

ZÁKLADY INFORMATIKY

1. [Úvod do informatiky](#) - pojem informace, vznik a vývoj teorie informace, osobnosti, přístupy, důvody pro vznik teorie informace.
2. Matematický aparát v teorii informace I. - [teorie pravděpodobnosti](#), náhodný jev, základní axiomy pravděpodobnosti, nezávislost a neslučitelnost, podmíněná pravděpodobnost, geometrická pravděpodobnost, Bayesův vzorec.
3. Matematický aparát v teorii informace II. - [kombinatorika](#) - permutace, variace, kombinace. [Základy statistiky](#) - náhodná veličina a její typy, hustota pravděpodobnosti, pravděpodobnostní funkce, distribuční funkce. Typy a grafické vyjádření rozdělení pravděpodobností.
4. [Číselné soustavy](#) - binární, oktálová, hexadecimální - převody mezi soustavami, aritmetika v soustavách. Polyadické a nepolyadické číselné soustavy.

1. písemná práce - test

Úvod do informatiky

- *pojem informace*
- *vznik a vývoj teorie informace*
- *osobnosti*
- *přístupy*
- *důvody pro vznik teorie informace*

Matematický aparát v teorii informace

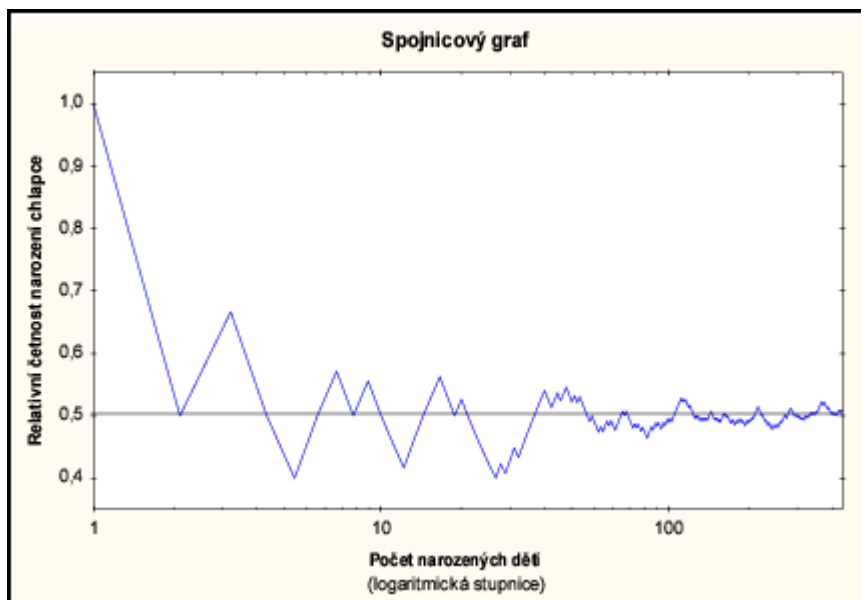
Teorie pravděpodobnosti

- náhodný jev
- základní axiomy pravděpodobnosti
- nezávislost a neslučitelnost
- podmíněná pravděpodobnost
- geometrická pravděpodobnost
- Bayesův vzorec

V praxi se často setkáváme s pokusy, jejichž výsledky nejsou jednoznačně předurčeny podmínkami, za kterých probíhají. Takové pokusy, které jsou (alespoň teoreticky) neomezeně mnohokrát opakovatelné, nazýváme **náhodné pokusy**. Jednoduchými příklady náhodných pokusů jsou: házení hracími kostkami nebo mincemi, tahání losů z osudí, míchání karet. I tyto jednoduché náhodné pokusy hrají důležitou roli v mnoha reálných situacích. Setkáváme se s nimi ve výběrových šetřeních, kdy třeba prostý náhodný výběr provádíme vhodnou technikou losování. Můžeme pak určit pravděpodobnosti, s jakými dostáváme výběry strukturou odlišné od struktury základního souboru (populace). Avšak teprve náhodné pokusy jiného druhu činí z počtu pravděpodobnosti důležitou vědu. Téměř všechny experimenty prováděné v biologii a medicíně mají svou náhodnou stránku a počtem pravděpodobnosti se zabýváme proto, abychom ji vhodně vyjádřili či zachytili. Přístupným způsobem pro pracovníky v biomedicínských oborech jsou širší základy teorie pravděpodobnosti sepsány v publikaci K. Zváry a J. Štěpána [22].

Náhodným jevem rozumíme jakékoli tvrzení o výsledku, o kterém lze po uskutečnění pokusu či pozorování rozhodnout, zda je či není pravdivé. Náhodné jevy označujeme velkými písmeny latinské abecedy, ke kterým můžeme připisovat indexy. Zabýváme-li se třeba náhodným jevem A "narození chlapce", můžeme říci o tomto jevu v okamžiku početí poměrně málo, ba ani těsně před porodem není předpověď pohlaví budoucího novorozence příliš spolehlivá. Předpokládejme, že postupně zaznamenáváme pohlaví narozených dětí a dostáváme následující posloupnost: $A, \neg A, A, \neg A, \neg A, A, A, \neg A, \dots$, kde $\neg A$ je náhodný jev, že se chlapec nenarodí, tj. "narození dívky". Četnost, s jakou nastává náhodný jev A pro libovolně dlouhou posloupnost pozorování, můžeme charakterizovat podílem r/n , kde n je délka posloupnosti (rozsah výběru) a r je počet narozených chlapců. Číslo r nazýváme **absolutní četnost** a podíl r/n **relativní četnost** výskytu náhodného jevu A ve výběru o rozsahu n . Grafické znázornění relativních četností výskytu náhodného jevu A (narození chlapce) v závislosti na rozsahu výběru n je uvedeno na obrázku 3.1. Vidíme, že se vzrůstajícím rozsahem výběru se relativní četnosti ustalují v blízkosti hodnoty 0,5.

Obrázek 3.1: Relativní četnost jevu "narození chlapce" v závislosti na celkovém počtu novorozenců (logaritmická stupnice)



Na základě tohoto příkladu si lze představit, že každý náhodný jev A je charakterizován číslem $P(A)$, které je mírou častosti výskytu tohoto jevu a nazývá se **pravděpodobnost** náhodného jevu A . V následujících bodech shrneme základní vlastnosti pravděpodobnosti.

- Pravděpodobnost $P(A)$ náhodného jevu A nabývá hodnot mezi nulou a jedničkou, tj. $0 \leq P(A) \leq 1$. V případě, že A je **jistý jev** (jev A nastává vždycky), je pravděpodobnost $P(A) = 1$. V případě, že A je **nemožný jev** (jev A nikdy nenastane), je pravděpodobnost $P(A) = 0$.
- Při mnohonásobném nezávislém opakování náhodného pokusu je prakticky jisté, že se relativní četnost výskytu náhodného jevu A jen nepatrně liší od pravděpodobnosti $P(A)$.
- Když se pravděpodobnost $P(A)$ jen nepatrně liší od nuly, je prakticky jisté, že při jediném pokusu jev A nenastane.
- Když se pravděpodobnost $P(A)$ jen nepatrně liší od jedné, je prakticky jisté, že při jediném pokusu jev A nastane.

V praxi pravděpodobnosti náhodných jevů odhadujeme pomocí relativních četností. Kvalita těchto odhadů vzrůstá s rostoucím počtem provedených pokusů. Podrobnější informace z teorie pravděpodobnosti přístupnou formou lze nalézt v [22].

Všimněme si statistických údajů, ze kterých odhadujeme pravděpodobnost náhodného jevu "narození chlapce" v naší populaci. V tabulce 3.1 jsou uvedeny relativní četnosti "narození chlapce" určené z celkového počtu živě narozených dětí v Československu v letech 1966-1975. Je patrné, že relativní četnosti jevu "narození chlapce" se v jednotlivých letech téměř neliší. Lze tedy usuzovat, že ani pravděpodobnost, že narozené dítě bude chlapec, se průběhem let nemění a je o něco vyšší než 0,51.

Tabulka 3.1: Relativní četnosti "narození chlapce" z celkového počtu živě narozených dětí v Československu v letech 1966-1975

Rok	Počet živě narozených dětí	Relativní četnost jevu "narození chlapce"
1966	222 615	0,5131
1967	215 985	0,5146
1968	213 807	0,5138
1969	222 934	0,5145
1970	228 531	0,5125
1971	237 242	0,5134
1972	251 455	0,5133
1973	274 703	0,5144
1974	291 367	0,5135
1975	289 425	0,5108
Celkem	2 448 064	0,5133

V praxi většinou nesledujeme výskyt jednoho náhodného jevu, ale zajímáme se o více jevů současně a o jejich vzájemné interakce. Ze souvislosti pravděpodobnosti a relativní četnosti můžeme odvodit základní pravidla pro počítání s pravděpodobnostmi. Vzájemné vztahy mezi náhodnými jevy jsou vyjádřeny v následujících symbolech. Náhodný jev $C = (A, B)$ nastává v případě, že nastanou oba jevy A a B současně. Náhodný jev $D = (A \text{ nebo } B)$ nastává v případě, že nastane alespoň jeden z jevů A a B , tj. buď jev A , nebo jev B , či oba jevy A a B současně (tj. jev C). Pravděpodobnost jevu $D = (A \text{ nebo } B)$ dovedeme vyjádřit pomocí pravděpodobností jevů A , B a $C = (A, B)$ jako

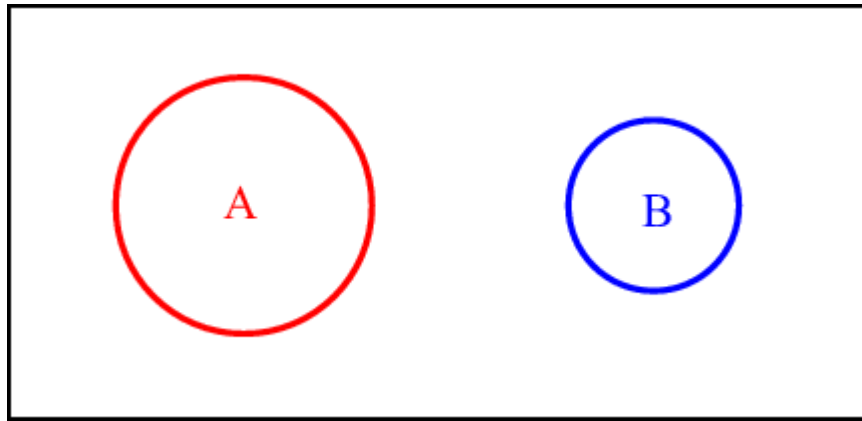
$$P(A \text{ nebo } B) = P(A) + P(B) - P(A, B) \quad (3.1)$$

Příklad 3.1 V náhodně vybrané skupině 140 mužů ve věku 40-50 let ohrožených ateriální hypertenzí se vyskytl rizikový faktor "zvýšený cholesterol" (jev A) ve 37 případech a rizikový faktor "kouření" (jev B) v 96 případech. Ve 31 případech jsme zjistili současný výskyt obou rizikových faktorů. Odhadněte pomocí relativních četností pravděpodobnosti výskytu jevů A , B , $C = (A, B)$ a $D = (A \text{ nebo } B)$.

Řešení: Pravděpodobnost výskytu faktoru "zvýšený cholesterol" je odhadnuta jako $P(A) = 37/140 = 0,2643$ a faktoru "kouření" jako $P(B) = 96/140 = 0,6857$. Odhad pravděpodobnosti současného výskytu obou faktorů je $P(A, B) = 31/140 = 0,2214$. Odhad pravděpodobnosti výskytu "zvýšeného cholesterolu" nebo "kouření" je $P(A \text{ nebo } B) = 0,2643 + 0,6857 - 0,2214 = 0,7286$. ◇

Často nám pro objasnění vzájemných souvislostí mezi pravděpodobnostmi náhodných jevů pomáhá jejich grafické znázornění pomocí Vennových diagramů. Na obrázcích [3.2](#) až [3.4](#) je znázorněno, že uvnitř obdélníku leží všechny možné výsledky náhodných pokusů či pozorování. Kruh s označením A reprezentuje jen takové výsledky, které vytvářejí jev A , podobně kruh s označením B reprezentuje výsledky, které vytvářejí jev B .

Obrázek 3.2: Neslučitelné jevy



Dva jevy A a B jsou *neslučitelné*, jestliže nemohou nastat oba současně (viz obrázek [3.2](#)). Současný výskyt jevů A a B je vyjádřen jevem $C = (A, B)$. Pro neslučitelné jevy A a B je jev $C = (A, B)$ nemožný a jeho pravděpodobnost je rovna nule, tj. $P(A, B) = 0$. Proto pravděpodobnost výskytu alespoň jednoho ze dvou neslučitelných jevů A a B se rovná součtu jejich pravděpodobností

$$P(A \text{ nebo } B) = P(A) + P(B) \quad (3.2)$$

Tento vztah se nazývá *pravidlem o sčítání pravděpodobností*.

Speciálním případem dvou neslučitelných jevů jsou jevy opačné, např. "narození chlapce" a "narození dívky". Při určování pohlaví novorozence jev "narození chlapce" nastává vždy, když nenastane jev "narození dívky". Obecně rozumíme *opačným (doplňkovým) jevem* k jevu A takový jev (značíme ho $\neg A$), který nastává právě tehdy, když jev A nenastává. Tedy $P(A \text{ nebo } \neg A) = 1$ a z pravidla o sčítání pravděpodobností dostaneme

$$P(\neg A) = 1 - P(A). \quad (3.3)$$

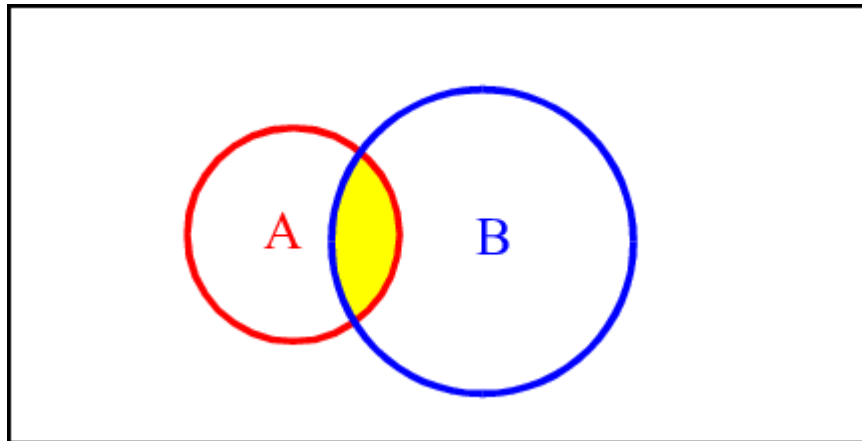
Příklad 3.2 Jestliže pravděpodobnost jevu A "narození chlapce" je rovna $P(A) = 0,51$, spočtěte pravděpodobnost jevu $\neg A$ "narození dívky".

