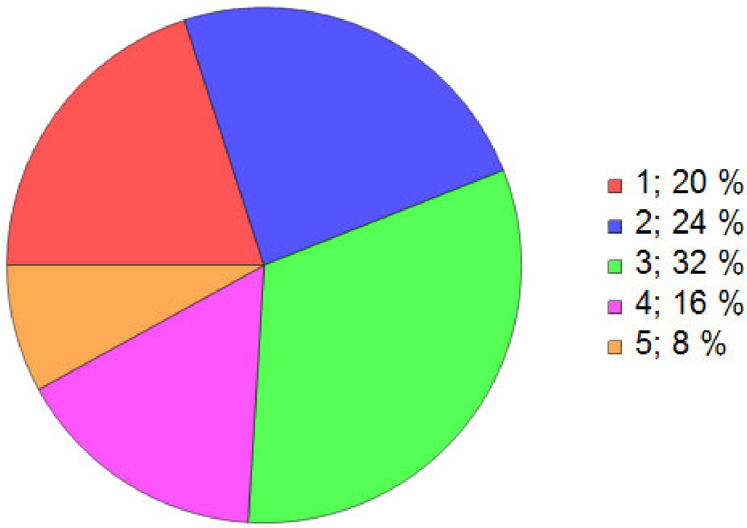
obr. 19

Na obr. 20 je pak zobrazen histogram se stejnou vypovídací hodnotou, jakou má polygon četností.



obr. 20

S využitím vztahu (42) je možné dopočítat relativní četnosti a zobrazit je v kruhovém diagramu (viz obr. 21).



obr. 21

3.3 Charakteristiky polohy a variability

Pro kvantitativní znaky lze zavést další charakteristiky, které jsou udány samostatnými čísly. Úplnou statistiku daného znaku podává rozdělení četnosti (viz kapitola 3.2), ale i čísla charakterizující polohu a variabilitu znaku jsou pro řadu aplikací důležitá.

3.3.1 Charakteristiky polohy

Nejčastěji používanou **charakteristikou polohy** znaku x je **aritmetický průměr**.

ARITMETICKÝ PRŮMĚR $\overline{x}$ ZNAKU x JE SOUČET HODNOT ZNAKU ZJIŠTĚNÝCH U VŠECH JEDNOTEK SOUBORU DĚLENÝ POČTEM VŠECH JEDNOTEK SOUBORU:

$$\overline{x} = \frac{1}{n}\sum_{i=1}^{n} x_i \,. \tag{46}$$

Počítáme-li aritmetický průměr z tabulky rozdělení četností, je nutné každou hodnotu x_j^* násobit její četností. Vztah (46) je proto nutné upravit do tvaru:

$$\overline{x} = \frac{1}{n}\sum_{j=1}^{r} x_j^* n_j \,. \tag{47}$$

Určování průměru má smysl pouze v těch statistických souborech, ve kterých jsou individuální odchylky jednotlivých prvků souboru nahodilé. Ve statistických souborech, ve kterých jsou individuální odchylky jednotlivých prvků souboru systematické, je vhodnější určovat **průměrný přírůstek**.

Soubory s nahodilými odchylkami jsou typicky data získaná během fyzikálního měření. Jednotlivé odchylky jsou dány nepřesným odečtením hodnoty z měřidla, nepřesným nastavením hodnoty elektrického proudu, … Systematická chyba (špatná metoda měření, řádová chyba při čtení údaje z měřidla, …) se velmi rychle v případě fyzikálního měření pozná a lze ji tedy eliminovat (změnou metody měření, pečlivějším čtením z přístroje, …).

V některých případech se statistický soubor skládá z více dílčích souborů A, B, C a D. Přitom známe počty jednotek n_A, n_B, n_C a n_D v dílčích souborech a průměry $\overline{x}_A$, $\overline{x}_B$, $\overline{x}_C$ a $\overline{x}_D$ znaku x v dílčích souborech. Na základě znalosti těchto dat chceme určit průměr $\overline{x}$ v celém souboru.

Dílčími soubory mohou být např. seznamy žáků dvou paralelních tříd v ročníku střední školy a znakem x známky z matematiky na pololetním vysvědčení v jednotlivých třídách. Za dílčí soubory lze považovat známky z jednoho předmětu (např. matematika), které mají různé váhy (domácí úkol, zkoušení, test, pololetní práce, …).

Hledaný průměr určíme tak, že součet hodnot sledovaného znaku v celém souboru vydělíme počtem všech jednotek v souboru. Dostaneme tak vztah:

$$\overline{x} = \frac{\overline{x}_A \cdot n_A + \overline{x}_B \cdot n_B + \overline{x}_C \cdot n_C + \overline{x}_D \cdot n_D}{n_A + n_B + n_C + n_D}. \tag{48}$$

Vztah (48) se nazývá **vážený průměr** čísel $\overline{x}_A$, $\overline{x}_B$, $\overline{x}_C$ a $\overline{x}_D$ s váhami n_A, n_B, n_C a n_D. Tento vztah je možné zobecnit i pro větší počet dílčích souborů, než jsou uvedené čtyři.

Nyní se pokusíme vysvětlit, proč je právě *průměr* dobrou charakteristikou polohy znaku. Uvažujme, že každá zjištěná hodnota znaku je součtem dvou složek:

1. složka charakteristická pro celý soubor obsahující určitou globální informaci o tomto souboru;
2. individuální odchylka dané jednotky souboru, která má víceméně náhodný charakter.

Vypočítáme-li průměr, tak první složka vynikne, protože individuální odchylky, které jsou kladné i záporné, se navzájem téměř odečtou.

Právě uvedené můžeme ilustrovat na fyzikálním měření. Uvažujme měření průměru válcové součástky (viz tab. 5), Při měření ve fyzice se běžně uvádí první a druhý sloupec, třetí sloupec je doplněn pro ilustraci faktu, že individuální odchylky (vznikající chybou měření) se navzájem mají tendenci odečíst.

Číslo měření	$\frac{d}{\text{mm}}$	$\frac{\overline{d} + \Delta d}{\text{mm}}$
1	25,3	25,4 − 0,1
2	25,7	25,4 + 0,3
3	24,9	25,4 − 0,5
4	24,8	25,4 − 0,6
5	25,7	25,4 + 0,3
6	25,5	25,4 + 0,1
7	25,6	25,4 + 0,2
8	24,6	25,4 − 0,8
9	26,0	25,4 + 0,6
10	25,8	25,4 + 0,4

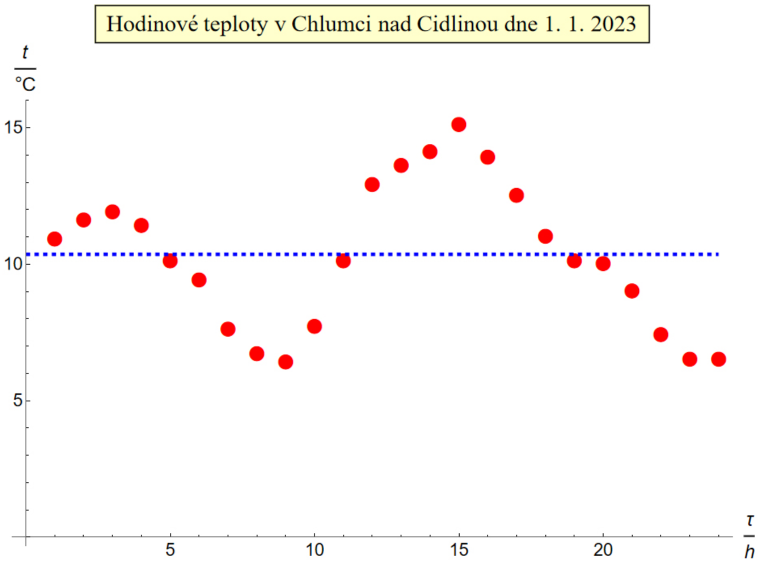
tab. 5

Z výše uvedeného důvodu není průměr vhodnou charakteristikou polohy v těch případech, v nichž individuální odchylky nejsou nahodilé, ale systematické. To je případ v časových řadách, v nichž data vykazují určitý trend, vývoj v čase; v tomto případě bývá lepším (zajímavějším) ukazatelem než průměr **průměrný přírůstek** (resp. **průměrný úbytek**) vyšetřovaného znaku za jedno časové období.

