MV011 Statistika I

2. Podmíněná pravděpodobnost

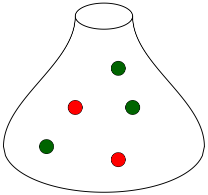
Jan Koláček (kolacek@math.muni.cz)

Ústav matematiky a statistiky, Přírodovědecká fakulta, Masarykova univerzita, Brno

Motivační příklad

Příklad 1

V pytlíku jsou 3 zelené a 2 červené kuličky. „Kamarád“ mi nabízí následující hru: „Vytáhneš si jednu kuličku, nebudeš ji vracet do pytlíku a pak vytáhneš ještě jednu. Za hru mi zaplatíš 35 Kč. Pokud však vytáhneš 2 zelené kuličky, zaplatím ti výhru 100 Kč.“ Mám si s ním zahrát?



$$P(1.\bullet \wedge 2.\bullet) = \underbrace{P(1.\bullet)}_{\frac{3}{5}} \cdot \underbrace{P(2.\bullet | 1.\bullet)}_{\frac{2}{4}} = 0,3$$

„Očekávaná“ výhra: $0,3 \cdot 100 = 30\,\text{Kč}$

Pokud by se kulička vracela: $P(1.\bullet \wedge 2.\bullet) = \underbrace{P(1.\bullet)}_{\frac{3}{5}} \cdot \underbrace{P(2.\bullet)}_{\frac{3}{5}} = 0,36$

Motivační příklady

Pokud bychom si předchozí přepsali pomocí jevů:
A . . . „1.●"
B . . . „2.●"
Nevracíme kuličku $\Rightarrow$ jevy A a B na sobě **závisí**
Pak

$$P(A \cap B) = P(A) \cdot P(B|A)$$

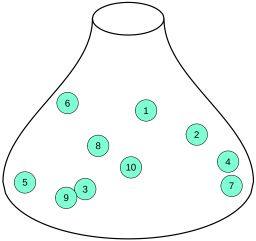
Vracíme kuličku $\Rightarrow$ jevy A a B jsou na sobě **nezávislé**, tj.

$$P(A \cap B) = P(A) \cdot P(B)$$

Motivační příklady

Příklad 2

V pytlíku je 10 kuliček označených čísly 1,...,10. „Kamarád" náhodně vytáhne jednu kuličku, prozradí nám jen, že její číslo je menší než 5 a řekne: „Vsaďte 1 Kč a když číslo bude sudé, dám vám 2 Kč." Vsadíme si?



Označme jevy:
A ... „číslo je sudé"
B ... „číslo je < 5"
Klasická PST: 4 možná čísla, z nich 2 sudá, tj.
$P(A|B) = \frac{2}{4} = 0,5$.

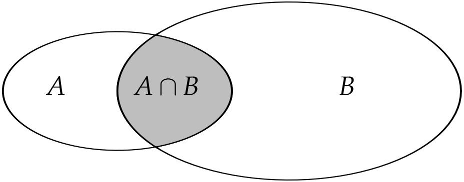
Otázka: Jaký je rozdíl mezi $P(A|B)$ a $P(A \cap B)$?

Motivační příklady

$A|B$ ... „číslo je sudé **za podmínky**, že je < 5“, tj. $P(A|B) = \frac{2}{4}$
$A \cap B$ ... „číslo je sudé **a současně** je < 5“, tj. $P(A \cap B) = \frac{2}{10}$

Otázka: Jak spolu souvisí $P(A|B)$ a $P(A \cap B)$?
A ... „číslo je sudé“, tj. $P(A) = \frac{5}{10}$
B ... „číslo je < 5“, tj. $P(B) = \frac{4}{10}$

$$\frac{2}{4} = P(A|B) = \frac{P(A \cap B)}{P(B)} = \frac{\frac{2}{10}}{\frac{4}{10}} = \frac{2}{4}$$



Podmíněná pravděpodobnost

Definice 1

Nechť $(\Omega, \mathcal{A}, P)$ je pravděpodobnostní prostor, $B \in \mathcal{A}$, $P(B) > 0$. Pak číslo

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

nazýváme **podmíněnou pravděpodobností jevu A za podmínky** (že nastal jev) B (**conditional probability**).