Řešení: Jevo $\neg A$ "narození dívky" je opačným jevem k jevu A "narození chlapce". Proto $P(\neg A) = 1 - P(A) = 1 - 0,51 = 0,49$. \diamond

Pravidlo o sčítání pravděpodobností lze snadno rozšířit na libovolný počet vzájemně neslučitelných jevů A_1, A_2, \dots, A_k . Označíme-li D výskyt aspoň jednoho z těchto jevů, tj. $D = (A_1 \text{ nebo } A_2 \text{ nebo } \dots \text{ nebo } A_k)$, potom pravidlo o sčítání pravděpodobností má tvar

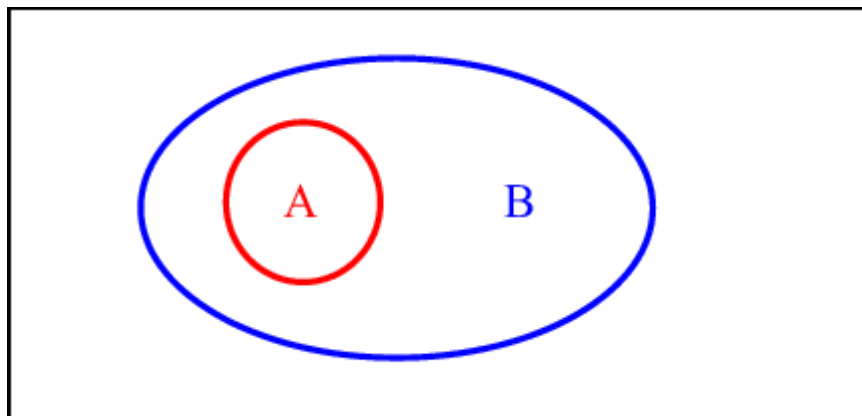
$$P(D) = P(A_1) + P(A_2) + \dots + P(A_k) = \sum_{i=1}^k P(A_i). \quad (3.4)$$

Obrázek 3.3: Slučitelné jevy



Na obrázku [3.3](#) jsou znázorněny dva jevy A a B , které se mohou vyskytovat současně. Tyto jevy tedy nejsou neslučitelné (tj. jsou *slučitelné*), a proto pro výpočet $P(A \text{ nebo } B)$ nelze použít pravidlo o sčítání pravděpodobností, ale obecnější vzorec ([3.1](#)).

Obrázek: Jevy, pro které je podmíněná pravděpodobnost $P(A | B) = 1$



V některých situacích se zajímáme o výskyt jevu A jen v případě, že nastal určitý jev B , který má kladnou pravděpodobnost (tj. může opravdu nastat). Víme-li že nastal jev B , může se tím změnit i pravděpodobnost výskytu jevu A . Všechny jevy neslučitelné s B se stanou nemožnými a jevy deterministicky určené B se stanou jistými (viz obr. [3.4](#), kde jev A nastává vždy, když nastane jev B). Ostatní jevy se mohou vyskytnout s pravděpodobnostmi, které mohou být odlišné od původních. Pravděpodobnosti jevů, zjištěné za podmínky výskytu jevu B , se nazývají podmíněné pravděpodobnosti vzhledem k jevu B . *Podmíněná pravděpodobnost* jevu A vzhledem k jevu B je definována jako

$$P(A | B) = \frac{P(A, B)}{P(B)}. \quad (3.5)$$

Tudíž pravděpodobnost současného výskytu dvou jevů A a B lze vyjádřit jako $P(A, B) = P(A | B)P(B)$.

Příklad 3.3 Odhadněte podmíněnou pravděpodobnost výskytu faktoru "zvýšený cholesterol" (jev A) za podmínky výskytu faktoru "kouření" (jev B) z údajů uvedených v příkladu [3.1](#).

Řešení: Odhad podmíněné pravděpodobnosti $P(A | B)$ spočteme, jestliže za pravděpodobnosti jevů A a B dosadíme jejich odhady pomocí relativních četností. Dostaneme

$$P(A | B) = \frac{0,2214}{0,7000} = 0,3163$$

Stejný výsledek musíme dostat, jestliže z celkového počtu 98 případů, ve kterých nastal jev B , stanovíme počet případů, ve kterých zároveň nastal jev A . Těchto případů je 31. Odhad podmíněné pravděpodobnosti je tedy

$$P(A | B) = \frac{31}{98} = 0,3163$$

a vyjadřuje relativní četnost jevu A mezi případy, kdy nastal jev B . \diamond

Dva jevy A a B jsou **nezávislé**, jestliže výskyt jednoho jevu neovlivňuje výskyt druhého jevu. Matematické vyjádření tohoto faktu zapíšeme pomocí podmíněné pravděpodobnosti jako $P(A | B) = P(A)$ nebo obdobně $P(B | A) = P(B)$. Vidíme tedy, že pro nezávislé jevy A, B platí

$$P(A, B) = P(A)P(B) \tag{3.6}$$

Tento vztah se nazývá **pravidlem o násobení pravděpodobností**.

Příklad 3.4 Zjistěte, zda faktory A a B uvedené v příkladu [3.1](#) se vyskytují nezávisle, jestliže vypočtené relativní četnosti považujeme za skutečné pravděpodobnosti.

Řešení: V případě nezávislosti faktorů A a B platí $P(A | B) = P(A)$. Z dat příkladu [3.1](#) dostáváme $P(A | B) = 0,3163$, což se liší od pravděpodobnosti $P(A) = 0,2643$. Jevy A a B tedy nejsou nezávislé. \diamond

Příklad 3.5 Označme A jev, že "první novorozenec narozený v příštím kalendářním roce v ČR je chlapec" a B jev, že "druhý novorozenec narozený v příštím kalendářním roce v ČR je chlapec". Vyloučíme-li vícečetné porody, spočtete pravděpodobnost jevu C , že "oba novorozenci jsou chlapci" za předpokladu, že pravděpodobnost narození chlapce je 0,51.

Řešení: Výskyt jevu A neovlivňuje výskyt jevu B , tudíž jevy A a B jsou nezávislé. Pravděpodobnost jevu $C = (A, B)$, že oba novorozenci jsou chlapci, je tedy podle [\(3.6\)](#) rovna

$$P(C) = P(A)P(B) = 0,51 \cdot 0,51 = 0,2601.$$

Příklad 3.6 Pravděpodobnost jevu A "osoba má pravé oko modré" je rovna 0,3 a pravděpodobnost jevu B "osoba má levé oko modré" je také rovna 0,3. Jestliže

pravděpodobnost, že "osoba má pravé oko modré" za podmínky, že nastal jev "osoba má levé oko modré" je rovna 1, spočtete pravděpodobnost jevu C "osoba má obě oči modré".

Řešení: Jevy A a B nejsou nezávislé, neboť $P(A) = 0,3$ a $P(A|B) = 1$. Proto pravděpodobnost jevu $C = (A, B)$ spočteme jako

$$P(C) = P(A|B)P(B) = 1 \cdot 0,3 = 0,3.$$

Pravděpodobnost, že "osoba má obě oči modré" je tedy rovněž 0,3. \diamond

Pravidlo o násobení pravděpodobností lze rozšířit na libovolný počet nezávislých jevů A_1, A_2, \dots, A_k . Označíme-li C jev, který spočívá v současném výskytu těchto jevů, tj. $C = (A_1, A_2, \dots, A_k)$, potom pravidlo o násobení pravděpodobností má tvar

$$P(C) = P(A_1, A_2, \dots, A_k) = P(A_1) \cdot P(A_2) \cdot \dots \cdot P(A_k) = \prod_{i=1}^k P(A_i) \quad (3.7)$$

Příklad 3.7 Za předpokladu, že pravděpodobnost narození chlapce je 0,51, spočtete, jaká je pravděpodobnost, že v sérii čtyř po sobě narozených dětí (vícečetné porody vyloučíme), bude právě jeden chlapec.

Řešení: Označte C jev, že mezi čtyřmi novorozenci je právě jeden chlapec. Konkrétní možnosti, které vytvářejí jev C , jsou dány jevy C_1, C_2, C_3 a C_4 , kde $C_1 = (A, \neg A, \neg A, \neg A)$ je jev, kdy chlapec se narodí jako první, a podobně zbývající jevy $C_2 = (\neg A, A, \neg A, \neg A)$, $C_3 = (\neg A, \neg A, A, \neg A)$ a $C_4 = (\neg A, \neg A, \neg A, A)$ vyjadřují, v jakém pořadí se chlapec narodí. Jevy C_1, C_2, C_3 a C_4 jsou vzájemně neslučitelné. Z pravidla o sčítání pravděpodobností dostaneme

$$P(C) = P(C_1 \text{ nebo } C_2 \text{ nebo } C_3 \text{ nebo } C_4) = P(C_1) + P(C_2) + P(C_3) + P(C_4).$$

Pravděpodobnosti jevů C_1, C_2, C_3 a C_4 jsou všechny stejné a jsou vypočteny pomocí pravidla o násobení pravděpodobností. Například

$$P(C_1) = P(A)P(\neg A)P(\neg A)P(\neg A) = 0,51 \cdot 0,49 \cdot 0,49 \cdot 0,49 = 0,06.$$

Tedy $P(C) = 4 \cdot 0,06 = 0,24$ je pravděpodobnost jevu, že mezi čtyřmi novorozenci bude právě jeden chlapec. \diamond

3.3 Bayesův vzorec

Předpokládejme, že náhodné jevy B_i , kde $i = 1, 2, 3, \dots, k$, jsou vzájemně neslučitelné a v každém pokusu nastává právě jeden z nich, takže musí platit

$$P(B_1 \text{ nebo } B_2 \text{ nebo } \dots \text{ nebo } B_k) = \sum_{i=1}^k P(B_i) = 1.$$

Známe-li podmíněné pravděpodobnosti $P(A|B_i)$ jevu A za podmínky výskytu jevu B_i pro $i = 1, 2, 3, \dots, k$, potom pravděpodobnost jevu A lze vyjádřit vztahem

$$P(A) = \sum_{i=1}^k P(A, B_i) = \sum_{i=1}^k P(A | B_i)P(B_i) \quad (3.8)$$

nazývaným **pravidlo o úplné pravděpodobnosti**.

Příklad 3.8 Předpokládejme, že pravděpodobnost "úrazu" (jev A) u "dítěte" (jev B_1) je $P(A | B_1) = 0,2$, pravděpodobnost úrazu u "osoby v reprodukčním věku" (jev B_2) je $P(A | B_2) = 0,1$ a pravděpodobnost úrazu u "osoby v postreprodukčním věku" (jev B_3) je $P(A | B_3) = 0,4$. Pravděpodobnosti, že osoba bude patřit do některé z těchto skupin, jsou $P(B_1) = 0,25$, $P(B_2) = 0,60$ a $P(B_3) = 0,15$. Spočítejte pravděpodobnost úrazu v dané populaci.

Řešení: Jevy B_1 , B_2 a B_3 jsou vzájemně neslučitelné a v každém případě nastává právě jeden z nich. Ze znalosti podmíněných pravděpodobností výskytu úrazu v jednotlivých věkových kategoriích obyvatelstva a ze znalostí pravděpodobností těchto kategorií spočteme pravděpodobnost úrazu v populaci jako

$$\begin{aligned} P(A) &= P(A | B_1)P(B_1) + P(A | B_2)P(B_2) + P(A | B_3)P(B_3) = \\ &= 0,20 \cdot 0,25 + 0,10 \cdot 0,60 + 0,40 \cdot 0,15 = 0,17 \end{aligned}$$

Pravděpodobnost úrazu v populaci je tedy 17 %. \diamond

Bayesův vzorec udává, jakým způsobem vypočítáme pravděpodobnosti $P(B_j | A)$ jevu B_j za podmínky, že nastal jev A , jestliže známe **apriorní** pravděpodobnosti $P(B_i)$ a **podmíněné** pravděpodobnosti $P(A | B_j)$ pro všechny jevy B_i , $i = 1, 2, 3, \dots, k$. Bayesův vzorec má tvar

$$P(B_j | A) = \frac{P(A | B_j)P(B_j)}{\sum_{i=1}^k P(A | B_i)P(B_i)}. \quad (3.9)$$

Odvození Bayesova vzorce provedeme snadno pomocí vztahů

$$\begin{aligned} P(B_j, A) &= P(A, B_j) = P(A | B_j)P(B_j), \\ P(B_j | A) &= \frac{P(B_j, A)}{P(A)} = \frac{P(A | B_j)P(B_j)}{P(A)}, \end{aligned}$$

$$\text{kde } P(A) = \sum_{i=1}^k P(A, B_i) = \sum_{i=1}^k P(A | B_i)P(B_i).$$

Příklad 3.9 Pravděpodobnost, že "osoba je kuřák" (jev A) ve skupině "osob s chronickou bronchitidou" (jev B_1) je $P(A | B_1) = 0,75$ a pravděpodobnost, že "osoba je kuřák" ve skupině "osob bez chronické bronchitidy" (jev B_2) je $P(A | B_2) = 0,50$. Pravděpodobnost "výskytu osoby s chronickou bronchitidou" v populaci budiž $P(B_1) = 0,40$ a pravděpodobnost "výskytu osoby bez chronické bronchitidy" v populaci $P(B_2) = 0,60$. Spočítejte pravděpodobnost výskytu chronické bronchitidy u kuřáka.