Jsou-li jednotlivá období očíslována 0, 1, 2, …, n, jsou-li x_0, x_1, x_2, …, x_n jim odpovídající hodnoty znaku a jsou-li $y_1 = x_1 - x_0$, $y_2 = x_2 - x_1$, …, $y_n = x_n - x_{n-1}$ přírůstky za jednotlivá období, pak pro průměrný přírůstek platí vztah

$$\overline{y} = \frac{1}{n}\sum_{i=1}^{n} y_i = \frac{x_n - x_0}{n}. \tag{49}$$

Ze vztahu (49) plyne, že k výpočtu průměrného přírůstku stačí znát celkový přírůstek za sledovaný počet období a není nutné znát přírůstky za jednotlivá období.



obr. 22

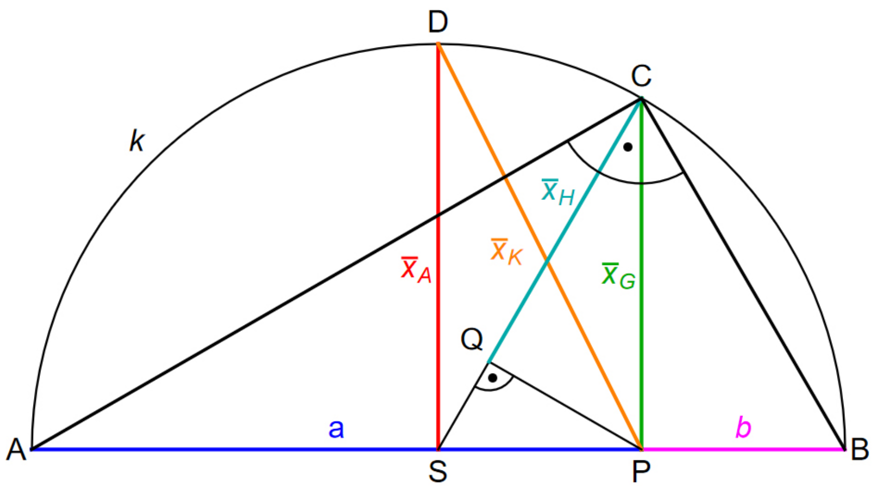
Na obr. 22 je zobrazen průběh průměrných hodinových teplot na stanici v Chlumci nad Cidlinou ze dne 1. 1. 2023. Průměrná teplota toho dne byla $10{,}4\ °\mathrm{C}$ (v grafu na obr. 22 je zobrazena modrou přerušovanou čarou), zatímco průměrný přírůstek (resp. úbytek) byl $-0{,}2\ °\mathrm{C}$.

Pro některé časové řady zobrazující zejména národohospodářské údaje (inflace, zisk, …) je vhodné udávat **průměrné tempo růstu** za jedno období. Tím je myšlen průměr podílů hodnot za dvě po sobě následující období, tedy průměr podílů $z_1 = \frac{x_1}{x_0}$, $z_2 = \frac{x_2}{x_1}$, …, $z_n = \frac{x_n}{x_{n-1}}$. V tomto případě se počítá **geometrický průměr** pomocí vztahu

$$\overline{z}_G = \sqrt[n]{z_1 \cdot z_2 \cdot \ldots \cdot z_n}\,. \tag{50}$$

Geometrický průměr se definuje pouze pro kladná čísla.

Dosazením do vztahu (50) pro průměrné tempo růstu získáme vztah $\overline{z}_G = \sqrt[n]{\frac{x_n}{x_0}}$.



obr. 23

Dalším typem průměru, který se pro kladná čísla x_1, x_2, …, x_n používá, je **harmonický průměr** definovaný vztahem

$$\overline{x}_H = \frac{1}{\frac{1}{n} \cdot \left(\frac{1}{x_1} + \frac{1}{x_2} + \ldots + \frac{1}{x_n} \right)}\,. \tag{51}$$

Posledním typem průměru, který lze zavést, ale který se běžně ve statistice příliš nepoužívá, je **kvadratický průměr** definovaný vztahem

$$\overline{x}_K = \sqrt{\frac{x_1^2 + x_2^2 + \ldots + x_n^2}{n}} \ . \tag{52}$$

Všechny průměry, které byly výše popsány, jsou schematicky pro dvě kladná čísla reprezentovaná úsečkami délek *a* a *b* zobrazeny na obr. 23. S využitím tohoto obrázku lze i na základě geometrických úvah a vztahů pro daná kladná čísla *a* a *b* odvodit vztahy (46), (50), (51) a (52).

Pro zobrazené průměry přitom platí vztah

$$\overline{x}_K \geq \overline{x}_A \geq \overline{x}_G \geq \overline{x}_H \ , \tag{53}$$

přičemž rovnost všech průměrů nastává pouze pro situaci, kdy $a = b$.

Dalšími charakteristikami polohy jsou **modus** a **medián**.

MODUS ZNAKU x SE ZNAČÍ $\mathrm{Mod}(x)$ A UDÁVÁ HODNOTU ZNAKU x S NEJVYŠŠÍ ČETNOSTÍ.

MEDIÁN ZNAKU x SE ZNAČÍ $\mathrm{Med}(x)$ A UDÁVÁ PROSTŘEDNÍ HODNOTU ZNAKU x, JSOU-LI HODNOTY $x_1, x_2, \ldots, x_n$ USPOŘÁDÁNY VZESTUPNĚ PODLE VELIKOSTI. PŘITOM PLATÍ:

$\mathrm{Med}(x) = x_{\frac{n+1}{2}}$ **JE-LI n LICHÉ;**

Tedy medián je skutečně prostřední hodnota v vzestupně uspořádaném souboru dat.

$\mathrm{Med}(x) = \frac{1}{2}\left(x_{\frac{n}{2}} + x_{\frac{n}{2}+1}\right)$ **JE-LI n SUDÉ.**

V tomto případě je medián průměrem dvou hodnot, které jsou „uprostřed“ vzestupně uspořádaného souboru dat.

Medián se užívá jako charakteristika polohy zejména v těch souborech, ve kterých hodnoty znaku u některých jednotek extrémně vybočují z řady ostatních hodnot (viz kapitola 3.3.3).

3.3.2 Charakteristiky variability

Každá charakteristika polohy (aritmetický průměr, modus, medián – viz kapitola 3.2.1) je určena číslem, kolem kterého jednotlivé hodnoty znaku kolísají. Míru tohoto kolísání vyjadřují **charakteristiky variability** (**charakteristiky proměnlivosti**) znaku.

Jako charakteristika variability k aritmetickému průměru (jakožto charakteristice polohy) se většinou používá **rozptyl**.

