# Počet pravděpodobnosti jako základ matematické statistiky

Pomocí metod popisné statistiky dokážeme přehledně shrnout informace, které se týkají výhradně objektů výběrového souboru. Pokud jsme však data získali na základě dobře navrženého výzkumného plánu, můžeme provádět induktivní úsudky o chování sledovaných proměnných v celém základním souboru. Metody statistické indukce se ovšem opírají o počet pravděpodobnosti.

## Počet pravděpodobnosti

Je to disciplína, která se zabývá studiem zákonitostí v náhodných pokusech. Matematickými prostředky modeluje situace, v nichž hraje roli náhoda. Pod pojmem náhoda rozumíme působení faktorů, které se živelně mění při různých provedeních téhož pokusu a nepodléhají naší kontrole.

Pokusem rozumíme jednorázové uskutečnění konstantně vymezeného souboru definičních podmínek. Předpokládáme, že pokus můžeme mnohonásobně nezávisle opakovat za dodržení definičních podmínek (ostatní podmínky se mohou měnit, proto různá opakování pokusu mohou vést k různým výsledkům). Dále předpokládáme, že opakováním pokusu vzniká opět pokus.
Deterministický pokus je takový pokus, jehož každé opakování vede k jedinému možnému výsledku. (Např. zahřívání vody na 100°C při atmosférickém tlaku 1015 hPa vede k varu vody.)
Náhodný pokus je takový pokus, jehož každé opakování vede k právě jednomu z více možných výsledků, které jsou vzájemně neslučitelné. (Např. hod kostkou vede k právě jednomu ze šesti možných výsledků.)

## Zavedení měřitelného prostoru

Neprázdnou množinu možných výsledků náhodného pokusu značíme $\Omega$ a nazýváme ji základní prostor. Možné výsledky značíme $\omega_1$, $\omega_2$, ...
Vymezená množina výsledků je náhodný jev, značíme ho symbolem A. Všechny možné náhodné jevy tvoří jevové pole **A**.
Dvojice ($\Omega$, **A**) se nazývá měřitelný prostor. $\Omega$ se nazývá jistý jev, $\emptyset$ nemožný jev.

## Vztahy mezi jevy

a) $A \subset B$ znamená, že jev A má za důsledek jev B.
Např. jev A je padnutí dvojky, jev B padnutí sudého čísla.

b) $A \cup B$ znamená nastoupení aspoň jednoho z jevů A, B
Např. jev A je padnutí lichého čísla, jev B padnutí šestky, $A \cup B$ znamená padnutí 1, 3, 5, 6.

c) $A \cap B$ znamená společné nastoupení jevů A, B
Např. jev A je padnutí čísla menšího než 3, jev B padnutí sudého čísla, $A \cap B$ znamená padnutí dvojky.

d) $A \setminus B$ znamená nastoupení jevu A za nenastoupení jevu B
Např. A je padnutí lichého čísla většího než 1, B padnutí trojky, $A \setminus B$ je padnutí pětky.

e) $\overline{A} = \Omega \setminus A$ znamená jev opačný k jevu A
Např. A je padnutí sudého čísla menšího než 6, $\overline{A}$ znamená padnutí lichého čísla nebo šestky.

f) $A \cap B = \emptyset$ znamená, že jevy A, B jsou neslučitelné
Např. A je padnutí jedničky, B je padnutí sudého čísla.

g) $\omega \in A$ znamená, že možný výsledek $\omega$ je příznivý nastoupení jevu A.
Např. A je padnutí násobku tří, $\omega$ je padnutí šestky.

**Některé vlastnosti operací s jevy**

a) Pro sjednocení a průnik jevů platí komutativní zákon, který pro dva jevy A, B má tvar:
$A \cup B = B \cup A$, $A \cap B = B \cap A$.

b) Pro sjednocení a průnik tří jevů A, B, C platí zákon asociativní:
$A \cup (B \cup C) = (A \cup B) \cup C$, $A \cap (B \cap C) = (A \cap B) \cap C$
distributivní: $A \cup (B \cap C) = (A \cup B) \cap (A \cap C)$, $A \cap (B \cup C) = (A \cap B) \cup (A \cap C)$

c) Pro sjednocení a průnik jevů opačných platí de Morganovy zákony, které pro dva jevy A, B zapíšeme takto: $\overline{A \cup B} = \overline{A} \cap \overline{B}$, $\overline{A \cap B} = \overline{A} \cup \overline{B}$.

**Příklad:** Náhodný pokus spočívá v hodu kostkou. Jev A znamená, že padne sudé číslo a jev B znamená, že padne číslo větší než 4.

a) Určete základní prostor $\Omega$.

b) Vypište možné výsledky příznivé nastoupení jevů A, B.

c) Pomocí operací s jevy vyjádřete následující jevy: padne liché číslo; nepadne číslo 1 ani 3, padne číslo 6; padne číslo 2 nebo 4.