Řešení: Pomocí Bayesova vzorce dostaneme, že pravděpodobnost výskytu chronické bronchitidy u kuřáka je

$$P(B_1 | A) = \frac{P(A | B_1)P(B_1)}{P(A | B_1)P(B_1) + P(A | B_2)P(B_2)} = \frac{0,75 \cdot 0,40}{0,75 \cdot 0,40 + 0,50 \cdot 0,60} = 0,50. \diamond$$

Bayesův vzorec se často používá v populačních etiologických studiích a v některých matematických modelech diagnostického, terapeutického či prognostického lékařského rozhodování. Bayesův vzorec tak nachází uplatnění v případech, kdy chceme vyhodnotit kvalitu diagnostického testu či skriningového testu (screening test). Implicitní předpoklad pro provádění skriningu je, že včasná detekce nemoci povede k příznivější prognóze nemoci, neboť bude včas zahájena léčba. Některé nemoci nejsou vhodnými kandidáty pro provádění skriningu. Aby byl skrining vhodný, musí být onemocnění závažné a léčba zahájená před rozvinutím příznaků musí být příznivější vzhledem ke snížení mortality či morbiditity než v případě, že onemocnění zachytíme již v pokročilém stadiu. Navíc prevalence nemoci v preklinickém stadiu musí být dostatečně vysoká v populaci, na které je skrining prováděn. Dalším problémem skriningu je, jak dobře vyvážit užitek včasné detekce nemoci u osob, které ji skutečně mají, a nepříznivé důsledky, které vzniknou tím, že skriningový test určí jako nemocné i ty osoby, které nemoc nemají. Ukážeme dále postupy jak hodnotit kvalitu skriningového testu pro detekci nemoci D .

Označme D^+ jev, že osoba nemoc má, a D^- jev, že osoba sledovanou nemoc nemá. Pozitivní výsledek skriningového testu označme T^+ a negativní výsledek T^- . Pravděpodobnost výskytu nemoci v populaci $P(D^+)$ se nazývá *prevalence (apriorní pravděpodobnost, pretestová pravděpodobnost)*. Naměřené kombinace výsledků skriningového testu pro nemocné a osoby bez sledované nemoci můžeme zapsat do následující tabulky (viz tab. 3.2). Přitom a je počet nemocných osob, u nichž test reagoval pozitivně, a c je počet nemocných osob, u nichž test reagoval negativně. Podobně b je počet osob bez nemoci s pozitivní reakcí na test a d je počet osob bez nemoci s negativní reakcí na test.

Tabulka 3.2: Výsledky skriningového testu

	Nemoc		
Výsledek skriningového testu	přítomna (D^+)	nepřítomna (D^-)	Celkem
T^+	a	b	$a + b$
T^-	c	d	$c + d$
Celkem	$a + c$	$b + d$	n

Senzitivita (sensitivity) a specificita (specificity) jsou dvě míry pro hodnocení skriningového testu.

Senzitivita SE je definována jako pravděpodobnost $P(T^+ | D^+)$, že test bude pozitivní u nemocných. Odhaduje se jako

$$SE = \frac{a}{a + c}. \quad (3.10)$$

Specifická SP je definována jako pravděpodobnost $P(T^- | D^-)$, že test je negativní u osob bez nemoci. Odhaduje se jako

$$SP = \frac{d}{b+d}. \quad (3.11)$$

Kromě senzitivity a specifity nás rovněž zajímají pravděpodobnosti, s jakými skrínigový test reaguje negativně u nemocných, resp. pozitivně u osob bez nemoci, nazývané nesprávná negativita (false negativity), resp. nesprávná pozitivita (false positivity).

Nesprávná pozitivita FP diagnostického testu se odhaduje jako

$$FP = \frac{b}{b+d}. \quad (3.12)$$

a **nesprávná negativita FN** diagnostického testu jako

$$FN = \frac{c}{a+c}. \quad (3.13)$$

Vidíme, že součet senzitivity a nesprávné negativity je jedna, tj. $SE + FN = 1$. Podobně i součet specifity a nesprávné pozitivity je jedna, tj. $SP + FP = 1$.

Zřejmě je žádoucí, aby skrínigový test byl vysoce senzitivní a vysoce specifický. Většinou to však není možné, a proto jde o to, jak vyvážit senzitivitu a specifitu skrínigového testu. Nesmíme také zapomínat, že náklady na skrínigový test nejsou jen náklady vztahované přímo k prováděnému skrínigu, ale také náklady, které vznikají vzhledem k dalším procedurám prováděných u těch osob, které ve skrínigovém testu reagovaly pozitivně.

Prediktivní hodnoty (predictive values) skrínigového testu měří, zda osoba podrobená skrínigovému testu je skutečně nemocná.

Prediktivní hodnota pozitivního testu PV^+ je pravděpodobnost $P(D^+ | T^+)$, že osoba je opravdu nemocná, když test reagoval pozitivně. Odhadujeme ji jako

$$PV^+ = \frac{a}{a+b}. \quad (3.14)$$

Podobně, **prediktivní hodnota negativního testu PV^-** je pravděpodobnost $P(D^- | T^-)$, že osoba nemá sledovanou nemoc při negativním výsledku testu. Odhaduje se jako

$$PV^- = \frac{d}{c+d}. \quad (3.15)$$

Senzitivita a specifická jsou charakteristiky samotného testu. Prediktivní hodnoty jsou ale velmi silně závislé na prevalenci nemoci. Podle Bayesova vzorce můžeme prediktivní hodnoty vyjádřit následovně

$$PV^+ = \frac{SE P(D^+)}{SE P(D^+) + (1 - SE)(1 - P(D^+))} \quad (3.16)$$

a

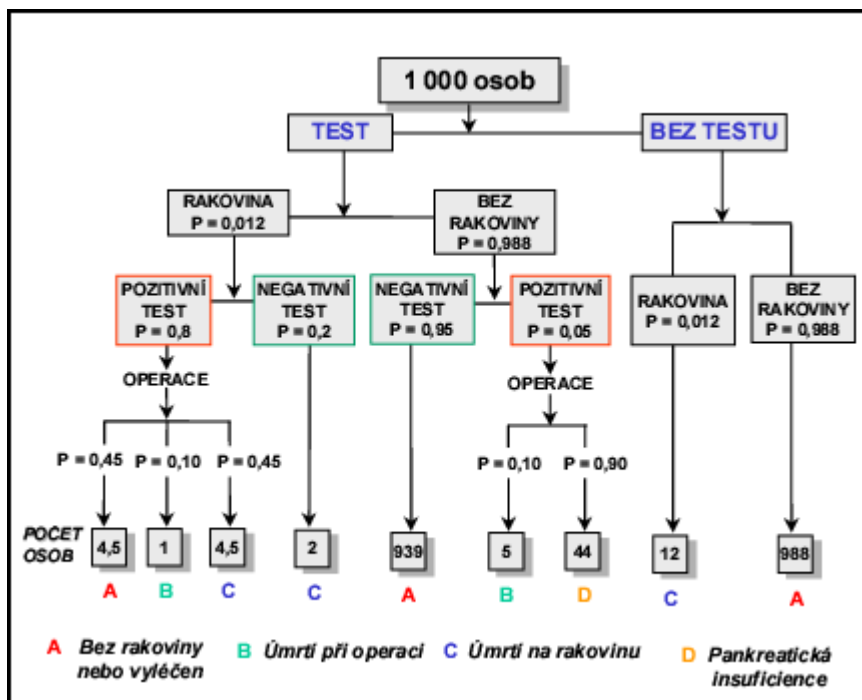
$$PV^- = \frac{SP(1 - P(D^+))}{SP(1 - P(D^+)) + (1 - SP)P(D^+)}. \quad (3.17)$$

Například použijeme-li test s 95% senzitivitou a 95% specifícou k předpovědi nemoci, která má prevalenci 1 %, potom pozitivní prediktivní hodnota je $PV^+ = 0,16$. To tedy znamená, že jenom 16 % lidí, u kterých byl test pozitivní, jsou skutečně nemocní, a 84 % z nich nemoc nemá, i když test vyšel pozitivně. Praktický význam diagnostického testu tedy záleží na senzitivitě, specifíce a prevalenci $P(D^+)$, které plně určují prediktivní hodnoty.

Když známe výsledek diagnostického testu, můžeme počítat *aposteriorní (potestovou) pravděpodobnost* výskytu nemoci D^+ . Tyto aposteriorní pravděpodobnosti jsou dány odpovídajícími prediktivními hodnotami, tj. pro pozitivní test PV^+ a pro negativní test PV^- a jsou předmětem prvořadého zájmu pro kliniku. Dobrý skrínigový či diagnostický test je takový, jehož výsledek zvyšuje kvalitu předpovědi o výskytu nemoci D^+ oproti předpovědi založené pouze na prevalenci nemoci. *Přesnost skrínigového testu* udává pravděpodobnost, s jakou test poskytuje správné závěry v populaci podrobené skrínigu. Odhadujeme ji jako podíl $(a + d)/n$.

Nevhodné užití skrínigového testu na příkladu detekce rakoviny pankreatu uvedli Sisson, Schoemaker a Ross (1976). Skrínig měl na základě výsledku testu poskytnout podklady pro rozhodnutí, zda osoba má či nemá rakovinu pankreatu. Prevalence nemoci v populaci byla $P(D^+) = 0,012$, senzitivita $SE = 0,8$ a specifická $SP = 0,95$. Pokud byla diagnostikována rakovina, byla u pacienta prováděna operace. Operace přinášela pacientovi riziko, že zemře, s pravděpodobností 0,10. Pravděpodobnost, že pacient po operaci stejně zemře na rakovinu, byla 0,45 a zlepšení jeho stavu po operaci nastalo také s pravděpodobností 0,45. V případě, že by byla operována osoba bez rakoviny, pravděpodobnost úmrtí kvůli operaci byla 0,10 a pravděpodobnost přežití po operaci, ale se zhoršeným zdravotním stavem kvůli pankreatické insuficienci, byla 0,90. Výsledky skrínigu jsou ukázány na obrázku [3.5](#). Vidíme, že pokud by se skrínig neprováděl, potom očekáváme, že z celkového počtu 1000 osob 12 osob zemře na rakovinu. V případě provedení skrínigu u 1000 osob očekávaný počet úmrtí vzroste na 12,5 a navíc 44 osob, které rakovinu pankreatu neměly, bude operováno a projeví se všechny negativní důsledky tohoto zásahu na jejich zdraví. Očekávané počty osob již nemusí být nutně celočíselné hodnoty.

Obrázek 3.5: Výsledky skrínigu



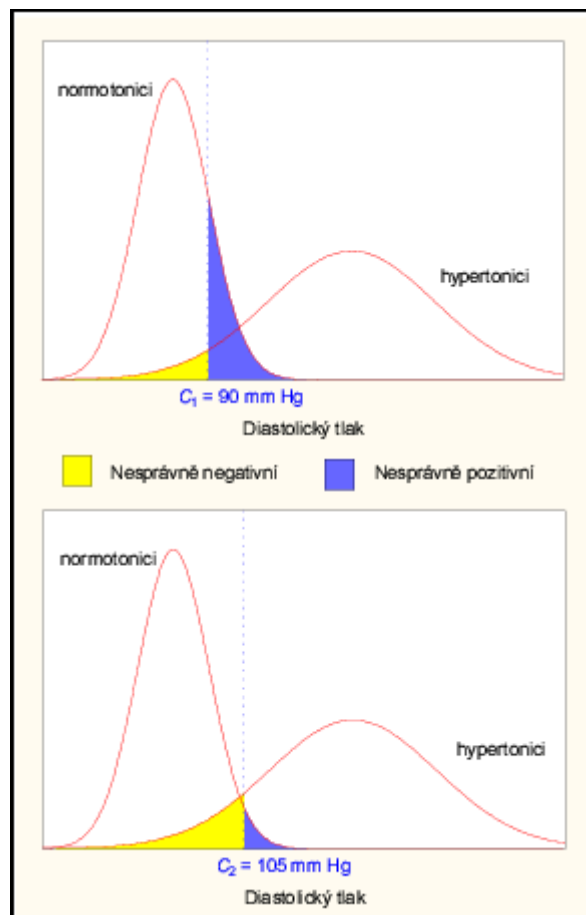
ROC křivka

Jak jsme viděli v předcházející části, senzitivita a specifita jsou charakteristiky samotného testu, ale prediktivní hodnoty jsou velmi ovlivněny tím, jak často se nemoc vyskytuje v populaci. Například test s 95% senzitivitou a 95% specificitou, který použijeme pro nemoc s prevalencí 1 %, vykáže 16% prediktivní hodnotu pozitivního testu (tedy 16 % lidí s pozitivním testem má skutečně uvažovanou nemoc), zatímco prediktivní hodnota negativního testu je 99,9 %. To znamená, že u osoby s negativním testem si můžeme být prakticky jisti, že uvažovanou nemoc nemá.

Když uvažujeme o senzitivě a specifitě jako o charakteristikách diagnostického testu, všimněme si podrobněji následujícího příkladu. Chceme zjistit, zda pacient je hypertonik, či normotonik, podle hodnoty jeho diastolického tlaku. Zvolme jako dělicí bod hodnotu $C_1 = 90$ mm Hg. Pacienty s tímto a vyšším tlakem budeme klasifikovat jako hypertoniky, ostatní pacienty jako normotoniky. Diastolický tlak lze považovat za spojitou veličinu s normálním rozdělením (viz 4.2) a jeho rozdělení ve skupině normotoniků a hypertoniků je znázorněno na obrázku 3.6. Na tomto obrázku vidíme vyznačené plochy, které ukazují podíl nesprávně pozitivních závěrů (normotonika zařadíme mezi hypertoniky) a podíl nesprávně negativních závěrů (hypertonika zařadíme mezi normotoniky). Tato situace je znázorněna

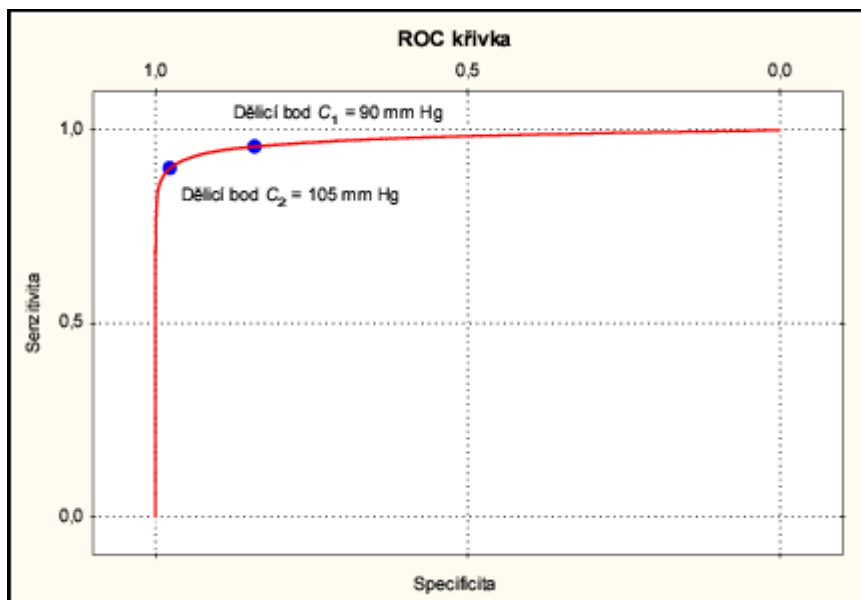
- pro dělicí bod $C_1 = 90$ mm Hg,
- pro dělicí bod $C_2 = 105$ mm Hg.