ROZPTYL s_x^2 ZNAKU x JE DEFINOVANÝ JAKO PRŮMĚR DRUHÝCH MOCNIN ODCHYLEK OD ARITMETICKÉHO PRŮMĚRU:

$$s_x^2 = \frac{1}{n}\sum_{i=1}^{n}\left(x_i - \overline{x}\right)^2 \ . \tag{54}$$

Je-li rozptyl počítán na základě tabulky četností, pak vztah (54) přejde na vztah

$$s_x^2 = \frac{1}{n}\sum_{j=1}^{r}\left(x_j^* - \overline{x}\right)^2 n_j \ . \tag{55}$$

Pokud provedeme ve vztahu (54) naznačené umocnění, získáme postupně:

$$s_x^2 = \frac{1}{n}\sum_{i=1}^{n}\left(x_i^2 - 2x_i\overline{x} + \overline{x}^2\right) = \frac{1}{n}\sum_{i=1}^{n}x_i^2 - 2\overline{x}\frac{1}{n}\sum_{i=1}^{n}x_i + \frac{1}{n}\cdot n\cdot\overline{x}^2 = \frac{1}{n}\sum_{i=1}^{n}x_i^2 - 2\overline{x}\cdot\overline{x} + \overline{x}^2 = \frac{1}{n}\sum_{i=1}^{n}x_i^2 - \overline{x}^2 \ .$$

Analogickou úpravu můžeme provést i se vztahem (55).

Další charakteristikou je **směrodatná odchylka**.

SMĚRODATNÁ ODCHYLKA s_x JE DEFINOVANÁ JAKO DRUHÁ ODMOCNINA Z ROZPTYLU:

$$s_x = \sqrt{\frac{1}{n}\sum_{i=1}^{n}\left(x_i - \overline{x}\right)^2} \ . \tag{56}$$

Směrodatná odchylka charakterizuje variabilitu znaku ve stejných jednotkách, v jakých byly udány hodnoty znaku, zatímco rozptyl je vyjádřen ve druhé mocnině těchto jednotek. Proto je pro technická měření (fyzika, chemie, …) vhodnější používat směrodatnou odchylku naměřených dat.

Bezrozměrným číslem (vyjadřujícím se též v procentech), které charakterizuje variabilitu znaku x, je **variační koeficient**.

VARIAČNÍ KOEFICIENT v_x JE DEFINOVÁN JAKO PODÍL SMĚRODATNÉ ODCHYLKY A ARITMETICKÉHO PRŮMĚRU:

$$v_x = \frac{s_x}{\bar{x}} \cdot 100\,\% \,. \tag{57}$$

Z definice je zřejmé, že variační koeficient má smysl pouze tehdy, pokud jsou hodnoty znaku x nezáporné.

V případě, že statistický soubor bude tvořen naměřenými daty (fyzikálními, chemickými, …), představuje variační koeficient relativní chybu (relativní odchylku) daného měření.

Poslední charakteristikou variability jsou **kvantily**.

KVANTIL x_p URČUJE HODNOTU ZNAKU, PRO KTEROU PLATÍ, ŽE NEJMÉNĚ p PROCENT PRVKŮ DANÉHO STATISTICKÉHO SOUBORU MÁ HODNOTU MENŠÍ NEBO ROVNOU x_p A $100 - p$ PROCENT PRVKŮ SOUBORU MÁ HODNOTU VĚTŠÍ NEBO ROVNOU x_p.

V praxi se používají tyto kvantily:

1. x_{50} je medián;

Definice uvedená výše říká, že medián je „prostřední“ prvek vzestupně uspořádaného souboru. To ale znamená, že 50 % hodnot souboru je menších nebo rovných mediánu a zbývajících 50 % hodnot je větších nebo rovných mediánu.

2. x_{25} je dolní kvartil;
3. x_{75} je horní kvartil;
4. $x_1, x_2, x_3, ..., x_{99}$ jsou percentily.

Při různých soutěžích, výběrových řízeních, přijímacích zkouškách, … se používají k popisu úspěšnosti daného uchazeče právě percentily. Je-li tedy např. percentil uchazeče 89 (tj. je na 89tém percentilu), znamená to, že 89 % uchazečů je horších než on a pouze 11 % je lepších než on. Pokud je navíc předem jasné, že úspěšných v daném řízení bude nejlepších 90 % účastníků, má tento účastník jisté, že je mezi úspěšnými.

Budeme-li pracovat se vzestupně uspořádaným statistickým souborem, který bude mít n statistických jednotek, je kvantil x_p roven prvku stojícímu na k-tém místě v souboru. Přitom k získáme ze vztahu

$$k = \frac{n \cdot p}{100} \tag{58}$$

tak, že výsledek uvedeného podílu zaokrouhlíme vždy nahoru.

3.3.3 <u>Konkrétní ukázka</u>

Vrátíme-li se nyní k seznamu známek z matematiky (44) uvedenému v kapitole 3.2.2, můžeme dopočítat další charakteristiky tohoto souboru.

Průměrná známka (vypočtená podle vztahu (46) je 2,68.

Medián známek je 3 - jedná se o třináctou známku ve vzestupně seřazeném seznamu známek (45), kterých je celkem 25. Pokud budeme chápat medián jako kvantil x_{50}, pak můžeme podle vztahu (58) spočítat pořadí známky, která je mediánem: $k = \frac{25 \cdot 50}{100} = \frac{1250}{100} = 12{,}5$; po zaokrouhlení nahoru dostáváme tedy $k \doteq 13$. Získali jsme tedy stejné pořadí jako s využitím definice mediánu.

Modus známek je 3, protože trojka se vyskytuje v seznamu známek nejčastěji (viz též tabulka četností na obr. 18).

Rozptyl známek je podle vztahu (54) roven 1,42.

Směrodatná odchylka je na základě vztahu (56) rovna 1,19.

Variační koeficient známek vypočtený podle vztahu (57) je roven 0,44, tedy 44 %.

Z praktického hlediska je někdy vhodnější udávat medián než průměr. V některých případech má totiž lepší vypovídací hodnotu.

Uvažujme tento soubor čísel: 1, 1, 1, 2, 3, 3, 4, 4, 5, 6, 8, 120, 150, 200, 242.

Průměr těchto čísel je 50 a medián je 4.

V případě souboru čísel 44, 45, 46, 47, 48, 48, 49, 50, 51, 52, 52, 53, 54, 55, 56, který má stejný počet statistických jednotek, je průměr 50 a medián také 50.

U druhé řady čísel mají průměr a medián stejnou vypovídací hodnotu. A to ne proto, že jsou obě hodnoty stejné, ale proto, že jsou přibližně stejné jako jednotlivá čísla v tomto souboru (směrodatná odchylka je rovna 3,56).

U první řady je ale většina čísel výrazně menší, než je průměr těchto čísel. Proto je medián lepší charakteristikou tohoto souboru, protože na základě něj lze získat představu o hodnotách čísel (obecně o prvcích daného statistického souboru): víme, že na prostředním místě vzestupně seřazeného souboru čísel je číslo 4, které je typickým číslem souboru, zatímco průměr je více než 10krát vyšší; směrodatná odchylka je pro tuto řadu rovna 80,89.

I z toho důvodu je např. údaj o průměrné mzdě obyvatel v dané zemi nevhodný. Výrazně lepší by bylo udávat medián platu obyvatel dané země.

3.4 Korelace

Minulé úvahy byly vedeny pro případ, kdy jsme popisovali:

1. jeden znak souboru;
2. více znaků, které jsme popisovali odděleně.