**Řešení:**

ad a) $\Omega = \{\omega_1, ..., \omega_6\}$, kde možný výsledek $\omega_i$ znamená, že padne číslo i , i = 1, ..., 6

ad b) $A = \{\omega_2, \omega_4, \omega_6\}$,  $B = \{\omega_5, \omega_6\}$

ad c) $\bar{A} = \{\omega_1, \omega_3, \omega_5\}$; $A \cup B = \{\omega_2, \omega_4, \omega_5, \omega_6\}$; $A \cap B = \{\omega_6\}$; $A \setminus B = \{\omega_2, \omega_4\}$

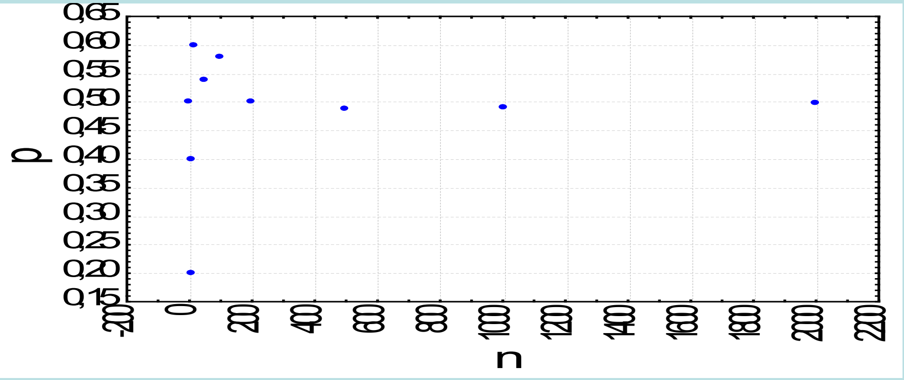
## Pravděpodobnostní prostor

**Motivace:** Provádíme opakovaně nezávisle týž náhodný pokus a v každém pokusu sledujeme nastoupení jevu A, kterému říkáme úspěch. Označme n celkový počet pokusů a N(A) počet těch pokusů, kdy nastal úspěch. S rostoucím n pozorujeme, že relativní četnost úspěchu $\frac{N(A)}{n}$ se blíží číslu P(A), které považujeme za pravděpodobnost úspěchu. (Tento poznatek je znám jako **empirický zákon velkých čísel**).

### Ilustrace empirického zákona velkých čísel

Provádíme n nezávislých hodů mincí. Padnutí líce považujeme za úspěch. Budeme sledovat závislost relativní četnosti úspěchu na počtu pokusů. (Počet pokusů volíme 2, 5, 10, 20, 50, 100, 200, 500,1000, 2000.)

| n | 2 | 5 | 10 | 20 | 50 | 100 | 200 | 500 | 1000 | 2000 |
|---|---|---|---|---|---|---|---|---|---|---|
| p | 0,5 | 0,2 | 0,4 | 0,6 | 0,54 | 0,58 | 0,5 | 0,488 | 0,49 | 0,4975 |



Vzniká otázka, jak zavést pravděpodobnost, aby byla „zidealizovaným“ protějškem relativní četnosti. Zdálo by se vhodné zavést pravděpodobnost takto:

$$P(A)=\lim_{n\to\infty}\frac{N(A)}{n}.$$

Jde o tzv. **statistickou definici pravděpodobnosti**.

**Příklad:** U 1000 náhodně vybraných voličů byly zjišťovány volební preference. Výsledky máme v tabulce:

| Preferovaná strana | pohlaví | | celkem |
|---|---|---|---|
| | žena | muž | |
| ABC | 153 | 130 | 283 |
| XYZ | 220 | 194 | 414 |
| ostatní | 157 | 146 | 303 |
| celkem | 530 | 470 | 1000 |

Pomocí relativních četností odhadněte pravděpodobnosti těchto jevů:

a) náhodně vybraný volič je žena, která nepreferuje stranu XYZ,
b) náhodně vybraný volič je buď žena nebo jedinec, který preferuje ostatní politické strany.

**Řešení:**

ad a) Všech žen je 530, těch, které nepreferují stranu XYZ, je 530 – 220 = 310, tedy odhad pravděpodobnosti daného jevu je

$$\frac{310}{1000}=0{,}31$$

ad b) Zajímá nás relativní četnost sjednocení dvou jevů, a to jevů $B_1$ … „náhodně vybraný volič je žena“ a $B_2$ … „náhodně vybraný volič preferuje ostatní politické strany“. Z vlastností relativní četnosti plyne

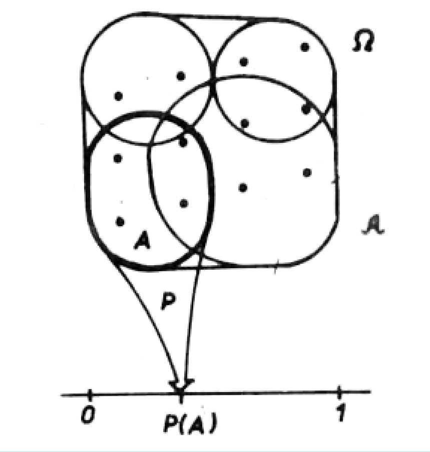
$$p(B_1\cup B_2)=p(B_1)+p(B_2)-p(B_1\cap B_2)=\frac{530}{1000}+\frac{303}{1000}-\frac{157}{1000}=\frac{676}{1000}=0{,}676$$

Z matematického hlediska však statistická definice definice není v pořádku, protože počet pokusů je vždy konečný a nelze se přesvědčit o existenci uvedené limity. Proto ve 30. letech 20. století ruský matematik A. A. Kolmogorov (1903 – 1987) vybudoval **axiomatickou teorii pravděpodobnosti.**

Axiomatická teorie pravděpodobnosti zavádí pravděpodobnost jako funkci, která každému jevu přiřazuje číslo mezi 0 a 1 a přitom je zidealizovaným protějškem relativní četnosti. Má tedy všechny vlastnosti relativní četnosti a kromě toho některé další vlastnosti, které vyplývají z vnitřních potřeb matematické teorie.