Obrázek 3.6: Grafické zobrazení vlivu dělicí hranice na nesprávně pozitivní a nesprávně negativní závěry: a) při dělicím bodu $C_1 = 90$ mm Hg je vyšší nesprávná pozitivita testu b) při dělicím bodu $C_2 = 105$ mm Hg je vyšší nesprávná negativita testu.



Naším cílem je najít takový dělicí bod, abychom docílili vhodné rovnováhy mezi nesprávně pozitivními a nesprávně negativními závěry. Z obrázku 3.6 můžeme snadno porovnat důsledky volby dělicího bodu C_1 oproti bodu C_2 . Volba dělicího bodu je tedy závislá na tom, jaké důsledky přináší nesprávná rozhodnutí a jak často se objevují. Proto při rozhodování o volbě dělicího bodu hrají roli i váhy, které přisuzujeme nesprávným rozhodnutím. Na obrázku 3.7 je nakreslena křivka, která pro zvolený dělicí bod dopočítává senzitivitu a specifitu testu. Nazývá se **ROC (Receiver Operating Characteristic) křivka**. (Nepřehlédněte poněkud nestandardní volbu měřítka na vodorovné ose!) Na ROC křivce uvádíme dva body, které přísluší zvoleným dělicím bodům $C_1 = 90 \text{ mm Hg}$ a $C_2 = 105 \text{ mm Hg}$.

Obrázek 3.7: ROC křivka a dělicí body C_1 a C_2



Šance, pravděpodobnost a věrohodnost

Řekneme-li že šance $O(A)$ (odds) závodního koně na první místo v dostihovém závodě (jev A) je 1 ku 4, znamená to, že kůň závod vyhraje s pravděpodobností $P(A) = \frac{1}{5} = 0,20$. Abychom vyjádření pomocí šance převedli na vyjádření pomocí pravděpodobnosti, sečteme vlastně čísla $1 + 4 = 5$ a dostaneme tak jmenovatel zlomku pro vyjádření pravděpodobnosti výhry, tj. $1/5$.

Pro libovolný náhodný jev A tedy platí: **šance** $O(A)$ výskytu jevu A je

$$O(A) = \frac{P(A)}{P(\neg A)} = \frac{P(A)}{1 - P(A)} \quad (3.18)$$

a

$$P(A) = \frac{O(A)}{1 + O(A)}. \quad (3.19)$$

V medicíně často používáme šance pro výpočet **podílu šancí (odds ratio) OR**, který udává podíl šance, že se vyskytne nějaký jev A za určité podmínky (jev B), k šanci, že se jev A vyskytne, když podmínka neplatí (jev $\neg B$). Podíl šancí se tedy vypočte jako

$$OR = \frac{O(A|B)}{O(A|\neg B)}, \quad (3.20)$$

kde samozřejmě

$$O(A|B) = \frac{P(A|B)}{P(\neg A|B)} \text{ a } O(A|\neg B) = \frac{P(A|\neg B)}{P(\neg A|\neg B)}.$$

Řekneme-li například, že ve statistické studii o rakovině plic bylo zjištěno, že šance na výskyt rakoviny plic (jev A) u kuřáků (jev B) je 5 ku 4 ($5/4$) a šance na výskyt rakoviny u nekuřáků (jev $\neg B$) je 1 ku 8 ($1/8$), potom podíl šancí je

$$\frac{\frac{5}{4}}{\frac{1}{8}} = \frac{4\bar{0}}{4} = 1\bar{0},$$

což znamená, že šance dostat rakovinu plic je $10\times$ větší u kuřáků než u nekuřáků.

Příbuzným pojmem k podílu šancí je **věrohodnostní poměr (likelihood ratio) LR** , který udává podíl pravděpodobnosti, že se vyskytne nějaký jev A za určité podmínky (jev B), k pravděpodobnosti, že se jev A vyskytne, když podmínka neplatí (jev $\neg B$). Má-li například pacient náhlou ztrátu paměti (jev A), chceme znát věrohodnostní poměr výskytu jevu A v případě, že má mozkový nádor (jev B), tj. podíl pravděpodobnosti, s jakou ztráta paměti vzniká při nádoru mozku, k pravděpodobnosti, s jakou vzniká v ostatních případech ($\neg B$). Věrohodnostní poměr je tedy podíl podmíněných pravděpodobností

$$LR = \frac{P(A|B)}{P(A|\neg B)}. \quad (3.21)$$

Použití věrohodnostního poměru je možná praktičtější než počítat pravděpodobnost mozkového nádoru. Věrohodnostní poměr užíváme i při hodnocení skrínigových a diagnostických testů. Například věrohodnostní poměr pozitivního skrínigového testu je dán jako $P(T^+ | D^+) / P(T^+ | D^-)$. Podobně věrohodnostní poměr negativního testu spočteme jako $P(T^- | D^+) / P(T^- | D^-)$.

Kombinatorika

- *Permutace*
- *Variace*
- *Kombinace*

Úvod

Kombinatorika hraje v rozvoji matematického myšlení výraznou roli. Její význam je zejména v rozvoji logického myšlení a obecných kombinačních schopností, v neposlední řadě ji lze považovat za základ pro následné řešení různých pravděpodobnostních problémů.

Je učivem hlavně středních škol, kde se omezuje na klasickou problematiku vytváření skupin předmětů a určování počtů všech skupin, které splňují určité podmínky. Na základní škole se s ní setkávají pouze žáci navštěvující školy s rozšířenou výukou matematiky.

Různé kombinatorické úlohy se vyskytují často v matematických olympiádách a dalších soutěžích.

Na druhém stupni nespécializovaných základních škol se kombinatorické úlohy řeší také, ale pouze intuitivně, úsudkem nebo dosazováním hodnot bez použití vzorců a obecných kombinatorických pravidel.

Co je to vlastně kombinatorika?

Kombinatorika je část matematiky, která se (jak je z názvu jasné) zabývá kombinováním všeho možného, např. můžeme při sportovním turnaji kombinovat družstva, můžeme je přiřazovat do různých skupin atd.

Vznik kombinatoriky asi nelze přesně zařadit do nějakého historického období. Vytvořila se průběžně s potřebou člověka, který chtěl znát odpovědi na svoje nejrůznější otázky, které se týkaly tohoto tématu.

Kombinatorika na rozdíl od mnohých jiných částí matematiky nepochází z Řecka. První zmínky o úlohách z kombinatoriky nacházíme v Indii. Například v lékařském spise Susruta se jeho čtenáři už v 6. století před n. l. mohli dočíst, že šesti různými základními příchutěmi se dá dosáhnout celkem 63 chutí.

Výsledky úloh v tomto období autoři nacházejí vypsáním všech možností, takže nevíme, zda znali i nějaké všeobecné vzorce. Ty už ale můžeme předpokládat u Varahamihira, který, chystajíc se vyrábět parfémy, uvažoval, že vždy když smíchá 4 ze 16 základních ingrediencí, tak dostane 1820 nadějných voňavek, což zřejmě nemohl zjistit vypisováním všech možností.

První kniha

Pak přišla mystická židovská kniha s hebrejským názvem Sefer Yetzirah. Ta tvrdila: „Ze dvou kamenů postavíš dva domy, ze třech kamenů postavíš šest domů, ... atd.“ Její autor tedy už ve 3. století našeho letopočtu nehovoří o ničem jiném, než o faktoriálech.

Mnozí další autoři židovského a islámského světa se zabývali hlavně úlohami o počtu slov, které je možné sestavit z daného počtu písmen v abecedě, stále jim však chyběla zobecnění. Až Abraham ibn Ezra (1090 – 1167), rabín žijící ve Francii, se na to zřejmě nemohl dívat, a tak pozorováním hvězd podrobně odvodil pravidlo na výpočet k -prvkových kombinací ze 7 prvků. Udělal to proto, že ho zajímal počet všech možných konjunkcí sedmi planet, které v té době podle něho pozoroval.

Od 13. století se už v mnohých pracích objevují i kombinatorické důkazy a matematici odvozují vztahy daleko složitější, než běžně používáme.

Kombinatorika v hrách

Jako matematická disciplína se kombinatorika začala objevovat přibližně v 17. století. O její rozvoj se krom jiných zasloužili Pascal, Fermat, Bernoulli, Leibniz, Euler či Laplace.

Předmětem jejich zkoumání byly hazardní hry. Velký význam měla úloha o rozdělení sázky, kterou Pascalovi předložil jeho přítel, vášnivý hráč, de Méré. Šlo o „Zápas“ hlava – orel, který se hraje do 6 vyhraných partií. Problém vznikl, když musel být přerušen v době, kdy jeden hráč měl 5 a druhý 4 vyhrané partie. Jak tedy rozdělit vsazené peníze? Bylo jasné, že rozdělení v poměru 5:4 by nebylo spravedlivé. Pascal použil metody kombinatoriky a řešil tento problém v obecném případě, kdy jednomu hráči zbývá ještě vyhrát r partií a druhému s partií. Touto úlohou se zabýval i Pierre Fermat, ale ten došel k jinému řešení.

Další rozvoj kombinatoriky je spojen se jmény Jakob Bernoulli, G. W. Leibniz a Leonhard Euler. I u nich byly hlavními aplikace na různé hry (loto, pasíans atd.).

V západních kulturách ji matematici objevili též v souvislosti s hazardními hrami. Tehdy v životě privilegovaných vrstev společnosti zaujímaly hazardní hry význačné místo. V kartách a kostkách se vyhrávaly a prohrávaly brilianty, zlato, paláce a statky, koně i drahé šperky. Také byly rozšířeny rozmanité loterie. Proto se kombinatorické úlohy zpočátku týkaly především těchto her. Řešily se například problémy kolika způsoby může při daném počtu vržených kostek padnout určitý počet ok, nebo kolika způsoby lze získat dva krále v jisté karetní hře. Tyto problémy byly hybnou silou v rozvoji nejen kombinatoriky, ale také teorie pravděpodobnosti, které se rozvíjely souběžně.

Jedním z prvních, kdo začal počítat různé kombinace při hře v kostky, byl italský matematik Niccolo Tartaglia. Sestavil tabulku, v níž je uvedeno, kolika způsoby může padnout na r kostkách s ok. Z počátku se zde vyskytovaly různé nedostatky. Nepřihlížel např. k tomu, že jeden a týž součet ok lze získat různými způsoby (např. $1+3+4=4+2+2$).

Je však dobré si uvědomit, že kombinatorika, stejně jako každá jiná matematická disciplína, je úzce propojena s ostatními disciplínami. Určitě nejvíce s pravděpodobností, která je vlastně na kombinatorice (a statistice) založená. Statistika poskytuje data a kombinatorika aparát. Ovšem není to jen pravděpodobnost, kombinatoriku můžeme použít všude, kde je to jen trochu možné. Přestavitelé nejrůznějších specializací potřebují mnohdy řešit úkoly, v nichž se zkoumají rozmanité kombinace sestavené z písmen, číslic a jiných objektů. Vedoucí dílny má například rozdělit několik druhů práce obráběcím strojům, agronom musí umístit osevy zemědělských kultur na několik polí, zástupce ředitele školy sestavit rozvrh hodin, vědec-chemik prozkoumat možná spojení mezi molekulami a atomy, lingvista uvážít různé varianty významu písmen neznámého jazyka atd.

Svémi metodami a pojmy se uplatňuje i v řadě dalších odvětví matematiky, zejména v algebře (v teorii grup a jejich reprezentací), teorii čísel, teorii pravděpodobnosti, teorii her, v geometrii (při zkoumání jejích základů), ale i v topologii a matematické analýze. Pro partie kombinatoriky rozvíjené ve 20. století se také používá název kombinatorická analýza.

Základní pojmy kombinatoriky

Kombinatorika zkoumá skupiny (podmnožiny) prvků vybraných z jisté základní množiny. Podle toho, zda se prvky v jednotlivých skupinách mohou či nemohou opakovat, rozdělujeme skupiny prvků na skupiny s opakováním a skupiny bez opakování. Dále rozlišujeme, zda vybrané skupiny jsou uspořádané či nikoli. Vybíráme tedy k prvků z daných n **konečné** množiny \mathbf{N} ($k \in \mathbf{N}$ a $n \in \mathbf{N}$) všech přirozených čísel a tvoříme (ne)uspořádané k -tice.

K nalezení všech možností využíváme základních pravidel kombinatoriky (kombinatorické pravidlo součtu a součinu), definovaných základních pojmů (permutace, variace, kombinace) nebo jednoduše výpis všech možností.

Kombinatorické pravidlo součtu

Prvním pravidlem kombinatoriky je **kombinatorické pravidlo součtu**. To je možné použít tehdy, když se nám podaří rozdělit zkoumané skupiny do několika tříd, přičemž každá skupina patří právě do jedné třídy. Je zřejmé, že pak je celkový počet skupin roven součtu počtů skupin ve všech třídách (za podmínky, že ani jedna z uvažovaných skupin nepatří do dvou nebo více tříd, tzn. že třídy jsou disjunktní).

Jestliže množina A_1 obsahuje n_1 prvků, množina A_2 má n_2 prvků, ..., množina A_k má n_k prvků a jestliže každé dvě z množin A_1, A_2, \dots, A_k , jsou disjunktní (tzn. průnik libovolných dvou množin je prázdný, tj. $A_i \cap A_j = \emptyset$ pro $i \neq j$ kde $i, j = 1, 2, \dots, k$), pak počet všech prvků

sjednocení množin $A_1 \cup A_2 \cup \dots \cup A_k = \bigcup_{i=1}^k A_i$ je roven součtu $n_1 + n_2 + \dots + n_k = \sum_{i=1}^k n_i$.