Dvěma oddělenými znaky mohou být známky žáků jedné třídy z matematiky a češtiny, délka a šířka školních hřišť v Praze, …

Nyní se budeme vyšetřovat popis dvojice znaků (x, y); výsledkem provedeného statistického šetření jsou přitom v souboru s délkou n data ve tvaru (x_1, y_1), (x_2, y_2), …, (x_n, y_n).

Může se jednat o:

data evropských zemí o roční spotřebě alkoholu na jednoho obyvatele a data o ročním úmrtí na cirhózu jater;

data o výsledcích státních maturitních zkoušek z matematiky a počtu žáků, kteří navštěvovali technicky zaměřené obory;

data o výskytu hrabošů na území jednotlivých krajů České republiky a počtu poničených hektarů půdy osetých obilím;

data o volební účasti obyvatelů jednotlivých krajských měst v prezidentských volbách a barvě domů dotazovaných voličů;

…

Kromě charakteristik polohy a variability, počítaných pro každý z obou znaků odděleně, je nutné v tomto případě znát i **míru statistické závislosti obou znaků**. Zvolíme-li za znaky polohy průměry $\overline{x}$ a $\overline{y}$ (viz vztah (46)) obou sledovaných znaků a za charakteristiky variability směrodatné odchylky s_x a s_y těchto znaků (viz vztah (56)), pak za společnou charakteristiku se velmi často volí **koeficient korelace** r_{xy}, který je definován vztahem:

$$r_{xy} = \frac{\frac{1}{n}\cdot\sum_{i=1}^{n}\left(x_i-\overline{x}\right)\cdot\left(y_i-\overline{y}\right)}{s_x\cdot s_y}. \qquad (59)$$

Čitatele zlomku lze postupně upravit do tvaru: $\frac{1}{n}\cdot\sum_{i=1}^{n}\left(x_i-\overline{x}\right)\cdot\left(y_i-\overline{y}\right)=$

$$=\frac{1}{n}\cdot\sum_{i=1}^{n}\left(x_iy_i-x_i\overline{y}-\overline{x}y_i+\overline{x}\cdot\overline{y}\right)=\frac{1}{n}\cdot\sum_{i=1}^{n}\left(x_iy_i\right)-\frac{1}{n}\cdot\overline{y}\cdot\sum_{i=1}^{n}x_i-\frac{1}{n}\cdot\overline{x}\cdot\sum_{i=1}^{n}y_i+\overline{x}\cdot\overline{y}=$$

$$=\frac{1}{n}\cdot\sum_{i=1}^{n}\left(x_iy_i\right)-\overline{x}\cdot\overline{y}-\overline{x}\cdot\overline{y}+\overline{x}\cdot\overline{y}=\frac{1}{n}\cdot\sum_{i=1}^{n}\left(x_iy_i\right)-\overline{x}\cdot\overline{y}.$$

Na základě těchto úprav lze vztah (59) psát ve tvaru:

$$r_{\mathrm{xy}} = \frac{\frac{1}{n} \cdot \sum_{i=1}^{\mathrm{n}} \left(x_{\mathrm{i}} y_{\mathrm{i}} \right) - \overline{x} \cdot \overline{y}}{s_{\mathrm{x}} \cdot s_{\mathrm{y}}}, \qquad (60)$$

který je pro manuální výpočty vhodnější (jednodušší).

Z obou vztahů (59) i (60) vyplývá, že koeficient korelace je bezrozměrná charakteristika daného statistického souboru; přitom platí $r_{\mathrm{xy}} \in \langle -1; 1 \rangle$. Krajních hodnot intervalu nabývá koeficient korelace tehdy, je-li mezi znaky *x* a *y* **funkční závislost** (tedy nejenom statistická), a to lineární závislost ve tvaru:

$$y = a \cdot x + b \qquad (61)$$

pro $a \in \mathbb{R} \setminus \{0\}$ a $b \in \mathbb{R}$. Znaménko koeficientu *b* přitom určuje, zda koeficient korelace bude nabývat hodnoty 1 nebo -1.

Vzhledem k rozdílu v čitateli zlomku ve vztahu (60) je zřejmé, že koeficient korelace r_{xy} může nabývat kladné i záporné hodnoty:

Je-li statistická závislost mezi znaky *x* a *y* taková, že nadprůměrným hodnotám *x* zpravidla odpovídají nadprůměrné hodnoty *y*, pak bude v čitateli zlomku ve vztahu (59) většina součinů kladných, a tedy i koeficient korelace bude **kladný**.

V souboru zaměstnanců podniku znaky počet let v podniku a roční příjem; v souboru evropských zemí spotřeba cigaret na jednoho obyvatele a počet úmrtí za rok na rakovinu plic; …

Jestliže nadprůměrným hodnotám znaku *x* odpovídají zpravidla podprůměrné hodnoty znaku *y* a naopak, bude v čitateli zlomku ve vztahu (59) většina součinů záporná, a tedy i koeficient korelace bude **záporný**.

V souboru obyvatel České republiky míra nezaměstnanosti v jednotlivých okresech a výše vkladů obyvatel u peněžních ústavů; v souboru úmrtí osob v daném roce spotřeba cigaret a věk, kterého se zemřelá osoba dožila; …

Není-li mezi znaky *x* a *y* žádná závislost, budou mít kladné i záporné součiny v čitateli zlomku ve vztahu (59) tendenci se navzájem rušit; koeficient korelace tedy bude **blízký nule**.

V souboru volebního průzkumu: volební účast a barva domu daného respondenta; v souboru zaměstnanců podniku: tělesná hmotnost a výše měsíčního příjmu; …

Ačkoliv jsme výše uvedené úvahy prováděli pouze na základě rozboru čitatele zlomku vztahu (59), je jmenovatel uvedeného vztahu nutný proto, aby koeficient korelace nezávisel na násobku jednotky, v níž sledované znaky *x* a *y* uvádíme.

Bude-li sledovaným znakem *x* měsíční příjem obyvatel a znakem *y* spotřeba pohonných hmot na čerpacích stanicích, tak koeficient korelace musí být nezávislý na tom, zda měsíční příjem bude uveden v korunách nebo v tisících korun. Popsanou změnou jednotky se čitatel vztahu (59) 1000krát zmenší, ale současně se 1000krát zmenší jeho jmenovatel. Koeficient korelace zůstane tedy beze změny.

Mírně odlišná situace nastane, pokud sledované znaky mohou nabývat pouze určitých hodnot. Uvažujme situaci, kdy znak *x* nabývá *r* různých hodnot označených symboly x_1^*, x_2^*, …, x_{r}^* a znak *y* nabývá pouze *s* různých hodnot označených symboly y_1^*, y_2^*, …, y_{s}^*. Do této skupiny patří i případy, kdy hodnoty znaků *x* a *y* sdružují do určitých intervalů; hodnoty x_{j}^* a y_{k}^* jsou pak středy těchto intervalů.

x \ *y*	y_1^*	y_2^*	…	y_{s}^*
x_1^*	n_{11}	n_{12}	…	$n_{1\mathrm{s}}$
x_2^*	n_{21}	n_{22}	….	$n_{2\mathrm{s}}$
…	…	…	…	…
x_{r}^*	$n_{\mathrm{r}1}$	$n_{\mathrm{r}2}$	….	n_{rs}

tab. 6

Pro každou možnou dvojici hodnot $\left\{x_{\mathrm{j}}^{*}; y_{\mathrm{k}}^{*}\right\}$ zjistíme, kolikrát se vyskytla mezi daty $\left\{x_1; y_1\right\}$, $\left\{x_2; y_2\right\}$, …, $\left\{x_{\mathrm{n}}; y_{\mathrm{n}}\right\}$; to znamená, že určíme její četnost n_{jk} v souboru n jednotek. Získáme tím **rozdělení četností dvojice znaků** (x, y) – viz tab. 6.