Pravděpodobností rozumíme reálnou množinovou funkci P: **A** → R, která splňuje následující tři axiómy:
každému jevu přiřazuje nezáporné číslo (axióm nezápornosti),
jistému jevu přiřazuje číslo 1 (axióm normovanosti),
sjednocení neslučitelných jevů přiřazuje součet pravděpodobností těchto jevů (axióm spočetné aditivity).
Trojice (Ω, **A**, P) se nazývá pravděpodobnostní prostor. Je to matematický model jednorázového provedení náhodného pokusu.

## Ilustrace pravděpodobnostního prostoru




Systém axiómů pravděpodobnosti je bezesporný (tj. na každém měřitelném prostoru lze sestrojit pravděpodobnost) a neúplný (tj. na každém měřitelném prostoru, jehož jevové pole není minimální, lze sestrojit pravděpodobností více).

## Vlastnosti pravděpodobnosti

Nechť $(\Omega, \mathcal{A}, P)$ je pravděpodobnostní prostor. Pak pro libovolné jevy $A, A_1, A_2, \ldots \in \mathcal{A}$ platí následujících 14 vlastností:

P1: $P(\emptyset) = 0$

P2: $P(A) \geq 0$ (nezápornost – axióm)

P3: $P(A_1 \cup A_2) + P(A_1 \cap A_2) = P(A_1) + P(A_2)$

P4: $1 + P(A_1 \cap A_2) \geq P(A_1) + P(A_2)$

P5: $P(A_1 \cup A_2) \leq P(A_1) + P(A_2)$ (subaditivita)

P6: $A_1 \cap A_2 = \emptyset \Rightarrow P(A_1 \cup A_2) = P(A_1) + P(A_2)$ (aditivita)

P7: $P(A_2 \setminus A_1) = P(A_2) - P(A_1 \cap A_2)$

P8: $A_1 \subseteq A_2 \Rightarrow P(A_2 \setminus A_1) = P(A_2) - P(A_2)$ (subtraktivita)

P9: $A_1 \subseteq A_2 \Rightarrow P(A_1) \leq P(A_2)$ (monotonie)

P10: $P(\Omega) = 1$ (normovanost – axióm)

P11: $P(A) + P(\bar{A}) = 1$ (komplementarita)

P12: $P(A) \leq 1$

P13: $A_i \cap A_j = \emptyset$ pro $i \neq j \Rightarrow P(A_1 \cup A_2 \cup \ldots) = P(A_1) + P(A_2) + \ldots$ (spočetná aditivita – axióm)

P14: $P\left(\bigcup_{i=1}^{n} A_i\right) = \sum_{i=1}^{n} P(A_i) - \sum_{i=1}^{n-1}\sum_{j=i+1}^{n} P(A_i \cap A_j) + \sum_{i=1}^{n-2}\sum_{j=i+1}^{n-1}\sum_{k=j+1}^{n} P(A_i \cap A_j \cap A_k) - + \ldots + (-1)^{n-1} P(A_1 \cap \ldots \cap A_n)$

(Pro neslučitelné jevy $A_1, \ldots, A_n$ dostáváme $P\left(\bigcup_{i=1}^{n} A_i\right) = \sum_{i=1}^{n} P(A_i)$.)

(Vlastnosti P1, ..., P12 odpovídají vlastnostem relativní četnosti z popisné statistiky, vlastnost P14 je známa jako věta o sčítání pravděpodobností.)

**Klasická pravděpodobnost**

V Kolmogorovově axiomatické definici se nic nepraví o tom, jak na daném měřitelném prostoru konkrétně pravděpodobnost zavést. V případě, že základné prostor je konečný a všechny možné výsledky mají stejnou šanci na uskutečnění, můžeme použít klasickou pravděpodobnost:

Klasická pravděpodobnost je funkce, která jevu A přiřazuje číslo $P(A) = \frac{m(A)}{m(\Omega)}$, kde m(A) je počet možných výsledků příznivých nastoupení jevu A a m(Ω) je počet všech možných výsledků.

**Příklad na klasickou pravděpodobnost** (s využitím vlastností pravděpodobnosti): V dodávce 100 kusů výrobků nemá požadovaný průměr 10 kusů, požadovanou délku 20 kusů a současně nemá požadovaný průměr i délku 5 kusů. Jaká je pravděpodobnost, že náhodně vybraný výrobek z této dodávky má požadovaný průměr i délku?

**Řešení:**

Jev A spočívá v tom, že výrobek má požadovaný průměr a jev B v tom, že výrobek má požadovanou délku. Počítáme

$P(A \cap B) = P\left(\overline{\bar{A} \cup \bar{B}}\right) = 1 - P(\bar{A} \cup \bar{B}) =$

$1 - [P(\bar{A}) + P(\bar{B}) - P(\bar{A} \cap \bar{B})] = 1 - \left(\frac{10}{100} + \frac{20}{100} - \frac{5}{100}\right) = 0{,}75.$

**Opakované závislé pokusy - hypergeometrické rozložení pravděpodobností**

Máme N objektů, mezi nimi je M objektů označeno, ($0 < M < N$). Náhodně bez vracení vybereme n objektů ($0 < n < N$).