Kombinatorické pravidlo součinu

Druhé pravidlo, které nazýváme **kombinatorickým pravidlem součinu**, je poněkud složitější. Při sestavování skupin o dvou prvcích je často známo, kolika způsoby můžeme vybrat první prvek a kolika způsoby prvek druhý, přitom počet způsobů výběru druhého prvku nezávisí na tom, jak byl vybrán první prvek. Necht' první prvek je možno vybrat m způsoby a druhý prvek n způsoby. Pak skupinu těchto prvků (m, n) lze vybrat $m \cdot n$ způsoby.

Jestliže množina A_1 obsahuje n_1 prvků, množina A_2 má n_2 prvků, množina A_k má n_k prvků, pak počet všech možných uspořádaných k -tic, jejichž první složkou je libovolný prvek množiny A_1 , druhou složkou libovolný prvek množiny A_2 , ..., k -tou složkou libovolný prvek množiny A_k , je roven součinu $n_1 \cdot n_2 \cdot \dots \cdot n_k$.

Příklad 1:

Kolik existuje dvojciferných přirozených čísel takových, v nichž se nevyskytuje stejná číslice?

Řešení 1a :

Určíme, kolik existuje dvojciferných čísel a poté z nich vyloučíme ta, v nichž se vyskytuje stejná číslice (to jsou čísla 11, 22, 33, 44, 55, 66, 77, 88 a 99). Dvojciferná čísla jsou od 10 do 99, takže jich je 90 (vzali jsme čísla 1 až 99, těch je 99, a odečetli jsme 9 jednociferných čísel, ta jsou: 1, 2, 3, 4, 5, 6, 7, 8 a 9). Nyní od našich 90 dvojciferných čísel odečteme dalších 9 (to jsou čísla 11, 22, ..., 99) a dostáváme výsledek $90 - 9 = 81$.

Poznámka 1:

Z prvního řešení lze vyzorovat tzv. kombinatorické **pravidlo součtu**. Mějme konečné množiny A_1, A_2, \dots, A_n , které mají po řadě p_1, p_2, \dots, p_n prvků. Jsou-li každé dvě množiny navzájem disjunktní, tzn. neobsahují žádný společný prvek, pak počet prvků množiny $A_1 \cup A_2 \cup \dots \cup A_n$ je $p_1 + p_2 + \dots + p_n$.

Poznámka 2:

Množina je soubor nějakých objektů, například čísel. Pod množinou A_1 si tedy můžeme představit třeba všechna jednociferná čísla. Velikost této množiny je $p_1 = 9$ (čísla 1 až 9 - nulu nepočítáme). Pod množinou A_2 si můžeme představit třeba množinu všech dvojciferných čísel. Její velikost je 90 (čísla 10 až 99). Znak znamená sjednocení množin, čímž vzniká další množina. Takže množina $A_1 \cup A_2$ je množina všech dvoj nebo jednociferných čísel. Její velikost je 99 (čísla 1 až 99), a to je právě pravidlo součtu, které je navíc zobecněno pro n množin.

Řešení 1b :

Položme si otázku: kolika různými číslicemi může takové dvojciferné číslo začínat? Zřejmě 9 číslicemi (jsou to 1, 2, 3, 4, 5, 6, 7, 8 a 9). A kolika může pokračovat? No mohlo by pokračovat

deseti číslicemi (0, 1, 2, 3, 4, 5, 6, 7, 8 a 9), ale pokud chceme jen ta, ve kterých se neopakují číslice, můžeme použít všechny číslice kromě té, která je na prvním místě, tedy můžeme použít 9 číslic. Tzn. celkový počet takových dvojciferných čísel je $9 \cdot 9 = 81$.

Poznámka 3:

Z druhého řešení můžeme vypožorovat kombinatorické **pravidlo součinu**. Chceme vytvořit uspořádanou k -tici (dvojici, trojici, čtveřici, pětičici,...). Pro první člen máme na výběr z p_1 prvků, pro druhý z p_2 prvků atd., pro k -tý z p_k prvků.

Pak počet všech možných uspořádaných k -tic je $p_1 \cdot p_2 \cdot \dots \cdot p_k$. Vysvětlení: V našem příkladu jsme chtěli utvořit uspořádanou dvojici ($k = 2$). Pro první cifru jsme měli na výběr z devíti číslic, pro druhou cifru taktéž z devíti čísel. Takže počet takových uspořádaných dvojic je $9 \cdot 9 = 81$.

Permutace

Permutace bez opakování

Permutace je vlastně obměna pořadí. Představte si žáky seřazené při nástupu v tělocviku. Když navzájem prohodíme dva (nebo více) žáky, vytvořili jsme novou permutaci (novou obměnu). V tomto případě je permutace navíc charakteristická tím, že se v ní každý prvek vyskytuje právě jednou (takže ani nechybí, ani se nevyskytuje vícekrát).

Zajímavá je otázka kolik takových permutací lze vytvořit z n prvků? Pro názornost si ukážeme všechny permutace ze tří prvků ($n = 3$). Mějme například Petra (P), Martinu (M) a Hanku (H) a postavíme je do řady:

P, M, H	M, P, H	H, P, M
P, H, M	M, H, P	H, M, P

Podařilo se nám sestavit šest různých řad (šest permutací) a žádná jiná již neexistuje.

Pro zobecnění počtu permutací z n prvků použijeme pravidlo součinu. Chceme sestavit uspořádanou n -tici, přičemž máme k dispozici celkem n prvků. Ptejme se: z kolika prvků máme na výběr pro první člen n -tice? Zřejmě můžeme použít všech n prvků. Z kolika prvků máme na výběr pro druhý člen n -tice? Zřejmě všechny kromě toho prvního, tedy $(n - 1)$ prvků. Z kolika prvků máme na výběr pro třetí člen n -tice? Zřejmě všechny kromě prvního a druhého, tedy $(n - 2)$, atd. Použitím pravidla součinu zjistíme, že počet všech permutací z n prvků je $n \cdot (n - 1) \cdot (n - 2) \cdot \dots \cdot 2 \cdot 1$.

Měli bychom si počet permutací z n prvků nějak označit, protože komu by se chtělo psát tak dlouhé součiny? Většinou se tento počet označuje $P(n)$ nebo P_n . Matematici si všimli, že počet permutací z n prvků se používá až příliš často, a proto si vymysleli ještě další symbol, a sice vykřičník, tedy $P(n) = n!$ a pojmenovali ho **faktoriál čísla n** . Takže faktoriál n je definován následovně $n! = n \cdot (n - 1) \cdot (n - 2) \cdot \dots \cdot 2 \cdot 1$. Navíc ještě definovali faktoriál nuly $0! = 1$. Proč to tak udělali si povíme až později, když budeme mluvit o variacích. Aby se faktoriál nestal jen pojmem, ukažme si několik prvních faktoriálů:

$$0! = 1$$

$$1! = 1$$

$$2! = 2 \cdot 1 = 2$$

$$3! = 3 \cdot 2 \cdot 1 = 6$$

$$4! = 4 \cdot 3 \cdot 2 \cdot 1 = 24$$

$$5! = 5 \cdot 4 \cdot 3 \cdot 2 \cdot 1 = 120$$

$$6! = 6 \cdot 5 \cdot 4 \cdot 3 \cdot 2 \cdot 1 = 720$$

$$7! = 7 \cdot 6 \cdot 5 \cdot 4 \cdot 3 \cdot 2 \cdot 1 = 5\,040$$

Faktoriály velice rychle rostou, což je jeden z důvodů, proč je lepší používat zápis $100!$ než zápis číselné hodnoty tak velkého faktoriálu.

Příklad 2

Mějme 6 různých závodních aut, označme si je třeba A, B, C, D, E, F. Jaký je

a) počet všech možných pořadí, v nichž auta projedou cílem?

b) počet všech možných pořadí, v nichž auto A projede cílem dříve než B?

c) počet všech možných pořadí, v nichž auto B přijede hned po autu A?

Řešení

a) Jedná se vlastně o permutaci ze šesti prvků, takže již bez vysvětlování všech možných pořadí je $P(6) = 6! = 720$.

b) Rozdělme si všechna pořadí do dvou skupin. V první skupině budou ta, kde je A před B a ve druhé skupině ta, kde je naopak B před A. Každému pořadí z první skupiny odpovídá právě jedno pořadí z druhé skupiny takové, že prohodíme A a B, takže tyto skupiny jsou stejně velké. Odtud je jasné, že počet hledaných možností je $1/2 \cdot P(6) = 6! / 2 = 360$.

c) Uděláme takovouto úvahu. Auta A, B musí projet bezprostředně po sobě, takže je to totéž, jakoby zavedlo jen jedno auto AB, tedy počet všech takových pořadí je počet permutací z pěti prvků (ze dvou prvků A, B vznikl jediný prvek AB), což je $5! = 120$.

Příklad 3

Mějme n různobarevných korálků, které budeme navlíkat na niť. Její konce poté svážeme, takže dostaneme kruh (něco jako náhrdelník). Kolika způsoby lze korálky do kruhu uspořádat? (Uspořádání, které se liší jen otočením kruhu nepovažujeme za různé)

Řešení

Nejdříve určíme počet všech uspořádání. Tedy jako bychom korálky navlíkali do řady, nikoli do kruhu. Těch je $n!$ Ovšem několik uspořádání je kruhu shodných. Provedme následující úvahu. Uvažujme nějaké uspořádání v kruhu zvolme si libovolný korálek, o kterém prohlásíme, že je první a ostatní korálky očíslováme třeba ve směru hodinových ručiček. Teď celé uspořádání pootočíme ve směru hodinových ručiček o jeden korálek (takže první se dostane na místo druhého, druhý na místo třetího, atd.), čímž dostaneme shodné uspořádání. Takto můžeme uspořádání pootočit n -krát a vždy dostaneme shodné uspořádání. Když jsme ale korálky navlíkali do řady, všechna tato shodná uspořádání jsme započítali.

$$\text{Výsledek tedy je } \frac{n!}{n} = \frac{n \cdot (n-1) \cdot (n-2) \cdot \dots \cdot 2 \cdot 1}{n} = (n-1) \cdot (n-2) \cdot \dots \cdot 2 \cdot 1$$

Pro názornost ještě uvedeme shodná uspořádání v kruhu pro čtyři korálky:

12	41	34	23
43	32	21	14

Permutace s opakováním

Permutace k prvků s opakováním z n prvků je každá uspořádaná k -tice sestavená z těchto n prvků tak, že se v ní některé ze zvolených prvků mohou opakovat, přičemž 1. prvek se opakuje k_1 -krát, 2. prvek se opakuje k_2 -krát, ..., n -tý prvek k_n -krát.

Počet těchto permutací znamená kolik různých k -tic lze takto utvořit. Značíme ho $P'(k_1, k_2, \dots, k_n)$ a počítáme podle vzorce

$$P'(k_1, k_2, \dots, k_n) = \frac{k!}{k_1! \cdot k_2! \cdot \dots \cdot k_n!}, \text{ pro } k_i \text{ platí, že } \sum_{i=1}^n k_i = n.$$

Variace

Variace bez opakování

Variace je také, podobně jako permutace, obměna pořadí. Ovšem je zde důležitý rozdíl. V permutaci jsme měli k dispozici n prvků a vytvářeli jsme n -členné skupiny. Vytváříme-li variaci, máme k dispozici n prvků a vytváříme k -členné skupiny. Jinými slovy vybíráme k prvků ve stanoveném pořadí z daných n prvků. Je tedy vidět, že permutace je speciální případ variace, kdy

$k = n$. Většinou se požaduje, aby bylo $k \leq n$, což ale není naprosto nutné. Co by se tedy stalo, kdyby bylo $k > n$? Když si uvědomíme, že vlastně z n prvků chceme sestavit k -tici a přitom n je menší než k , je jasné, že žádnou sestavit nemůžeme. (Zkuste vytvořit ze tří číslic pěticiferné číslo - nepodaří se vám to).

Variace bez opakování je navíc charakteristická tím, že každý prvek se v ní vyskytuje nejvýše jednou. Tzn., že buď ve variaci je, nebo tam není, ale nesmí tam být víckrát. Dále se budeme zabývat případem, kdy $k \leq n$. Opět je užitečné zjistit počet všech k -členných variací. Použijeme k tomu pravidlo součinu.

Ptáme se: z kolika prvků máme na výběr pro první člen k -tice? Zřejmě můžeme použít všech n prvků. Z kolika prvků máme na výběr pro druhý člen k -tice? Zřejmě všechny kromě první, protože ten jsme již použili, tedy $(n - 1)$ prvků. Z kolika prvků máme na výběr pro třetí člen k -tice? Zřejmě $(n - 2)$ prvků. A z kolika prvků máme na výběr pro poslední k -tý člen k -tice? Zřejmě všechny kromě těch, které jsme již použili. Použili jsme jich $(k - 1)$, takže pro poslední člen k -tice můžeme vybírat ze zbylých $[n - (k - 1)]$ prvků. Odtud je již jasný počet k -členných variací z n prvků: $n \cdot (n - 1) \cdot (n - 2) \cdot \dots \cdot (n - k + 1)$.

Pro počet k -členné variace bez opakování z n prvků byl zvolen symbol $V(k, n)$ nebo též $V_k(n)$. Protože jsme již zavedli pojem faktoriálu a umíme jej vypočítat, můžeme počet variací bez opakování vyjádřit jeho použitím

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

$$\text{takže } V(k, n) = \frac{n!}{(n - k)!}$$

Tento vzorec by měl platit i pro permutace, protože, jak jsme si již řekli, permutace je jen speciální případ variace.

Dosadíme tedy $k = n$: $V(n, n) = P(n) = n! / (n - n)! = n! / 0!$, přitom ale víme, že $P(n) = n!$

Proto definujeme $0! = 1$

Poznámka:

Pro případ $k > n$ dostáváme po dosazení do vzorce $V(k, n)$ jeden činitel nulový, takže i celý součin a tím i celá variace je rovna nule. Toto odpovídá naší úvaze v úvodu odstavce o variacích. Konkrétně pro $k = 5$, $n = 3$ dostáváme $V(5, 3) = 3 \cdot 2 \cdot 1 \cdot 0 \cdot (-1) = 0$.