Další postup při hledání koeficientu korelace je tento:

1. sečteme četnosti v jednotlivých řádcích tab. 6, čímž získáme rozdělení četností samotného znaku *x*, a z tohoto rozdělení vypočítáme průměrnou hodnotu $\overline{x}$ a směrodatnou odchylku s_{x};
2. sečteme četnosti v jednotlivých sloupcích tab. 6, čímž získáme rozdělení četností samotného znaku *y*, a z tohoto rozdělení vypočítám průměrnou hodnotu $\overline{y}$ a směrodatnou odchylku s_{y};
3. vypočítáme součet $\sum_{i=1}^{\mathrm{n}}\left(x_{\mathrm{i}}y_{\mathrm{i}}\right)$; každý ze součinů $x_{\mathrm{i}}y_{\mathrm{i}}$ je přitom roven některému součinu $x_{\mathrm{j}}^{*}y_{\mathrm{k}}^{*}$ a to s četností n_{jk}, takže můžeme psát
$$\sum_{i=1}^{\mathrm{n}}\left(x_{\mathrm{i}}y_{\mathrm{i}}\right)=x_1^{*}y_1^{*}n_{11}+x_1^{*}y_2^{*}n_{12}+\ldots+x_1^{*}y_{\mathrm{s}}^{*}n_{1\mathrm{s}}+\ldots+x_{\mathrm{r}}^{*}y_1^{*}n_{\mathrm{r}1}+x_{\mathrm{r}}^{*}y_2^{*}n_{\mathrm{r}2}+\ldots+x_{\mathrm{r}}^{*}y_{\mathrm{s}}^{*}n_{\mathrm{rs}};$$
4. vypočítáme výraz $=\frac{1}{n}\cdot\sum_{i=1}^{\mathrm{n}}\left(x_{\mathrm{i}}y_{\mathrm{i}}\right)$ a spolu s vypočtenými hodnotami $\overline{x}$, $\overline{y}$, s_{x} a s_{y} dosadíme do vztahu (60) a získáme tak koeficient korelace.

3.5 Normální rozdělení

3.5.1 Úvod

Normální rozdělení (někdy též uváděné jako Gaussovo rozdělní) je jedno z nejdůležitějších rozdělení pravděpodobností spojité náhodné veličiny. Jeho důležitost vyplývá z tzv. **centrální limitní věty**, která zhruba řečeno tvrdí, že součet (nebo aritmetický průměr) velkého počtu libovolných vzájemně nezávislých a nepříliš se vzájemně lišících náhodných veličin se vždy podobá normálnímu rozdělení náhodné veličiny.

Za takové rozdělení lze považovat např. naměřené údaje dané fyzikální veličiny. Tyto údaje budou „skoro stejné“, takže se navzájem nebudou příliš odlišovat, naměřené údaje budou „kolísat na obě strany“ od skutečné hodnoty měřené veličiny a současně bude (pro korektní proměření dané veličiny) provedeno dostatečné množství měření. Ačkoliv naměřené hodnoty budou diskrétní (budou to „jednotlivá čísla“), bude mít taková sada naměřených dat přibližně normální rozdělení, které je definováno (viz dále) pro spojité funkce.

Proto normální rozdělení za určitých podmínek poměrně dobře aproximuje jiná pravděpodobnostní rozdělení (a to jak diskrétních, tak i spojitých veličin), ačkoliv málokteré rozdělení pravděpodobností náhodných veličin vyskytující se v praxi je přesně normální rozdělení.

Náhodné chyby způsobené velkým počtem malých, neznámých a vzájemně nezávislých příčin, mají v důsledku centrální limitní věty také přibližně normální rozdělení. Proto se normálnímu rozdělení také někdy říká **zákon chyb**.

3.5.2 Hustota pravděpodobnosti

Budeme-li uvažovat veličinu *X* se spojitým rozdělením pravděpodobnosti, může tato veličina nabývat spojitě rozložených hodnot. To znamená, že příznivými případy pro vyšetřování pravděpodobnosti budou všechny hodnoty *x* z určitého intervalu $\langle a; b\rangle$. Při vyšetřování pravděpodobnosti $P(x)$, že veličina *X* nabývá hodnot $x\in\langle a; b\rangle$ nastanou ale potíže. Má-li být „součet“ pravděpodobností $P(x)$ pro $x\in\langle a; b\rangle$ roven jedné, nastane problém. Hodnot *x* z daného intervalu je nekonečně mnoho (dokonce nespočetně mnoho) a i kdyby byla pravděpodobnost $P(x)$ jakkoliv minimální, nebude součet všech uvažovaných pravděpodobností nikdy konečný. Proto musí být jednotlivé pravděpodobnosti $P(x)$ nulové. Tím ale nezískáme žádné zajímavé informace.

Proto je nutné rozdělení náhodné veličiny *X* charakterizovat nikoliv pravděpodobností, ale **hustotou pravděpodobnosti**.

HUSTOTA PRAVDĚPODOBNOSTI $w(x)$ **VELIČINY** X **V BODĚ** x **INTERVALU** $\langle a; b\rangle$ **JE DEFINOVÁNA VZTAHEM**

$$w(x) = \lim_{\Delta x \to 0} \frac{\Delta p(x)}{\Delta x}, \tag{62}$$

KDE $\Delta p(x)$ JE PRAVDĚPODOBNOST, ŽE VELIČINA X NABÝVÁ LIBOVOLNOU HODNOTU Z INTERVALU $\langle x; x+\Delta x\rangle$.

V PŘÍPADĚ, ŽE $x = a$ RESP. $x = b$, SE JEDNÁ VE VZTAHU (62) O LIMITU ZPRAVA RESP. ZLEVA.

Situace je podobná jako při popisu hmotnosti jednotlivých částí tělesa, u kterého předpokládáme spojité rozložení hmotnosti v jeho objemu. Ptát se, jaká je hmotnost jednoho bodu tělesa, nedává smysl, protože hmotnost jednoho bodu tělesa by byla nulová. Je ale zajímavé, ptát se na hmotnost $\Delta m(\vec{r})$ určité části tělesa o objemu ΔV umístěného v bodě s polohovým vektorem $\vec{r}$. (Bude-li uvažovanou částí tělesa malý kvádr, pak polohový vektor bude odpovídat jednomu z vrcholů tohoto uvažovaného kvádru.) Pro hustotu $\rho(\vec{r})$ tělesa v bodě popsaném polohovým vektorem $\vec{r}$ pak platí vztah $\rho(\vec{r}) = \lim_{\Delta V \to 0} \frac{\Delta m(\vec{r})}{\Delta V}$, což je analogie vztahu (62).

Funkce $w(x)$ přestavuje spojité rozložení náhodné veličiny *X*, protože určuje hustotu pravděpodobnosti pro libovolnou hodnotu x náhodné veličiny *X*. Pro další úvahy je vhodné předpokládat, že funkce $w(x)$ je na intervalu $\langle a; b\rangle$ spojitá.

3.5.3 Distribuční funkce

Distribuční funkce (funkce rozdělení, zleva kumulovaná pravděpodobnost) je funkce udávající pravděpodobnost, že hodnota veličiny *X* je menší nebo rovina než zadaná hodnota.