Pravděpodobnost, že ve výběru je právě x označených objektů ($\max\{0, M-N+n\} \le x \le \min\{n, M\}$):

$$P_{N,M,n}(x) = \frac{\binom{M}{x}\binom{N-M}{n-x}}{\binom{N}{n}}.$$

Pravděpodobnost, že ve výběru je nejvýše $x_1$ označených objektů:

$$\sum_{x=\max(0,M-N+n)}^{x_1} P_{N,M,n}(x).$$

Pravděpodobnost, že ve výběru je aspoň $x_0$ označených objektů:

$$\sum_{x=x_0}^{\min(n,M)} P_{N,M,n}(x).$$

**Příklad na hypergeometrické rozložení:** Máme skupinu 20 lidí, mezi nimi 4 muže. Vybíráme bez vracení (tj. nikdo nemůže být vybrán opakovaně) pětici z této skupiny. Jaký je nejpravděpodobnější počet mužů mezi vybranými?

**Řešení:**

Vypočítáme všechny možné pravděpodobnosti a najdeme počet mužů s největší pravděpodobností. Vzhledem k technice výběru (bez vracení) jde o hypergeometrické rozložení pravděpodobností s parametry N = 20, M = 4, n = 5.

Počítáme

$$P_{N,M,n}(x) = \frac{\binom{M}{x}\binom{N-M}{n-x}}{\binom{N}{n}}$$

pro x = 0, 1, 2, 3, 4.

$$P_{20,4,5}(0) = \frac{\binom{4}{0}\binom{20-4}{5-0}}{\binom{20}{5}} \doteq 0{,}281\ldots, \quad P_{20,4,5}(1) = \frac{\binom{4}{1}\binom{20-4}{5-1}}{\binom{20}{5}} \doteq 0{,}469\ldots,$$

$$P_{20,4,5}(2) = \frac{\binom{4}{2}\binom{20-4}{5-2}}{\binom{20}{5}} \doteq 0{,}216\ldots, \quad P_{20,4,5}(3) = \frac{\binom{4}{3}\binom{20-4}{5-3}}{\binom{20}{5}} \doteq 0{,}031\ldots,$$

$$P_{20,4,5}(4) = \frac{\binom{4}{4}\binom{20-4}{5-4}}{\binom{20}{5}} \doteq 0{,}001\ldots$$

(Pro kontrolu – součet vypočítaných pravděpodobností je 1.)

Vidíme, že nejvyšší pravděpodobnost – 0,4696 – je dosažena v případě, kdy mezi pěticí vybraných osob je právě 1 muž.

**Výpočet pomocí systému STATISTICA:**

Otevřeme nový datový soubor s pěti proměnnými P0, P1, P2, P3, P4 a jedním případem.

Do Dlouhého jména proměnné P0 napíšeme =Combin(4;0)*Combin(16;5)/Combin(20;5)

Do Dlouhého jména proměnné P1napíšeme =Combin(4;1*Combin(16;4/Combin(20;5)

Do Dlouhého jména proměnné P2napíšeme =Combin(4;2*Combin(16;3/Combin(20;5)

Do Dlouhého jména proměnné P3napíšeme =Combin(4;3*Combin(16;2/Combin(20;5)

Do Dlouhého jména proměnné P4napíšeme =Combin(4;4*Combin(16;1/Combin(20;5)

Dostaneme tabulku:

| | 1 P0 | 2 P1 | 3 P2 | 4 P3 | 5 P4 |
|---|---|---|---|---|---|
| 1 | 0,281 | 0,469 | 0,216 | 0,03( | 0,001 |

Vidíme, že nevyšší pravděpodobnost je P1, tedy nejpravděpodobnější počet mužů je jeden.

## Stochasticky nezávislé jevy

Za stochasticky nezávislé považujeme takové jevy, kdy informace o nastoupení jednoho jevu nijak neovlivní šance, s nimiž očekáváme nastoupení druhého jevu.

V popisné statistice jsme zavedli četnostní nezávislost dvou množin $G_1$, $G_2$ v daném výběrovém souboru pomocí multiplikativního vztahu $p(G_1 \cap G_2) = p(G_1)\,p(G_2)$.

V počtu pravděpodobnosti řekneme, že jevy $A_1, A_2 \in \mathcal{A}$ jsou stochasticky nezávislé, jestliže $P(A_1 \cap A_2) = P(A_1)\,P(A_2)$. Pro tři jevy budeme požadovat, aby i jevy $A_1 \cap A_2$ a $A_3$ byly stochasticky nezávislé, což vede ke vztahu $P(A_1 \cap A_2 \cap A_3) = P(A_1)\,P(A_2)\,P(A_3)$. Tak můžeme pokračovat pro libovolný počet jevů, tedy jevy $A_1, \ldots, A_n \in \mathcal{A}$ jsou stochasticky nezávislé, jestliže platí systém multiplikativních vztahů:

$\forall\, i < j$ $\quad P(A_i \cap A_j) = P(A_i)\,P(A_j)$ (dvojmístný multiplikativní vztah)

$\forall\, i < j < k$ $\quad P(A_i \cap A_j \cap A_k) = P(A_i)\,P(A_j)\,P(A_k)$ (trojmístný multiplikativní vztah)

$\vdots$

$P(A_1 \cap \ldots \cap A_n) = P(A_1) \ldots P(A_n)$ (n-místný multiplikativní vztah)

Jevy $A_1, A_2, \ldots \in \mathcal{A}$ jsou stochasticky nezávislé, jestliže pro všechna přirozená $n \geq 2$ jsou stochasticky nezávislé jevy $A_1, \ldots, A_n \in \mathcal{A}$.