Příklad 4

Máme k dispozici pět barev (třeba modrá, bílá, červená, žlutá a černá) a chceme z nich vytvořit trojbarevnou vlajku, která je složená ze tří vodorovných pruhů (podobně jako ruská, německá, maďarská, lucemburská či ázerbájdžánská). Navíc chceme, aby každý pruh měl jinou barvu (rakouská vlajka nepřípadá v úvahu).

- Kolik takových vlajek můžeme sestavit?**
- Kolik z nich má bílý pruh?**
- Kolik z nich má bílý pruh uprostřed?**
- Kolik z nich nemá bílý pruh?**

Řešení

- Jedná se o trojčlennou variaci z pěti prvků, takže všech možností je $V(3, 5) = 5 \cdot 4 \cdot 3 = 60$
- Můžeme ze zbylých barev vybrat vždy dvojici, kterou pak doplníme bílou barvou. Těchto dvojic je $V(2, 4) = 4 \cdot 3 = 12$. Bílou barvu lze doplnit buďto nad oba pruhy nebo mezi ně nebo pod ně (3 možnosti umístění). Takže celkem vlajek z bílým pruhem je $3 \cdot V(2, 4) = 36$.
- Vlastně jsme tuto část již vyřešili v b). Máme jen jednu možnost, kam bílou barvu doplnit, a sice jediné mezi dvojici barev, takže je jasný výsledek 12.
- Využijeme kombinatorické pravidlo součtu. Počet vlajek, které nemají bílou barvu si označíme třeba x . Víme, že počet vlajek, které bílou barvu nemají a počet vlajek, které bílou barvu mají nám musí dát počet všech vlajek (pravidlo součtu). Vlajek, které mají

bílou barvu je 36 (viz výše) a počet všech vlajek je 60, tedy musí platit: $x + 36 = 60$, odtud $x = 24$, což je hledaný výsledek.

Variace s opakováním

k -členná variace s opakováním z n prvků je každá uspořádaná k -tice sestavená z těchto n prvků, přičemž všechny prvky v ní nemusí být různé (tj. mohou se opakovat).

Tzn. že z množiny n objektů sestavujeme všechny možné skupiny o k předmětech, tzv. k -tice. Přitom se v těchto skupinách mohou předměty téhož druhu vyskytovat i vícekrát.

Počet těchto variací znamená kolik různých k -tic lze takto utvořit. Značíme ho $V'(k, n)$ a počítáme podle vzorce $V'(k, n) = n^k$.

Kombinace

Kombinace se význačně liší od permutací a variací. V kombinacích nám totiž **nezáleží na pořadí prvků**. Vezmeme atypický příklad. Známa je např. alpská kombinace (sjezd, slalom a obří slalom). Vůbec však nezáleží na tom, která z disciplín se pojede jako první, druhá a která jako poslední. Je to naprosto jedno, což se přesně hodí ke kombinacím chápaným matematicky.

Kombinace bez opakování

K -členná kombinace bez opakování z n prvků je každá neuspořádaná k -tice (množina k prvků) vybraná z daných n prvků. V množině se žádné prvky neopakují (každý prvek se v ní vyskytuje nejvýše jednou). Opět se většinou požaduje stejně jako u variací $n \cdot k \cdot \dots$, ale není to naprosto nezbytné. Počet těchto kombinací znamená kolik různých k -tic lze takto utvořit. Značíme ho

$K(k, n)$ a počítáme podle vzorce
$$K(k, n) = \frac{V(k, n)}{k!} = \binom{n}{k} = \frac{n!}{k!(n-k)!}$$

Poznámka:

Symbol $\binom{n}{k}$ čteme „ n nad k “ a nazýváme ho **kombinačním číslem**. Pro každé $n, k \in \mathbb{N}_0$, $k \leq n$ platí:

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

Pokusme se nyní počet k -členných kombinací z n prvků odvodit.

Umíme určit počet uspořádaných k -tic, které lze sestavit z n prvků - to je variace. Ovšem několik těchto k -tic se liší pouze pořadím prvků. Kolik jich je? Vezmeme si libovolnou k -tici a vytvoříme všechny její obměny pouze s jejími prvky (tedy permutaci). Zřejmě se všemi k -ticemi, které jsme právě vytvořili, se bude lišit pouze pořadím prvků.

Odtud je již jasné, že počet všech k -členných kombinací na n prvcích je $V(k, n) / P(k)$.

Základy statistiky

- náhodná veličina a její typy
- hustota pravděpodobnosti
- pravděpodobnostní funkce
- distribuční funkce
- typy a grafické vyjádření rozdělení pravděpodobností

Náhodná veličina a rozdělení pravděpodobnosti

Statistik je ten, kdo s hlavou v rozpálené troubě a s nohama v nádobě s ledem na dotaz, jak se cítí, odpoví: "V průměru se cítím dobře."

anonym

Diskrétní a spojitá náhodná veličina

Většina náhodných pokusů a pozorování prováděných v biologii a medicíně má výsledek vyjádřený reálným číslem. Tato čísla vytvářejí hodnoty reálné náhodné veličiny. Náhodné veličiny označujeme zpravidla velkými písmeny z konce latinské abecedy, například X, Y, Z , a jejich hodnoty odpovídajícími malými písmeny, například x, y, z . K základním charakteristikám náhodné veličiny řadíme průměr (μ) a rozptyl (σ^2) - viz dále. Jak často určité hodnoty náhodné veličiny nastávají, je exaktně matematicky popsáno pomocí rozdělení pravděpodobnosti. V praxi se zpravidla setkáváme s náhodnými veličinami dvojího typu - diskrétními a spojitými náhodnými veličinami.

Diskrétní náhodná veličina X může nabýt jen konečného x_1, x_2, \dots, x_k nebo spočetného x_1, x_2, x_3, \dots počtu hodnot. Každé hodnotě x_i je přiřazena pravděpodobnost $P(X = x_i) > 0$ a součet těchto pravděpodobností pro všechny hodnoty x_i je roven jedné. Pravděpodobnosti $P(X = x_i)$ charakterizují diskrétní pravděpodobnostní rozdělení. Pro diskrétní náhodnou veličinu X s konečným počtem hodnot spočteme **průměr (střední hodnotu)**

$$\mu = \sum_{i=1}^k x_i P(X = x_i) \quad (4.1)$$

a **rozptyl**

$$\sigma^2 = \sum_{i=1}^k (x_i - \mu)^2 P(X = x_i). \quad (4.2)$$

Druhá odmocnina z rozptylu se nazývá **směrodatná (standardní) odchylka** $\sigma \geq 0$.

V případě, že diskrétní náhodná veličina nabývá spočetného počtu hodnot, dostáváme obdobně

$$\mu = \sum_{i=1}^{\infty} x_i P(X = x_i) \quad (4.3)$$

a

$$\sigma^2 = \sum_{i=1}^{\infty} (x_i - \mu)^2 P(X = x_i). \quad (4.4)$$

Příklad 4.1 Vylučovatelství skupinově specifických substancí ABH je podmíněno dominantní alelou *Se*, nevylučovatelství je podmíněno recesivní alelou *se*. Jestliže rodiče jsou heterozygotní vylučovatelé (*Se, se*), jejich potomek může být nevylučovatel (*se, se*), homozygotní vylučovatel (*Se, Se*) nebo heterozygotní vylučovatel (*Se, se*) s pravděpodobnostmi uvedenými v tabulce [4.1](#).

Tabulka 4.1: Vylučovatelství skupinově specifických substancí ABH

Genotyp potomka	Počet alel <i>Se</i>	Pravděpodobnost
Nevylučovatel (<i>se, se</i>)	0	0,25
Heterozygotní vylučovatel (<i>Se, se</i>)	1	0,50
Homozygotní vylučovatel (<i>Se, Se</i>)	2	0,25

Vidíme, že výsledky pozorování lze popsat buď pomocí nominálního znaku "genotyp potomka", rozděleného do tří tříd: nevylučovatel (*se, se*), heterozygotní vylučovatel (*Se, se*), homozygotní vylučovatel (*Se, Se*), či méně podrobně pomocí dichotomického znaku - vylučovatel, nevylučovatel. Můžeme však také sledovat kvantitativní znak "počet alel *Se*", jehož hodnoty lze interpretovat jako výsledky náhodných pozorování. "Počet alel *Se*" je diskrétní náhodná veličina X nabývající hodnot $x = 0, 1, 2$ s pravděpodobnostmi $P(X = 0) = 0,25$, $P(X = 1) = 0,5$, $P(X = 2) = 0,25$, tudíž s průměrem (střední hodnotou)

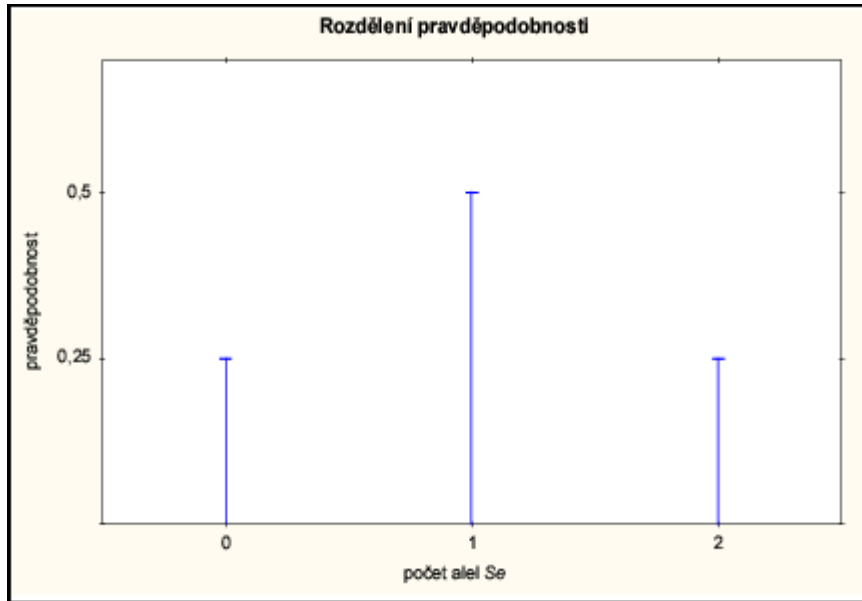
$$\mu = 0 \cdot 0,25 + 1 \cdot 0,50 + 2 \cdot 0,25 = 1$$

a rozptylem

$$\sigma^2 = (0 - 1)^2 \cdot 0,25 + (1 - 1)^2 \cdot 0,5 + (2 - 1)^2 \cdot 0,25 = 0,50.$$

Její pravděpodobnostní rozdělení je znázorněno tyčkovým grafem na obrázku [4.1](#).

Obrázek 4.1: Rozdělení pravděpodobnosti "počtu alel Se" v genotypu potomka heterozygotních vylučovatelů

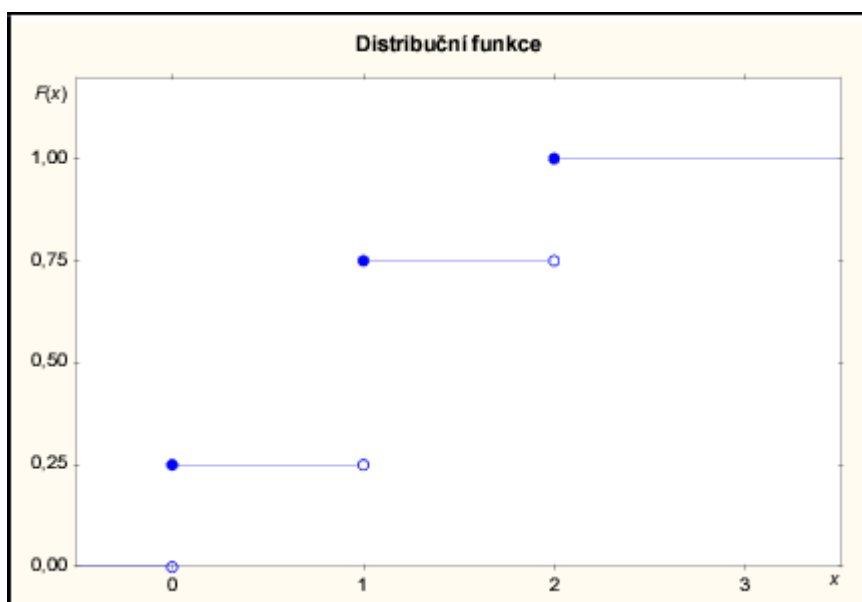


Velmi často popisujeme rozdělení pravděpodobnosti náhodné veličiny X pomocí distribuční funkce $F(x)$. *Distribuční funkce $F(x)$* vyjadřuje pro každé reálné x pravděpodobnost, že náhodná veličina X nabude hodnoty menší nebo rovné x , tedy

$$F(x) = P(X \leq x), \quad \text{pro } -\infty < x < +\infty. \quad (4.5)$$

Na obrázku [4.2](#) je graficky znázorněna distribuční funkce diskrétní náhodné veličiny X "počet alel Se". \diamond

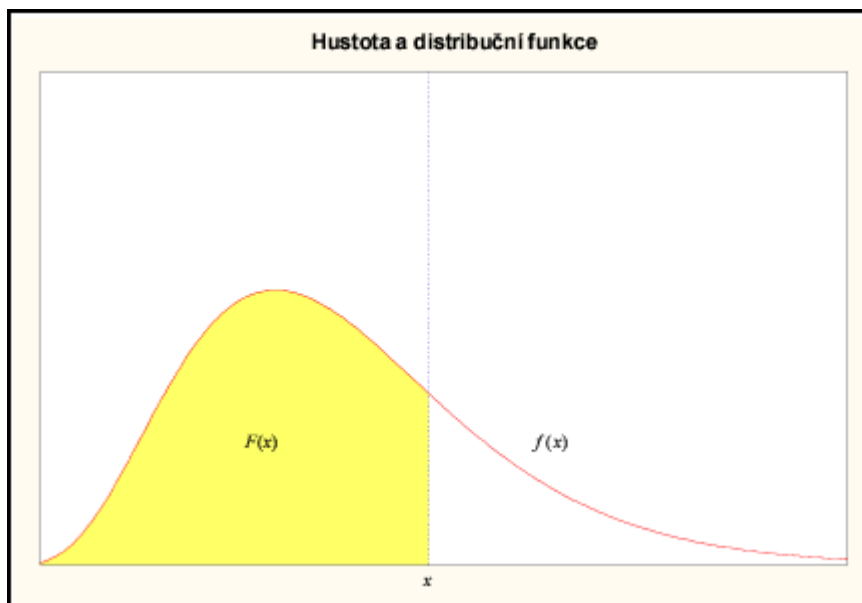
Obrázek 4.2: Distribuční funkce $F(x)$ náhodné veličiny X "počet alel Se" v genotypu potomka heterozygotních vylučovatelů



Jestliže je náhodná veličina spojitá, nemá smysl uvažovat pravděpodobnosti jednotlivých hodnot x , protože tyto pravděpodobnosti jsou obecně rovny nule. Například pravděpodobnost, že výška dospělého muže je *přesně* 180 cm, je rovna nule, protože v obecně nekonečné populaci muži s výškou 180 cm tvoří zcela zanedbatelnou část. Avšak uvažujeme-li malý interval kolem středu 180 cm, bude již nenulová pravděpodobnost, že naměříme výšku z tohoto intervalu. V praxi měření výšky provádíme většinou s přesností na centimetry, a proto naměřená výška 180 cm znamená, že skutečná výška muže leží přibližně v rozmezí od 179,5 cm do 180,5 cm.