DISTRIBUČNÍ FUNKCE $F(x)$ VELIČINY X V BODĚ x INTERVALU $\langle a; b\rangle$ JE DEFINOVÁNA VZTAHEM

$$F(x) = \int_a^x w(t)\,\mathrm{d}t. \tag{63}$$

Vztah (63) tedy udává pravděpodobnost, že hodnota veličiny *X* leží v intervalu $\langle a; x\rangle$, což je v souladu s úvodní větou této kapitoly. Na intervalu $(a; b)$ je tedy hustota pravděpodobnosti $w(x)$ derivací distribuční funkce $F(x)$.

Tato skutečnost vyplývá přímo ze vztahu (63). Fakt, že jsme se při popisování funkce $w(x)$ jako derivace funkce $F(x)$ omezili na otevřený interval, souvisí se spojitostí a derivací funkce (abychom nemuseli vyšetřovat krajní body uvažovaného intervalu zvlášť).

Vzhledem k tomu, že integrál ve vztahu (63) udává pravděpodobnost, že hodnota veličiny *X* leží v intervalu $\langle a; x\rangle$, pak

$$F(b) = \int_a^b w(t)\,\mathrm{d}t = 1, \tag{64}$$

protože integrál z hustoty pravděpodobnosti na intervalu $\langle a; b\rangle$ udává pravděpodobnost, že hodnota veličiny *X* bude ležet právě v tomto intervalu; jedná se tedy o jistý jev.

Z hlediska teorie funkcí je třeba si uvědomit, že vztah (64) představuje vlastně **normovací podmínku**.

Normovací podmínka zaručuje, že pravděpodobnost popsaná vlastně funkcí $F(x)$ bude nabývat „běžných hodnot“, tj. bude ležet v intervalu $\langle 0; 1\rangle$.

Vztah (64) současně říká, že funkce $w(x)$ je na svém definičním oboru omezená.

Integrál ve vztahu (64) lze interpretovat také tak, že určuje obsah plochy pod grafem funkce $w(x)$ na intervalu $\langle a; b\rangle$. Je-li uvažovaný obsah plochy je konečné číslo, je sama funkce $w(x)$ omezená.

Je vhodné distribuční funkci dodefinovat na množinu všech reálných čísel, protože s takovými funkcemi se snáze pracuje. Abychom dodrželi její význam (udává pravděpodobnost), nabízí se toto dodefinování:

$$F(x)=0 \text{ pro } x\in(-\infty;a), \tag{65}$$

$$F(x)=1 \text{ pro } x\in(b;\infty).$$

Analogicky, jako byly definovány charakteristiky polohy a variability pro diskrétní veličiny (viz kapitola 3.3), lze tyto charakteristiky definovat i pro spojité veličiny.

STŘEDNÍ HODNOTA $\overline{x}$ VELIČINY X JE DEFINOVÁNA VZTAHEM

$$\overline{x}=\int_a^b x\cdot w(x)\,\mathrm{d}x. \tag{66}$$

ROZPTYL s^2 VELIČINY X JE DEFINOVÁN VZTAHEM

$$s^2=\int_a^b (x-\overline{x})^2\cdot w(x)\,\mathrm{d}x. \tag{67}$$

MODUS VELIČINY X (TJ. NEJPRAVDĚPODOBNĚJŠÍ HODNOTA VELIČINY X) JE TAKOVÁ HODNOTA VELIČINY X, PRO NÍŽ JE HUSTOTA PRAVDĚPODOBNOSTI NEJVĚTŠÍ.

S využitím distribuční funkce F lze definovat (a to snáze než u veličin s diskrétním rozdělením) medián veličiny X. Pro medián $\mathrm{Med}(x)$ platí

$$F(\mathrm{Med}(x))=\frac{1}{2}. \tag{68}$$

Analogicky je možné definovat p-kvantil x_p veličiny X:

$$F(x_\mathrm{p})=p, \tag{69}$$

který je definován pro libovolnou hodnotu $p\in(0;1)$.

Vztahy (68) a (69) jsou definovány korektně díky vlastnostem distribuční funkce F resp. funkce hustoty pravděpodobnosti w. Funkce w je nezáporná spojitá funkce definovaná na uzavřeném intervalu. Proto je i distribuční funkce F na stejném uzavřeném intervalu spojitá a rostoucí a nabývá hodnot z intervalu $\langle 0;1\rangle$ (viz definice funkce F a její dodefinování vztahy (65)). Taková funkce podle jedné z vět o spojitých funkcí na uzavřeném intervalu nabývá všech mezihodnot mezi minimální a maximální funkční hodnotou na uvažovaném uzavřeném intervalu. Proto na intervalu $\langle a;b\rangle$ existuje alespoň jedna hodnota x_p splňující vztah (69). Díky faktu, že funkce F je navíc rostoucí (ryze monotónní na intervalu $\langle a;b\rangle$), existuje hodnota x_p právě jedna.

Proto je vztah (69) definován korektně; vztah (68) je pak jeho speciálním případem.

3.5.4 Normální rozdělení

Nyní se již můžeme věnovat normálnímu rozdělení náhodné veličiny.

VELIČINA X S NORMÁLNÍM ROZDĚLENÍM JE TAKOVÁ NÁHODNÁ VELIČINA, JEJÍŽ HUSTOTA PRAVDĚPODOBNOSTI JE DEFINOVÁNA VZTAHEM

$$w(x)=\frac{1}{\sigma\cdot\sqrt{2\pi}}\cdot\mathrm{e}^{-\frac{(x-x_0)^2}{2\sigma^2}}, \tag{70}$$

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

Grafem této funkce je tzv. **Gaussova křivka** (viz obr. 24, obr. 26 a obr. 28).

DISTRIBUČNÍ FUNKCE VELIČINY X S NORMÁLNÍM ROZDĚLENÍM JE FUNKCE TVARU

$$F(x)=\int_{-\infty}^{x}\frac{1}{\sigma\cdot\sqrt{2\pi}}\cdot\mathrm{e}^{-\frac{(t-x_0)^2}{2\sigma^2}}\,\mathrm{d}t. \tag{71}$$

Distribuční funkce normálního rozdělení se často nazývá **errorfunkce** a je běžnou součástí řady počítačových programů.

Ve vztazích (70) a (71) je

$$\sigma = \sqrt{s^2} \tag{72}$$

směrodatná odchylka a $x_0 = \overline{x}$ je střední hodnota veličiny X.

Grafy hustot pravděpodobností a distribučních funkcí jsou zobrazeny na obrázcích:

1. pro $x_0 = 1$ a $\sigma = 0{,}8$ na obr. 24 a obr. 25;
2. pro $x_0 = 1$ a $\sigma = 0{,}2$ na obr. 26 a obr. 27;
3. pro $x_0 = 0$ a $\sigma = 1$ na obr. 28 a obr. 29.