(Lze snadno ukázat, že jev nemožný resp. jev jistý a libovolný jev jsou stochasticky nezávislé jevy. Jestliže v posloupnosti stochasticky nezávislých jevů nahradíme libovolný počet jevů jevy opačnými, stochastická nezávislost se neporuší. Rovněž tak průniky a sjednocení stochasticky nezávislých jevů jsou stochasticky nezávislé.)

**Příklad na stochasticky nezávislé jevy:** Nechť $A_1, A_2, A_3$ jsou stochasticky nezávislé jevy, $P(A_1) = 1/4$, $P(A_2) = 1/3$, $P(A_3) = 1/2$. Jaká je pravděpodobnost, že

a) nastane právě jeden z jevů $A_1, A_2, A_3$
b) nastanou právě dva z jevů $A_1, A_2, A_3$
c) nastanou nejvýše dva z jevů $A_1, A_2, A_3$ ?

**Řešení**:

ad a)

$$P\left((A_1 \cap \bar{A}_2 \cap \bar{A}_3) \cup (\bar{A}_1 \cap A_2 \cap \bar{A}_3) \cup (\bar{A}_1 \cap \bar{A}_2 \cap A_3)\right) =$$
$$= P(A_1)P(\bar{A}_2)P(\bar{A}_3) + P(\bar{A}_1)P(A_2)P(\bar{A}_3) + P(\bar{A}_1)P(\bar{A}_2)P(A_3) =$$
$$= \frac{1}{4}\left(1-\frac{1}{3}\right)\left(1-\frac{1}{2}\right) + \left(1-\frac{1}{4}\right)\frac{1}{3}\left(1-\frac{1}{2}\right) + \left(1-\frac{1}{4}\right)\left(1-\frac{1}{3}\right)\frac{1}{2} =$$
$$= \frac{1}{4}\cdot\frac{2}{3}\cdot\frac{1}{2} + \frac{3}{4}\cdot\frac{1}{3}\cdot\frac{1}{2} + \frac{3}{4}\cdot\frac{2}{3}\cdot\frac{1}{2} = \frac{2+3+6}{24} = \frac{11}{24}$$

ad b) $P(A_1)P(A_2)P(\bar{A}_3) + P(A_1)P(\bar{A}_2)P(A_3) + P(\bar{A}_1)P(A_2)P(A_3) = \frac{6}{24}$

ad c) $1-P(A_1)\,P(A_2)\,P(A_3) = \frac{23}{24}$

**Opakované nezávislé pokusy**: Nezávisle opakujeme týž náhodný pokus. Nechť jev $A_i$ znamená úspěch v i-tém pokusu, přičemž $P(A_i) = \vartheta$, i = 1, 2, ...

## 1. Binomické rozložení pravděpodobností

Pravděpodobnost, že v prvních n pokusech úspěch nastane právě x-krát ($0 < [illegible] < [illegible]$):

$$P_n(x) = [illegible].$$

K výpočtu v systému STATISTICA slouží funkce Binom(x; $\vartheta$; n)

Pravděpodobnost, že v prvních n pokusech úspěch nastane nejvýše $x_1$-krát ($0 < [illegible] < [illegible]$):

$$\sum_{x=0}^{x_1} P_n(x).$$

K výpočtu v systému STATISTICA slouží funkce IBinom($x_1$; $\vartheta$; n)

Pravděpodobnost, že v prvních n pokusech úspěch nastane aspoň $x_0$-krát ($0 < [illegible] < [illegible]$):

$$\sum_{x=x_0}^{n} P_n(x).$$

Výpočet lze provést takto: 1 - IBinom($x_0$ - 1; $\vartheta$; n)

Pravděpodobnost, že v prvních n pokusech úspěch nastane aspoň $x_0$-krát a nejvýše $x_1$-krát:

$$\sum_{x=x_0}^{x_1} P_n(x).$$

Výpočet lze provést takto: IBinom($x_1$; $\vartheta$; n) - IBinom($x_0$ - 1; $\vartheta$; n)

**Příklad na binomické rozložení pravděpodobností**: Firma se účastní čtyř nezávislých výběrových řízení. Pravděpodobnost, že uspěje v kterémkoliv z nich, je pro všechny konkurzy stejná a je rovna 0,7. Jaká je pravděpodobnost, že firma uspěje

a) právě 2x
b) aspoň 2x
c) nejvýše 2x?

**Řešení:** Počet pokusů n = 4, pravděpodobnost úspěchu $p$ = 0,7

ad a) $P_4(2) = \binom{4}{2} 0{,}7^2 0{,}3^2 = 0{,}264$

ad b) $P_4(2) + P_4(3) + P_4(4) = \binom{4}{2} 0{,}7^2 0{,}3^2 + \binom{4}{3} 0{,}7^3 0{,}3 + \binom{4}{4} 0{,}7^4 = 0{,}916$

ad c) $P_4(0) + P_4(1) + P_4(2) = \binom{4}{0} 0{,}3^4 + \binom{4}{1} 0{,}7 \cdot 0{,}3^3 + \binom{4}{2} 0{,}7^2 0{,}3^2 = 0{,}348$