Spojitá náhodná veličina X nabývá všech hodnot x z určitého intervalu. Její pravděpodobnostní rozdělení je popsáno pomocí reálné nezáporné funkce $f(x)$, která se nazývá *hustota (frekvenční funkce)* rozdělení. Pomocí hustoty počítáme hodnoty průměru a rozptylu spojitě náhodné veličiny. Distribuční funkci spojitě náhodné veličiny $F(x)$ lze graficky vyjádřit plochou pod hustotou $f(x)$, viz obrázek [4.3](#).

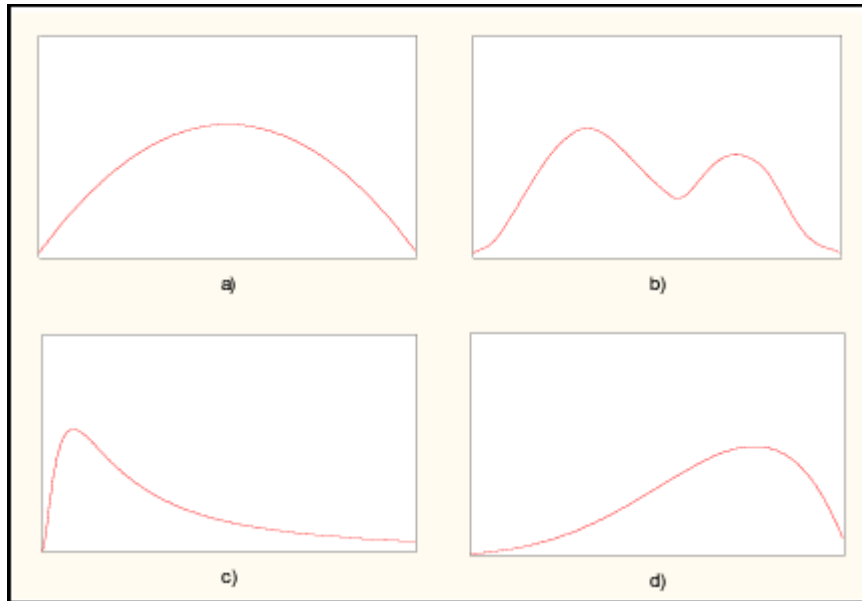
Obrázek 4.3: Velikost vybarvené plochy odpovídá hodnotě distribuční funkce F v bodě x



Podle tvaru rozeznáváme symetrické a asymetrické pravděpodobnostní rozdělení. Symetrické rozdělení často mívá "tělesná výška" ve sledované populaci osob. Asymetrické rozdělení je buď levostranně nebo pravostranně asymetrické. Pravostranně asymetrické rozdělení (má delší pravý chvost) často pozorujeme pro "tělesnou hmotnost" osob sledované populace. Pravděpodobnostní rozdělení jsou většinou jednovrcholová, ale mohou být i dvouvrcholová nebo obecně vícevrcholová. Zpravidla je vícevrcholovost způsobena nehomogenitou dat, například dvouvrcholovost pozorujeme u pravděpodobnostního rozdělení "tělesné výšky" pro populaci složenou z mužů a žen. Na obrázku [4.4](#) jsou uvedeny příklady některých spojitých rozdělení pravděpodobnosti.

Obrázek 4.4: Příklady spojitých rozdělení a) symetrické jednovrcholové

rozdělení, b) dvouvrcholové rozdělení, c) pravostranně asymetrické rozdělení, d) levostranně asymetrické rozdělení



4.2 Normální rozdění

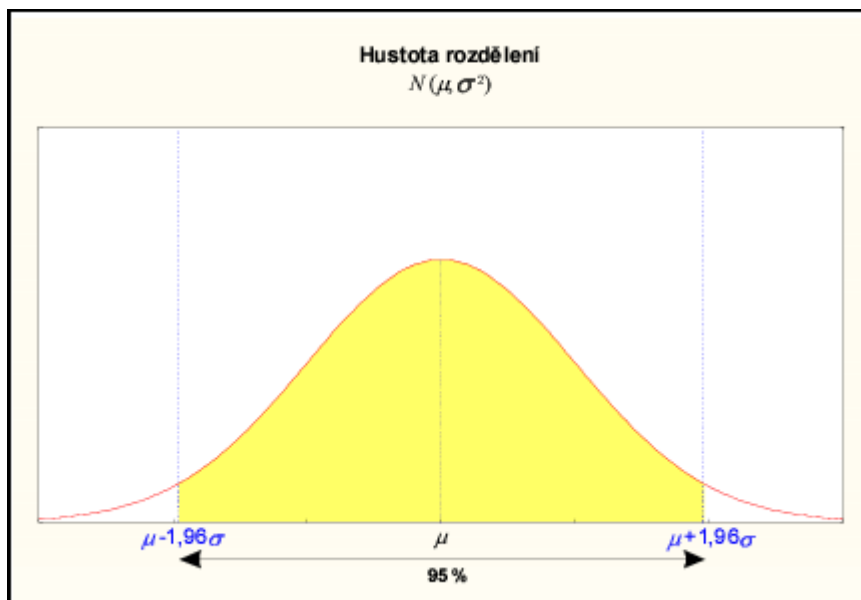
V praxi se často setkáváme s normálním rozděním u řady veličin popisujících výsledky biochemického vyšetření, např. u tělesné výšky, diastolického tlaku, vitální kapacity plic nebo sledování velikosti chyby měření. Normální rozdění je také známo jako Gaussovo rozdění (podle svého objevitele Gause). Slovo "normální" zde ovšem není použito ve svém obvyklém smyslu "obyčejné", "běžné", či v lékařském významu "bez nemoci". Jeho použití se vztahuje k staršímu významu "řídící se zákonem nebo modelem".

Hustota normálního rozdění veličiny X má tvar

$$f(x) = \frac{1}{\sigma\sqrt{2\pi}} e^{-\frac{1}{2}\left(\frac{x-\mu}{\sigma}\right)^2}, \quad (4.6)$$

kde $\pi = 3,141\dots$ a $e = 2,718\dots$ jsou matematické konstanty a μ a $\sigma > 0$ jsou konstanty určující polohu křivky (4.6) na ose x (μ) a její "roztážení" podél osy x (σ), tj. průměrnou hodnotu a míru variability. Takovým konstantám se říká *parametry*. Známe-li parametry μ a σ , je normální rozdění plně určeno. To, že veličina X má normální rozdění s průměrem μ a rozptylem σ^2 , se proto symbolicky zapisuje jako $X \sim N(\mu, \sigma^2)$. Pro veličinu X s normálním rozděním lze histogram výsledků velkého počtu n nezávislých pozorování vyrovnat křivkou (4.6) - viz obrázek 5.1.

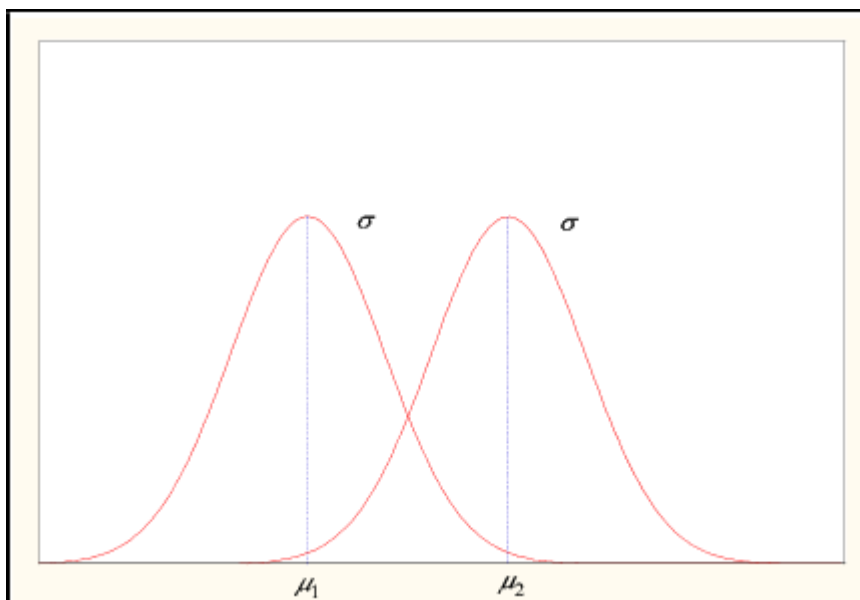
Obrázek 4.5: Hustota normálního rozdění



Grafické znázornění normálního rozdělení je dáno touto symetrickou jednovrcholovou hustotou, která je zvonovitého tvaru a nikde neprotíná vodorovnou osu (viz obrázek [4.5](#)). Parametr ležící pod vrcholem hustoty je průměr μ , parametr σ je směrodatná odchylka a jeho druhá mocnina rozptyl σ^2 (variance) veličiny X . Plocha pod křivkou hustoty normálního rozdělení je rovna jedné. Pravděpodobnost, že náhodná veličina nabude hodnot z určitého intervalu, je rovna ploše pod hustotou nad tímto intervalem. Například pro interval s hranicemi $\mu - 1,96\sigma$ a $\mu + 1,96\sigma$ má tato plocha velikost 0,95. Náhodná veličina X nabývá tedy hodnot z tohoto intervalu s 95% pravděpodobností a pouze s 5% pravděpodobností leží její hodnoty mimo uvedený interval (viz obrázek [4.5](#)).

Parametr μ , průměr náhodné veličiny, určuje polohu rozdělení na číselné ose. Na obrázku [4.6](#) jsou zakresleny dvě hustoty normálního rozdělení s různými průměry a stejnými směrodatnými odchylkami. Vzhledem k symetrii normálního rozdělení je parametr současně modem a mediánem. **Modus** je hodnota náhodné veličiny, která se vyskytuje nejčastěji. U spojitého rozdělení ji hledáme pod vrcholem hustoty. **Medián** je prostřední hodnota v tom smyslu, že rozděluje plochu pod hustotou na dvě stejně velké části. Průměr, modus a medián se nazývají **míry polohy**. Modus a medián se uplatňují zejména při popisu asymetrických a vícevrcholových rozdělení.

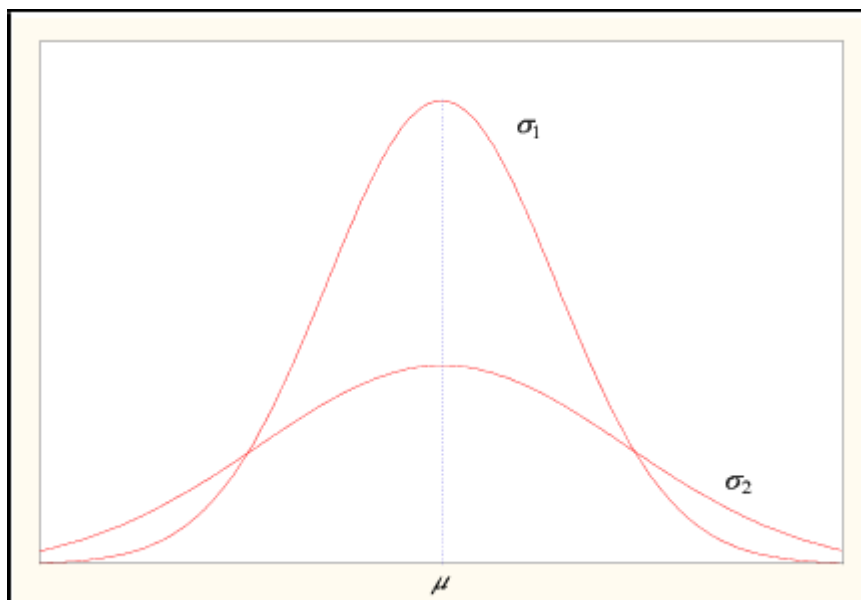
Obrázek: Hustoty normálních rozdělení s různými průměry $\mu_1 < \mu_2$ a stejnými směrodatnými odchylkami



Poznámka: Je-li veličina normálně rozdělena, pak průměr, medián a modus jsou shodné, tj. mají stejnou hodnotu. Je-li veličina zešikmena napravo (tj. má pravostranně asymetrické rozdělení neboli delší pravý chvost), pak $\text{modus} < \text{medián} < \text{průměr}$. Analogicky pro veličinu zkosenou doleva. (Viz obrázek [4.4.](#))

Parametr σ , směrodatná odchylka náhodné veličiny, určuje tvar hustoty normálního rozdělení. Čím je směrodatná odchylka větší, tím je hustota plošší, čím je menší, tím je křivka strmější. Na obrázku [4.7](#) jsou znázorněny dvě hustoty normálního rozdělení se stejnými průměry a různými směrodatnými odchylkami ($\sigma_1 < \sigma_2$). Vidíme, že pro větší směrodatnou odchylku se mohou snáze vyskytovat hodnoty náhodné veličiny vzdálenější od průměru. Proto směrodatnou odchylku σ (stejně jako rozptyl σ^2) řadíme mezi *míry variability*.

Obrázek: Hustoty normálních rozdělení se stejnými průměry a různými směrodatnými odchylkami $\sigma_1 < \sigma_2$



V normálním rozdělení:

1. téměř 70 % hodnot leží ve vzdálenosti menší než 1 směrodatná odchylka od průměru, přesněji
68,27 % leží mezi $\mu \pm 1\sigma$;
2. 95 % hodnot leží ve vzdálenosti menší než 2 směrodatné odchylky od průměru, přesněji
95 % leží mezi $\mu \pm 1,96\sigma$;
3. 99 % hodnot leží ve vzdálenosti menší než 3 směrodatné odchylky od průměru, přesněji
99 % leží mezi $\mu \pm 2,576\sigma$.