Pro $x_0 = 0$ a $\sigma = 1$ dostáváme tzv. **normované normální rozdělení** (**standardizované normální rozdělení**) s hustotou pravděpodobnosti danou vztahem

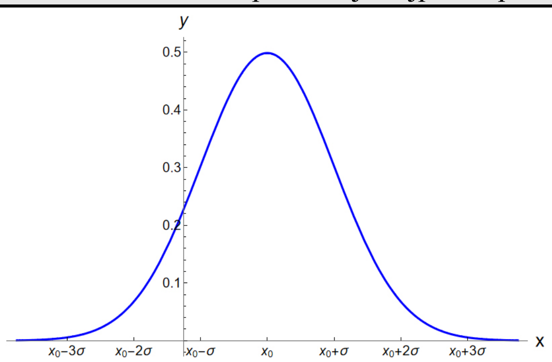
$$w(x) = \frac{1}{\sqrt{2\pi}} \cdot \mathrm{e}^{-\frac{x^2}{2}} . \tag{73}$$

Normované normální rozdělení pro náhodnou veličinu Y lze získat z normálního rozdělení veličiny X transformací podle vztahu

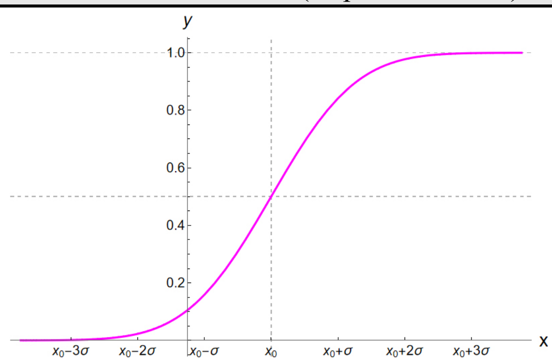
$$y = \frac{x - x_0}{\sigma} . \tag{74}$$

Jak je vidět z grafů funkcí hustoty pravděpodobnosti (obr. 24, obr. 26 a obr. 28), které se navzájem liší hodnotou směrodatné odchylky resp. rozptylu, určuje rozptyl míru rozptýlenosti jednotlivých měření (tj. jak jsou data „rozptýlena“ do stran od střední hodnoty).

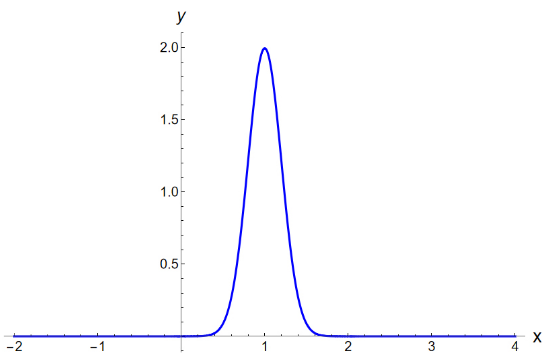
Střední hodnota pak určuje nejpravděpodobnější hodnotu daného měření (resp. souboru dat).



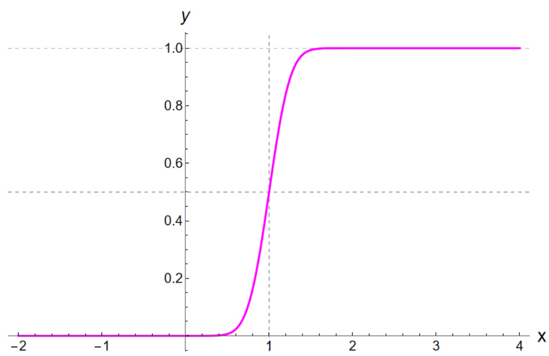
obr. 24



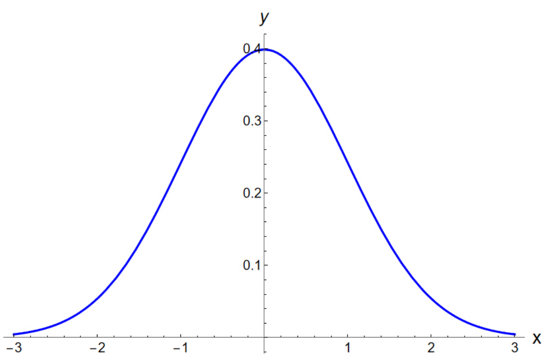
obr. 25



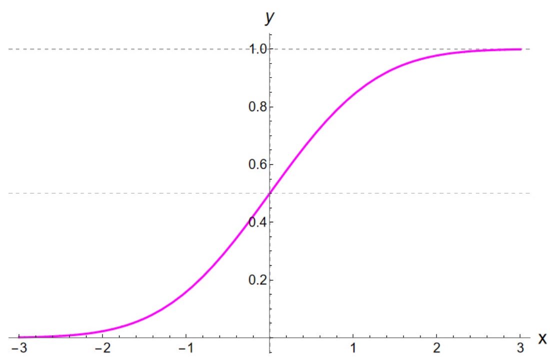
obr. 26



obr. 27



obr. 28



obr. 29

Pro normální rozdělení lze vypočítat pravděpodobnost, že hodnota náhodné veličiny X leží v daném intervalu ohraničeném násobky směrodatné odchylky σ:

1. pro $x \in (x_0 - \sigma; x_0)$ nebo $x \in (x_0; x_0 + \sigma)$ je $p(x) = \int_{x_0-\sigma}^{x_0} w(x)\,\mathrm{d}x = \int_{x_0}^{x_0+\sigma} w(x)\,\mathrm{d}x = 0{,}341$;
2. pro $x \in (x_0 - 2\sigma; x_0 - \sigma)$ nebo $x \in (x_0 + \sigma; x_0 + 2\sigma)$ je $p(x) = \int_{x_0-2\sigma}^{x_0-\sigma} w(x)\,\mathrm{d}x = \int_{x_0+\sigma}^{x_0+2\sigma} w(x)\,\mathrm{d}x = 0{,}136$;
3. pro $x \in (x_0 - 3\sigma; x_0 - 2\sigma)$ nebo $x \in (x_0 + 2\sigma; x_0 + 3\sigma)$ je $p(x) = \int_{x_0-3\sigma}^{x_0-2\sigma} w(x)\,\mathrm{d}x = \int_{x_0+2\sigma}^{x_0+3\sigma} w(x)\,\mathrm{d}x = 0{,}021$.

Výše uvedené integrály není možné počítat analyticky, ale je nutné řešit numericky. Proto jsou hodnoty těchto integrálů pro rozumné dělení osy x (viz grafy na obr. 24, obr. 26, obr. 28 a obr. 30) uvedeny v tabulkách pro využití při řešení matematických úloh i pro zpracování různých měření.

Ze symetrie grafu hustoty pravděpodobnosti (viz grafy na obr. 24, obr. 26, obr. 28 a obr. 30) vyplývá, že

$$w(x_0 - x) = w(x_0 + x)\,. \tag{75}$$

Analogicky pro distribuční funkci platí

$$F(x_0 - x) = 1 - F(x_0 + x)\,. \tag{76}$$

Vztahy (75) a (76) platí pro $x \in (-\infty; \infty)$ a speciálně platí i pro normované normální rozdělení, tj. pro rozdělení, kde $x_0 = 0$.

Vypočtené pravděpodobnosti lze vizualizovat i graficky. Výše uvedené integrály totiž určují obsah plochy pod grafem funkce hustoty pravděpodobnosti na uvedených intervalech. Vzhledem k tomu, že integrály mají význam pravděpodobnosti, lze vypočtené hodnoty integrálů chápat jako procentuální vyjádření pravděpodobnosti (viz obr. 30).