**Výpočet pomocí systému STATISTICA:**

Otevřeme nový datový soubor se třemi proměnnými P1, P2, P3 a o jednom případu.
Do Dlouhého jména proměnné P1 napíšeme =Binom(2;0,7;4)
Do Dlouhého jména proměnné P2 napíšeme =1-IBinom(1;0,7;4)
Do Dlouhého jména proměnné P3 napíšeme =IBinom(2;0,7;4)
Dostaneme tabulku:

| | 1<br>P1 | 2<br>P2 | 3<br>P3 |
|---|---|---|---|
| 1 | 0,26 | 0,91 | 0,34 |

**2. Geometrické rozložení pravděpodobností**

Pravděpodobnost, že prvnímu úspěchu bude předcházet x neúspěchů:

$$P(x) = (1-\vartheta)^x \vartheta$$

K výpočtu v systému STATISTICA slouží funkce Geom(x; $\vartheta$)

Pravděpodobnost, že prvnímu úspěchu bude předcházet nejvýše $x_1$ neúspěchů:

$$\sum_{x=0}^{x_1} P(x)$$

K výpočtu v systému STATISTICA slouží funkce IGeom($x_1$; $\vartheta$)

Pravděpodobnost, že prvnímu úspěchu bude předcházet aspoň $x_0$ neúspěchů:

$$1 - \sum_{x=0}^{x_0-1} P(x)$$

Výpočet lze provést takto: 1 - IGeom($x_0$-1; $\vartheta$)

**Příklad na geometrické rozložení pravděpodobností:** Jaká je pravděpodobnost, že při hře „Člověče, nezlob se!“ nasadíme figurku nejpozději při třetím hodu?

**Řešení:**

Počet neúspěchů: x = 0, 1, 2, pravděpodobnost úspěchu: $\vartheta$

$\sum \ldots = \sum \ldots = \sum \left(\frac{5}{6}\right)^x \frac{1}{6} = \mathrm{Geom}(2;1/6) = 0{,}421$

Pravděpodobnost, že figurku nasadíme nejpozději při třetím hodu, je 42,13%.

**Příklad na geometrické rozložení pravděpodobností:** Studenti biologie zkoumají barvu očí octomilek. Pravděpodobnost, že octomilka má bílou barvu očí, je 0,25, červenou 0,75. Jaká je pravděpodobnost, že až čtvrtá zkoumaná octomilka má bílou barvu očí?

**Řešení:**

Počet neúspěchů: x = 3, pravděpodobnost úspěchu: $\vartheta$

$P(3) = 0{,}75^3 \cdot 0{,}25 = \mathrm{Geom}(3;0{,}25) = 0{,}105$

Pravděpodobnost, že až čtvrtá zkoumaná octomilka má bílou barvu očí, je 10,55%.

**3. Negativní binomické rozložení pravděpodobností**

Pravděpodobnost, že k-tému úspěchu bude předcházet x neúspěchů:

$$P_k(x) = \binom{x + k - 1}{\;} \ldots (1 - \;)\ldots .$$

Pravděpodobnost, že k-tému úspěchu bude předcházet nejvýše $x_1$ neúspěchů:

$$\sum_{x=0}^{x_1} P_k(x).$$

Pravděpodobnost, že k-tému úspěchu bude předcházet aspoň $x_0$ neúspěchů:

$$1 - \sum_{x=0}^{x_0-1} P_k(x).$$

**Příklad na negativní binomické rozložení pravděpodobností:** Hráč hází kostkou tak dlouho, dokud mu nepadnou tři šestky. Jaká je pravděpodobnost, že bude  muset hodit kostkou 10 x?

**Řešení:**

Počet neúspěchů x = 7, protože v 7 z 10 hodů nepadne šestka. Pravděpodobnost úspěchu $\vartheta$

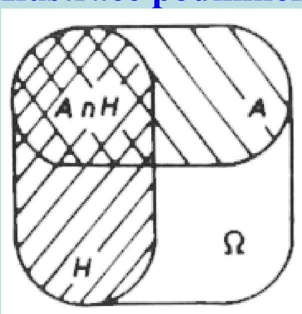
$$P_3(7) = \binom{3 + 7 - 1}{\;}\left(\;\right)^7\left(\;\right)^3 = \binom{9}{\;}\frac{5^7}{6^{10}} = 0{,}046$$

Hledaná pravděpodobnost je 4,65%.

## Podmíněná pravděpodobnost

Opakovaně nezávisle provádíme týž náhodný pokus a sledujeme nastoupení jevu A v těch pokusech, v nichž nastoupil jev H. Podmíněnou relativní četnost A za podmínky H jsme v popisné statistice zavedli vztahem $p(A/H) = \frac{p(A \cap H)}{p(H)}$ (za předpokladu, že p(H) > 0). Tato podmíněná relativní četnost se s rostoucím počtem pokusů ustaluje kolem konstanty $P(A/H) = \frac{P(A \cap H)}{P(H)}$, kterou považujeme za podmíněnou pravděpodobnost jevu A za podmínky H.

## Ilustrace podmíněné pravděpodobnosti




**Příklad na podmíněnou pravděpodobnost:** Jaká je pravděpodobnost, že při hodu kostkou padlo sudé číslo, je-li známo, že padlo číslo menší než 5?

**Řešení:** $\Omega = \{\omega_1, \ldots, \omega_6\}$, A … padlo sudé číslo, $A = \{\ldots\}$, H … padlo číslo menší než 5, $H = \{\ldots\}$, $A \cap H = \{\ldots\}$

$$P(A/H) = \frac{P(A \cap H)}{P(H)} = \frac{\frac{2}{6}}{\ldots} = \frac{1}{\ldots}$$

**Vlastnosti podmíněné pravděpodobnosti**

Je zřejmé, že jevy $A_1$, $A_2$ jsou stochasticky nezávislé, právě když
$P(A_1/A_2) = P(A_1)$ a právě když $P(A_2/A_1) = P(A_2)$.