Pro srovnání variability některých pravděpodobnostních rozdělení je vhodným ukazatelem **variační koeficient**, který je dán podílem směrodatné odchylky a průměru. Je bezrozměrný a obvykle se vyjadřuje v procentech, tj. $(\sigma / \mu) \cdot 100\%$. Variační koeficient používáme pro srovnání několika náhodných veličin s velmi odlišnými průměry nebo pro srovnání variability veličin měřených v různých jednotkách.

Důležitými ukazateli variability jsou tzv. **kvantily**. 100 P% kvantil x_p je číslo, které odděluje 100 P% nejmenších hodnot náhodné veličiny X , tj. distribuční funkce v bodě x_p je $F(x_p) = P$. Tedy 50% kvantil $x_{0,50}$ je totéž co medián. Dobrý popis rozdělení pravděpodobnosti dostaneme stanovením dostatečného počtu kvantilů. Kvantily zaznamenané po dvaceti pěti procentech nazýváme **kvartily**, po deseti procentech **decily** a po jednom procentu **percentily**. Tedy

25% kvantil je 1. kvartil (dolní kvartil), 10% kvantil je 1. decil a podobně 1% kvantil je 1. percentil. Medián je totéž co 50% kvantil, 2. kvartil, 5. decil nebo 50. percentil. S použitím kvantilů, decilů a percentilů se často setkáváme při prezentaci výsledků antropometrických studií.

Pro veličinu $X \sim N(\mu, \sigma^2)$ lze vypočítat hodnotu distribuční funkce $F(x) = P(X \leq x)$ pro libovolné x pomocí hustoty normálního rozdělení. Ve statistických tabulkách nalezneme vypočtené hodnoty distribuční funkce pouze pro normální rozdělení s průměrem 0 a rozptylem 1 (viz tabulku [A.2](#)). Veličinu s tímto normálním rozdělením označíme Z , tj. $Z \sim N(0,1)$ a rozdělení nazýváme **standardizované (normované) normální rozdělení**. Distribuční funkci standardizovaného normálního rozdělení označíme $\Phi(z) = P(Z \leq z)$. Libovolnou veličinu $X \sim N(\mu, \sigma^2)$ můžeme transformovat na veličinu $Z = (X - \mu) / \sigma$, která má standardizované normální rozdělení, tj. $Z \sim N(0,1)$. Podle pravidel pro úpravu nerovností lze snadno odvodit, že

$$\begin{aligned} F(x) &= P(X \leq x) = P\left(\frac{X - \mu}{\sigma} \leq \frac{x - \mu}{\sigma}\right) = \\ &= P\left(Z \leq \frac{x - \mu}{\sigma}\right) = \Phi\left(\frac{x - \mu}{\sigma}\right). \end{aligned}$$

Příklad 4.2 Výška v populaci chlapců ve věku 3,5-4 roky má normální rozdělení s průměrem $\mu = 102$ cm a směrodatnou odchylkou $\sigma = 4,5$ cm. Spočítejte, jaké procento chlapců v uvedeném věku má výšku menší nebo rovnou 93 cm.

Řešení: Pravděpodobnost, že výška nabude hodnoty menší nebo rovné 93 cm, je vyjádřena hodnotou distribuční funkce

$$\begin{aligned} F(93) &= P(X \leq 93) = P\left(Z \leq \frac{93 - 102}{4,5}\right) = \\ &= P(Z \leq -2) = \Phi(-2). \end{aligned}$$

V tabulce [A.2](#) najdeme $\Phi(-2) = 0,0228$. Tudiž pouze 2,28 % chlapců ve věku 3,5-4 roky má výšku menší nebo rovnou 93 cm. \diamond

Jak již bylo uvedeno, řada veličin v biologii a medicíně má Gaussovo normální rozdělení. Setkáváme se však také s rozděleními jiného typu, ale někdy lze takové veličiny vhodnou transformací převést opět k normálnímu rozdělení. Příkladem je logaritmická transformace. Řekneme, že veličina X má **logaritmicko-normální rozdělení** s parametry μ a σ , tj.

$X \sim LN(\mu, \sigma^2)$, jestliže veličina $Y = \ln X$ má normální rozdělení s parametry μ a σ^2 , tj. $Y \sim N(\mu, \sigma^2)$. Ze zkušenosti víme, že logaritmicko-normální rozdělení mívá např. tělesná hmotnost, doba přežití po jedné dávce ozáření nebo minimální smrtná dávka přípravku v homogenní skupině pokusných zvířat.

Již z dat můžeme usuzovat, zda je model normálního rozdělení vhodný. Velmi často jsou data zešikmena - buď kladně (dlouhý chvost napravo) nebo záporně (dlouhý chvost nalevo). V medicíně se s tím lze setkat při měření hematologických, hormonálních nebo biologických veličin. Co tedy s takovými daty dělat?

Obvyklý přístup

1. Ztransformujeme původní veličinu (např. pomocí logaritmu, druhé odmocniny či převrácené hodnoty) na novou veličinu, pro kterou je model normálního rozdělení přijatelný.
2. Analýzu potom provedeme na transformované veličině.
3. Výsledky analýzy (např. průměry či intervaly spolehlivosti) lze pro účely prezentace výsledků zpětně transformovat.

Pokud nenajdeme vhodnou transformaci na normální rozdělení, nabízí statistika jiné přístupy založené např. na neparametrických metodách.

4.3 Binomické rozdělení

Normální rozdělení je vhodné pro *spojité* veličiny, ale často musíme pracovat s veličinami, které jsou ve své podstatě *nespojité*. Existuje několik typů diskrétních rozdělení, nejčastější z nich se nazývá *binomické*.

Uvažujeme n statisticky nezávislých pokusů. V každém pokusu může sledovaný jev buď nastat (= "úspěch") nebo nenastat (= "neúspěch"). Odpovídající pravděpodobnosti označíme π a $(1 - \pi)$ a jsou v každém pokusu stejné. V tomto případě tedy π nemá význam Ludolfova čísla (3,141...)! Celkový počet úspěchů X v n nezávislých pokusech je *binomická veličina*. Tato náhodná veličina může nabývat pouze celočíselných hodnot od 0 do n .

Je-li v každém pokusu pravděpodobnost úspěchu π , potom pravděpodobnost, že v n nezávislých pokusech nastane přesně k úspěchů, je

$$P(X = k) = \binom{n}{k} \pi^k (1 - \pi)^{n-k}, \quad \text{pro } k = 0, 1, 2, \dots, n. \quad (4.7)$$

Binomický koeficient $\binom{n}{k}$ je počet k -členných kombinací z n objektů a je definován následovně:

$$\binom{n}{k} = \frac{n!}{k!(n-k)!}, \quad (4.8)$$

kde *faktoriál* $n!$ znamená

$$n! = n(n-1)(n-2) \cdots 1. \quad (4.9)$$

Například binomický koeficient $\binom{3}{1}$ je

$$\binom{3}{1} = \frac{3!}{1!2!} = \frac{3 \cdot 2 \cdot 1}{1 \cdot 2 \cdot 1} = 3.$$

Poznámka: $1! = 1$; $0! = 1$.

Binomické rozdělení náhodné veličiny X symbolicky značíme $X \sim Bi(n, \pi)$, kde n je přirozené číslo a $0 < \pi < 1$.

Pravděpodobnosti lze počítat rovněž podle rekurzivního vzorce

$$P(X = k + 1) = \frac{n - k}{k + 1} \cdot \frac{\pi}{1 - \pi} P(X = k). \quad (4.10)$$

Jestliže spočteme pravděpodobnosti $P(X = k)$ pro všechny hodnoty k , snadno ověříme, že jejich součet je roven jedné. Podle klasické binomické věty dostaneme

$$\sum_{k=0}^n P(X = k) = \sum_{k=0}^n \binom{n}{k} \pi^k (1 - \pi)^{n-k} = (\pi + (1 - \pi))^n = 1.$$

Průměrný počet úspěchů ve výběru o rozsahu n se získá vynásobením počtu nezávislých pokusů (někdy nazývaných Bernoulliho pokusy) pravděpodobností úspěchu v jednom pokusu. Průměr binomické veličiny je tedy $n\pi$.

Rozptyl binomické veličiny je $n\pi(1 - \pi)$, kde $(1 - \pi)$ je pravděpodobnost neúspěchu. Směrodatná odchylka je druhá odmocnina z rozptylu, tedy $\sqrt{n\pi(1 - \pi)}$.

Speciálním případem binomického rozdělení $Bi(n, \pi)$ je rozdělení $Bi(1, \pi)$, které se nazývá **alternativní**.

Příklad 4.3 Předpokládejme, že pravděpodobnost narození dívky je 0,49. Jaká je pravděpodobnost toho, že mezi třemi dětmi v rodině je právě jedna dívka?

Tabulka 4.2: Parametry binomického rozdělení v příkladu 4.3

Pokus	Úspěch	Neúspěch	Pravděpodobnost úspěchu	Počet pokusů	Počet úspěchů
-------	--------	----------	-------------------------	--------------	---------------

			π	n	k
narození dítěte	dívka	chlapec	0,49	počet dětí	počet dívek

Řešení: Jak je vidět z tabulky 4.2, počet narozených dívek v rodině je náhodná veličina s binomickým rozdělením. Pravděpodobnost, že mezi třemi dětmi je právě jedna dívka, tedy vypočteme jako

$$P(X = 1) = \binom{3}{1} 0,49^1 \cdot 0,51^2 = 3 \cdot 0,127 = 0,38. \diamond$$

Při použití tohoto postupu jsme předpokládali, že pokusy jsou nezávislé, tj. znalost výsledku prvního pokusu neovlivní pravděpodobnost v druhém pokusu atd. V našem příkladu se tedy nejedná o rodinu s dvojčaty či trojčaty.

Příklad 4.4 Jaká je pravděpodobnost, že v rodině s 8 dětmi jsou právě 3 dívky? Opět vylučujeme, že některé děti jsou z vícečetných těhotenství. (Znovu předpokládejme, že pravděpodobnost narození dívky je 0,49.)

Řešení:

$$P(k = 3) = \binom{8}{3} 0,49^3 \cdot 0,51^5 = \frac{8 \cdot 7 \cdot 6}{3 \cdot 2 \cdot 1} \cdot 0,118 \cdot 0,035 = 0,23. \diamond$$

V praxi se lze vyhnout nepříjemnému počítání binomických koeficientů

1. vyhledáním koeficientu v tabulkách,
2. použitím kalkulačky,
3. použitím počítače.

4.4 Poissonovo rozdělení

Toto diskrétní rozdělení vznikne buď jako limitní případ binomického rozdělení nebo tehdy, když události nějakého druhu nastávají náhodně v čase či prostoru. Je-li pravděpodobnost (π) nějaké výjimečné události (např. určité mutace genu) relativně malá a rozsah výběru poměrně velký, pak Poissonovo rozdělení v podstatě splývá s binomickým, ale je mnohem výhodnější pro počítání. Například když n je příliš velké - řekněme v tisících - pak je velmi obtížné vypočítat binomický

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

Uvažujme náhodnou veličinu X , která představuje počet výskytů nějaké výjimečné události (např. mutace) v daném intervalu (natažení DNA). Veličina X tedy může nabývat celočíselných hodnot od 0 do nekonečna. Necht' λ je konstanta označující průměrný počet událostí v intervalu (času nebo prostoru). Potom

$$P(X = x) = \frac{e^{-\lambda} \lambda^x}{x!} \tag{4.11}$$

a X je náhodná veličina s Poissonovým rozdělením s parametrem λ , kde symbol e je základ přirozeného logaritmu, $e \doteq 2,71828$. Symbolicky Poissonovo rozdělení značíme $X \sim Po(\lambda)$, $\lambda > 0$.

Poissonovo rozdělení je charakterizováno následujícími vlastnostmi.

1. Pravděpodobnost výskytu jedné události v daném intervalu (času nebo prostoru) je úměrná délce tohoto intervalu.
2. Události se vyskytují nezávisle jak ve stejném intervalu, tak mezi po sobě jdoucími intervaly.

Připomeňme, že průměr binomické náhodné veličiny je roven $n\pi$ a rozptyl je $n\pi(1-\pi)$. Když je π velmi malé, pak $1-\pi$ je blízko 1 a $n\pi(1-\pi)$ je přibližně rovno $n\pi$. V takové situaci je tedy průměr i rozptyl rozdělení totožný a může být nahrazen jediným parametrem λ . Skutečnost, že průměr je roven rozptylu (tj. parametru λ), je charakteristickou vlastností Poissonova rozdělení.

Příklad 4.5 Předpokládejme, že v určité populaci krys se vyskytuje albín s pravděpodobností $\pi = 0,001$, ostatní krysy jsou normálně pigmentované. Ve vzorku 100 krys náhodně vybraných z této populace určete pravděpodobnost, že vzorek a) neobsahuje albína, b) obsahuje právě jednoho albína.

Řešení: Pravděpodobnost výskytu albína je $\pi = 0,001$. Předpokládaný počet albínů ve výběru o rozsahu n je $\lambda = n\pi$ (průměr binomické náhodné veličiny), tj. v našem příkladu $\lambda = n\pi = 100 \cdot 0,001 = 0,1$. Počet řídkých událostí (tj. albínů) označme x . Potom podle vzorce (4.11)

$$\text{a) pro } x = 0 \text{ máme } \frac{e^{-0,1} \cdot 0,1^0}{0!} = \frac{e^{-0,1} \cdot 1}{1} = 0,9048,$$

$$\text{b) pro } x = 1 \text{ máme } \frac{e^{-0,1} \cdot 0,1^1}{1!} = \frac{e^{-0,1} \cdot 0,1}{1} = 0,09048.$$

Jak je vidět, pravděpodobnost, že ve vzorku 100 krys nebude žádný albín, je desetkrát vyšší než pravděpodobnost, že ve vzorku bude právě jeden albín. Pravděpodobnosti výskytu dvou a více albínů jsou již velmi malé. \diamond

Číselné soustavy

- *Polyadické a nepolyadické číselné soustavy*
- *binární, oktalová, hexadecimální*
- *převody mezi soustavami*
- *aritmetika v soustavách*