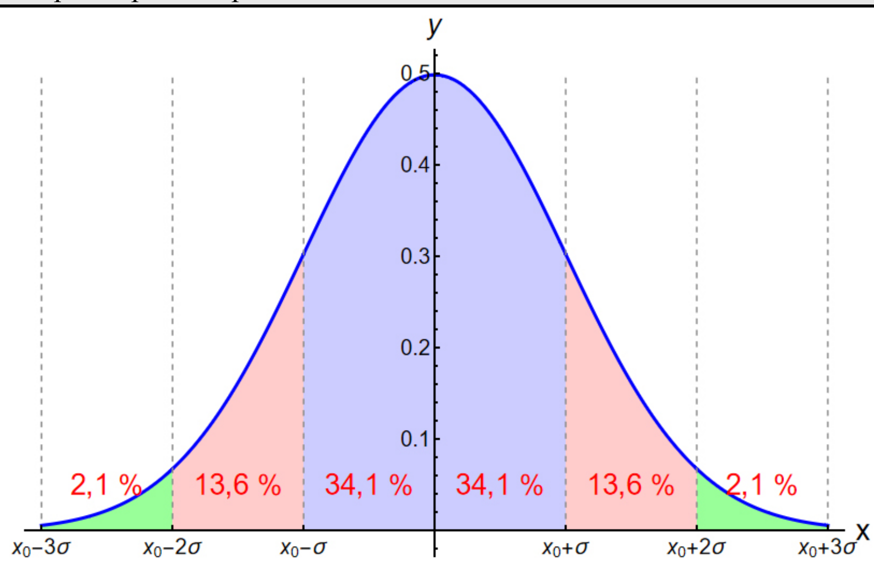
Z doplněného grafu zobrazeného na obr. 30 je patrné, že např. pravděpodobnost nalezení hodnoty veličiny X v intervalu $(x_0 - \sigma; x_0)$ je 0,341.

V intervalu $x \in (x_0 - 3\sigma; x_0 + 3\sigma)$ pak nalezneme hodnotu veličiny X s pravděpodobností 0,996. Tento interval je někdy též charakterizován **krajní chybou** κ, přičemž platí

$$\kappa = 3\sigma\,. \tag{77}$$

Ačkoliv graf zobrazený na obr. 30 odpovídá normovanému normálnímu rozdělení, výše uvedené výpočty pravděpodobností (resp. obsahů ploch pod grafem funkce hustoty pravděpodobnosti) platí obecně pro libovolné normální rozdělení.

Obecné normální rozdělení má oproti normovanému normálnímu rozdělení graf funkce hustoty pravděpodobnosti pouze posunut po ose x.



obr. 30

Této skutečnosti se využívá v matematice a dalších vědních disciplínách (fyzika, elektrotechnika, chemie, …), v nichž se využívá skutečnosti, že výše uvedené pravděpodobnosti jsou tabelované.

Kromě směrodatné odchylky σ se v některých aplikacích požívá i **pravděpodobná chyba** ϑ určující interval $(x_0 - \vartheta; x_0 + \vartheta)$, ve kterém leží hodnota veličiny X s pravděpodobností 0,5. S využitím numerických metod lze odvodit

$$\vartheta = 0{,}67449\,\sigma \doteq \frac{2}{3}\sigma\,. \tag{78}$$

3.5.5 Ukázka fyzikálních dat

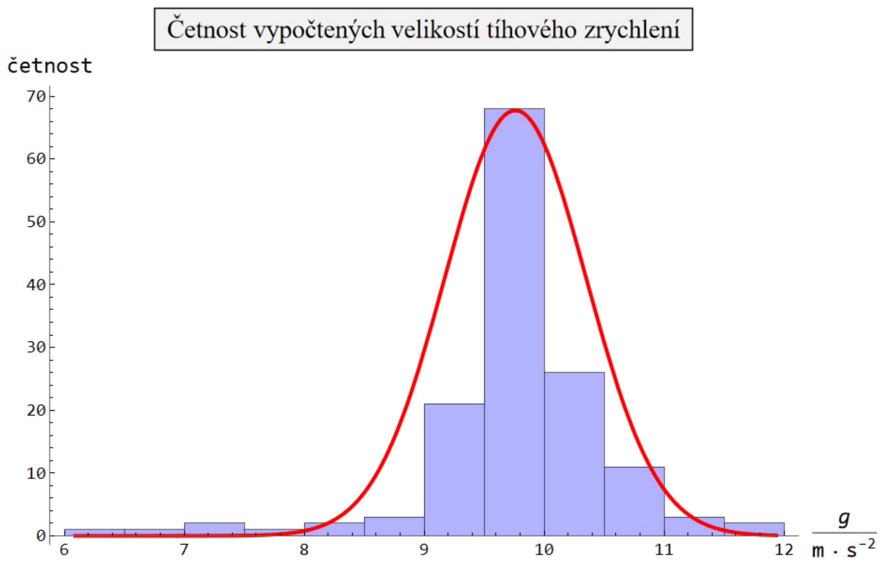
Jedním z mnoha měření, která lze v rámci středoškolské fyziky se žáky provést, je proměření parametrů matematického kyvadla. Jedná se o těleso, které má v dané soustavě zanedbatelné rozměry a které kýve se zanedbáním odporových sil na závěsu zanedbatelné hmotnosti.

Za matematické kyvadlo tedy lze považovat např. tenisák zavěšený na niti, ocelovou matici zavěšenou na niti, horolezce visícího na laně ze skalního převisu, …

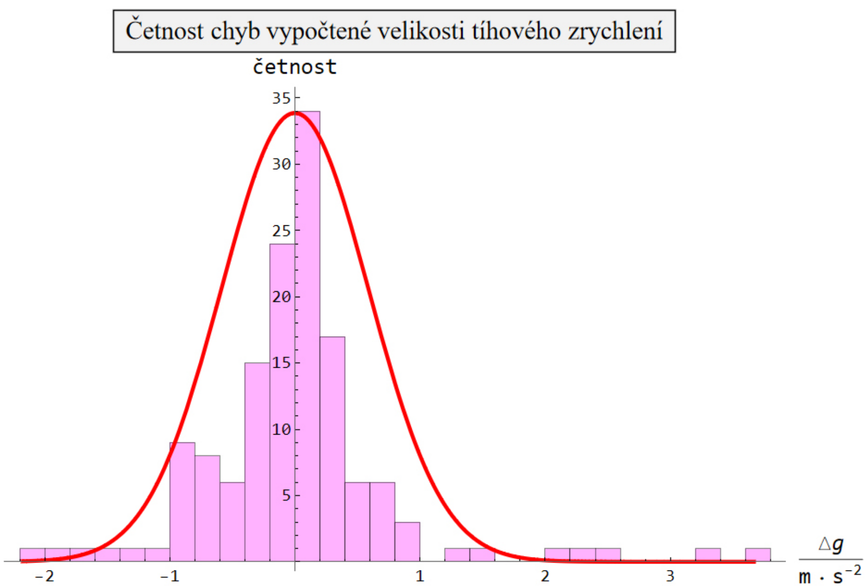
Perioda T kývajícího matematického kyvadlo s délkou závěsu l je dána vztahem

$$T = 2\pi\sqrt{\frac{l}{g}}\,, \tag{79}$$

kde g je velikost místního tíhového zrychlení; běžně udávaná velikost přitom je $g \doteq 9{,}81\,\mathrm{m\cdot s^{-2}}$.



obr. 31



obr. 32

Na základě proměření délek různých závěsů matematického kyvadla a jim odpovídajících period kmitání lze s využitím vztahu (79) vypočítat velikost místního tíhového zrychlení. Toto měření autor textu se žáky v hodinách fyziky provádí. Z dat získaných od žáků během několika let lze vypočtené hodnoty velikostí tíhového zrychlení zobrazit v histogramu zobrazeném na obr. 31. V grafu je kromě histogramu zobrazena i funkce hustoty pravděpodobnosti definovaná vztahem (70), která příslušným datům odpovídá.

V grafu zobrazeném na obr. 32 jsou pak zobrazeny četnosti chyb výpočtu velikosti tíhového zrychlení spolu s funkcí hustoty pravděpodobnosti tohoto rozdělení.

Je patrné, že oba typy zobrazených dat mají přibližně normální rozdělení.

S využitím zobrazených dat vychází velikost tíhového zrychlení $g = \left(9,8 \pm 0,5\right)\,\mathrm{m}\cdot\mathrm{s}^{-2}$.