Okamžitě z definice plyne:

$P(A_1 \cap A_2) = P(A_1)\,P(A_2/A_1)$ pro $P(A_1) > 0$,

$P(A_1 \cap A_2) = P(A_2)\,P(A_1/A_2)$ pro $P(A_2) > 0$.

Tento multiplikativní vztah lze zobecnit ve větu o násobení pravděpodobností:

$P(A_1 \cap A_2 \cap \ldots \cap A_n) = P(A_1)\,P(A_2/A_1)\,P(A_3/A_1 \cap A_2) \ldots P(A_n/A_1 \cap \ldots \cap A_{n-1})$ pro $P(A_1 \cap \ldots \cap A_{n-1}) \neq 0$.

**Příklad na větu o násobení pravděpodobností**: Ze sady 32 karet náhodně vytahujeme po jedné kartě, kterou nikdy nevracíme zpět. Jaká je pravděpodobnost, že eso se objeví až ve 4. tahu?

**Řešení**:

$A_i$ … v i-tém tahu nebylo vybráno eso, i = 1, 2, 3, 4

$$P(A_1 \cap A_2 \cap A_3 \cap \bar{A}_4) = P(A_1)\,P(A_2/A_1)\,P(A_3/A_1 \cap A_2)\,P(\bar{A}_4/A_1 \cap A_2 \cap A_3) =$$
$$= \frac{28}{32} \cdot \frac{27}{31} \cdot \frac{26}{30} \cdot \frac{4}{29} = 0{,}0911$$

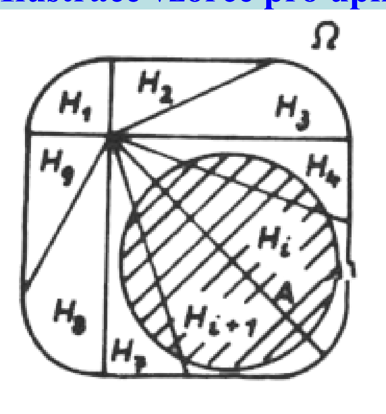
Eso se objeví až ve 4. tahu s pravděpodobností 0,0911.

## Vzorec pro úplnou pravděpodobnost a Bayesův vzorec

Jestliže $H_1, \ldots, H_n \in \mathcal{A}$ jsou jevy, které tvoří rozklad jistého jevu (tj. jsou neslučitelné a jejich sjednocením je celý základní prostor – říkáme, že tvoří úplný systém hypotéz), pak pravděpodobnost libovolného jevu $A \in \mathcal{A}$ lze vypočítat pomocí vzorce pro úplnou pravděpodobnost:

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

## Ilustrace vzorce pro úplnou pravděpodobnost




Podmíněnou pravděpodobnost libovolné hypotézy za podmínky, že nastal jev A - tzv. aposteriorní pravděpodobnost $P(H_k/A)$ - lze vypočítat pomocí Bayesova vzorce:

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

. (Původní pravděpodobnost $P(H_k)$ se nazývá apriorní pravděpodobnost).

Thomas Bayes (1702 – 1761): Anglický kněz a matematik

**Příklad na vzorec pro úplnou pravděpodobnost a Bayesův vzorec:** U jistého druhu elektrického spotřebiče se s pravděpodobností 0,01 vyskytuje výrobní vada. U spotřebiče s touto výrobní vadou dochází v záruční lhůtě k poruše s pravděpodobností 0,5. Výrobky, které tuto vadu nemají, se v záruční lhůtě porouchají s pravděpodobností 0,01. Jaká je pravděpodobnost, že

a) u náhodně vybraného výrobku nastane v záruční lhůtě porucha,

b) výrobek, který se v záruční lhůtě porouchá, bude mít dotyčnou výrobní vadu?

**Řešení:**

$H_1$ - výrobek má dotyčnou výrobní vadu

$H_2$ - výrobek nemá tuto výrobní vadu

A - výrobek se v záruční době porouchá

Pak je: $P(H_1) = 0{,}01$, $P(H_2) = 0{,}99$, $P(A/H_1) = 0{,}5$, $P(A/H_2) = 0{,}01$

ad a) $P(A) = P(H_1).P(A/H_1) + P(H_2).P(A/H_2) = 0{,}01.0{,}5 + 0{,}99.0{,}01 = 0{,}0149$

ad b) $P(H_1/A) = \frac{P(H_1).P(A/H_1)}{P(A)} = \frac{0{,}01.0{,}5}{0{,}0149} = 0{,}338$

**Využití Bayesova vzorce v medicíně při hodnocení kvality diagnostického testu**

Předpokládáme, že máme dvě skupiny objektů – jedna skupina objektů splňuje nějakou podmínku (pozitivní případy), druhá skupina nikoliv (negativní případy). Provedeme diagnostický test, který objekt označí buď jako pozitivní nebo jako negativní.

Zavedeme následující označení:

jev H … objekt je pozitivní
jev $\overline{H}$… objekt je negativní
jev A … test označí objekt za pozitivní
jev $\overline{A}$ … test označí objekt za negativní

Apriorní pravděpodobnost P(H) se nazývá prevalence a vyjadřuje pravděpodobnost výskytu pozitivních objektů v souboru všech objektů.
Podmíněná pravděpodobnost P(A/H)se nazývá senzitivita a vyjadřuje pravděpodobnost, že test dá kladný výsledek u pozitivního objektu.
Podmíněná pravděpodobnost P($\overline{A}$/$\overline{H}$)se nazývá specificita a vyjadřuje pravděpodobnost, že test dá záporný výsledek u negativního objektu.
Podmíněná pravděpodobnost P($\overline{A}$/H)se nazývá falešná negativita a vyjadřuje pravděpodobnost, že test dá záporný výsledek u pozitivního objektu.
Podmíněná pravděpodobnost PA/$\overline{H}$)se nazývá falešná pozitivita a vyjadřuje pravděpodobnost, že test dá kladný výsledek u negativního objektu.
Aposteriorní pravděpodobnost P(H/A) se nazývá prediktivní hodnota pozitivního testu a vyjadřuje pravděpodobnost, že objekt je skutečně pozitivní, když test dopadl pozitivně.
Aposteriorní pravděpodobnost P($\overline{H}$/$\overline{A}$) se nazývá prediktivní hodnota negativního testu a vyjadřuje pravděpodobnost, že objekt je skutečně negativní, když test dopadl negativně.

Uvedené podmíněné pravděpodobnosti neznáme, můžeme je pouze odhadnout pomocí výsledků testu, které zapíšeme do kontingenční tabulky:

| výsledek testu | podmínka | | celkem |
|---|---|---|---|
| | H (pozitivní) | $\bar{H}$ (negativní) | |
| A (pozitivní) | a | b | a+b |
| $\bar{A}$ (negativní) | c | d | c+d |
| celkem | a+c | b+d | n |

Odhad senzitivity: $P(A/H) = TPF = \frac{a}{a+c}$ (true positive fraction – relativní četnost správně klasifikovaných pozitivních případů).

Odhad specificity: $P(\bar{A}/\bar{H}) = TNF = \frac{d}{b+d}$ (true negative fraction – relativní četnost správně klasifikovaných negativních případů).

Odhad falešné negativity: $P(\bar{A}/H) = FNF = \frac{c}{a+c}$ (false negative fraction – relativní četnost nesprávně klasifikovaných pozitivních případů).

Odhad falešné pozitivity: $P(A/\bar{H}) = FPF = \frac{b}{b+d}$ (false positive fraction – relativní četnost nesprávně klasifikovaných negativních případů).

Je okamžitě zřejmé, že TPF + FNF = 1, TNF + FPF =1.

Odhady prediktivních hodnot pozitivního a negativního testu počítáme podle Bayesova vzorce.

$$P(H/A) = PVV = \frac{P(H)P(A/H)}{P(H)P(A/H) + P(\bar{H})P(A/\bar{H})} = \frac{P(H)\cdot TPF}{P(H)\cdot TPF + (1-P(H))(1-TNF)}$$

$$P(\bar{H}/\bar{A}) = PVN = \frac{P(\bar{H})P(\bar{A}/\bar{H})}{P(\bar{H})P(\bar{A}/\bar{H}) + P(H)P(\bar{A}/H)} = \frac{(1-P(H))\cdot TNF}{(1-P(H))\cdot TNF + P(H)(1-TPF)}$$

Vidíme, že praktický význam diagnostického testu záleží na prevalenci P(H), odhadu senzitivity TPF a odhadu specificity TNF. Tyto charakteristiky jsou tedy plně určeny prediktivními hodnotami PVV a PVN.

**Příklad:** Použijeme diagnostický test u předpovědi nemoci, která má prevalenci 1 % (tj. pravděpodobnost výskytu této nemoci v celé populaci je 0,01). Je známo, že tento test má 95 % senzitivitu (tj. s pravděpodobností 0,95 dá kladný výsledek u nemocného jedince ) a 95 % specificitu (tj. s pravděpodobností 0,95 dá záporný výsledek u zdravého jedince). Vypočtěte odhad prediktivní hodnoty pozitivního testu a negativního testu.

**Řešení:**

Odhad prediktivní hodnoty pozitivního testu:

$$P(H/A) = PV_+ = \frac{P(H)P(A/H)}{P(H)P(A/H) + P(\bar{H})P(A/\bar{H})} = \frac{P(H).TPF}{P(H).TPF + (1-P(H)).(1-TNF)} = \frac{0{,}01.0{,}95}{0{,}01.0{,}95 + 0{,}99.0{,}05} = 0{,}16$$

Znamená to, že jenom 16 % těch jedinců, u nichž vyšel test pozitivně, je skutečně nemocných danou chorobou, zatímco 84 % jedinců, kteří měli test pozitivní, chorobu nemá.

Odhad prediktivní hodnoty negativního testu:

$$P(\bar{H}/\bar{A}) = PV_- = \frac{P(\bar{H})P(\bar{A}/\bar{H})}{P(\bar{H})P(\bar{A}/\bar{H}) + P(H)P(\bar{A}/H)} = \frac{(1-P(H)).TNF}{(1-P(H)).TNF + P(H).(1-TPF)} = \frac{0{,}99.0{,}95}{0{,}99.0{,}95 + 0{,}01.0{,}05} = 0{,}999$$

Znamená to, že 99,9 % těch jedinců, u nichž vyšel test negativně, uvedenou chorobu skutečně nemá.