Z definice plyne

- $P(A \cap B) = P(A|B)P(B)$ platí i pro $P(B) = 0$, neboť $A \cap B \subseteq B$ a $P(B) = 0 \Rightarrow P(A \cap B) = 0$
- Symetricky platí také $P(A \cap B) = P(B|A)P(A)$

Podmíněná pravděpodobnost

Značení: Mějme pevně daný náhodný jev $B \in \mathcal{A}$, pro který platí $P(B) > 0$. Označme

$$P_B : \mathcal{A} \to \langle 0, 1 \rangle : P_B(A) = P(A|B)$$

Věta 2

P_B je pravděpodobnost na $(\Omega, \mathcal{A})$ pro každé $B \in \mathcal{A}$, pro které $P(B) > 0$.

Podmíněná pravděpodobnost

Příklad 3

Reklamní agent v nákupním centru má 25 balónků; 10 červených, 6 modrých a 9 zelených. Balónky rozdává náhodně zákazníkům. Jaká je pravděpodobnost, že prvním čtyřem lidem dá samé červené?

A_i … „i-tý zákazník dostane červený balónek“, $P(A_1 \cap A_2 \cap A_3 \cap A_4) =?$

$$P(A_1 \cap A_2 \cap A_3 \cap A_4) = \underbrace{P(A_1)}_{\frac{10}{25}} \underbrace{P(A_2|A_1)}_{\frac{9}{24}} \underbrace{P(A_3|A_1 \cap A_2)}_{\frac{8}{23}} \underbrace{P(A_4|A_1 \cap A_2 \cap A_3)}_{\frac{7}{22}}$$

$$= 0{,}0166$$

Podmíněná pravděpodobnost

Věta 3 (Věta o násobení pravděpodobností)

Platí

$$P\left(\bigcap_{i=1}^{n} A_i\right) = P(A_1) \cdot P(A_2|A_1) \cdot P(A_3|A_1 \cap A_2) \cdot \ \ldots \ \cdot P(A_n|A_1 \cap \cdots \cap A_{n-1})$$

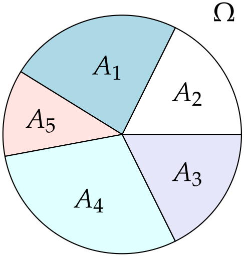
pro $P\left(\bigcap_{i=1}^{n-1} A_i\right) > 0.$

Úplný systém

Definice 4

Nechť $(\Omega, \mathcal{A}, P)$ je pravděpodobnostní prostor. Náhodné jevy $\{A_n\}_{n=1}^{\infty} \in \mathcal{A}$ tvoří **úplný systém jevů** na $(\Omega, \mathcal{A}, P)$, jestliže platí

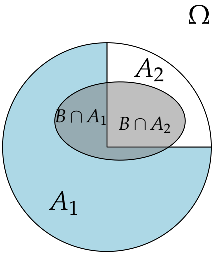
$$A_i \cap A_j = \emptyset, \text{ pro } i \neq j, \text{ a } \bigcup_{n=1}^{\infty} A_n = \Omega.$$



Příklad

Příklad 4

V krabici je 8 nových žárovek, o nichž víme, že 6 splňuje normu a 2 normu nesplňují. Víme, že žárovka, která splňuje normu, praskne v následujícím roce s pravděpodobností 0,2 a žárovka nesplňující normu praskne s pravděpodobností 0,5. Jaká je pravděpodobnost, že náhodně vybraná žárovka praskne?



$A_1 \ldots$ „ž. splňuje normu“, $P(A_1) = \frac{6}{8} = 0,75$

$A_2 \ldots$ „ž. je vadná“, $P(A_2) = \frac{2}{8} = 0,25$

$B \ldots$ „ž. praskne“, $\boxed{P(B) = ?}$

$B|A_1 \ldots$ „dobrá ž. praskne“, $P(B|A_1) = 0,2$

$B|A_2 \ldots$ „vadná ž. praskne“, $P(B|A_2) = 0,5$

$B = \{B \cap A_1\} \cup \{B \cap A_2\} \ldots$ disjunktní sjednoc.

$$\boxed{P(B)} = P(\{B \cap A_1\}) + P(\{B \cap A_2\})$$

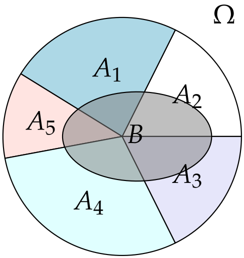
$$= \underbrace{P(B|A_1)}_{0,2} \underbrace{P(A_1)}_{0,75} + \underbrace{P(B|A_2)}_{0,5} \underbrace{P(A_2)}_{0,25} = \boxed{0,275}$$

Věta o úplné pravděpodobnosti

Věta 5 (Vzorec pro úplnou pravděpodobnost)

Nechť posloupnost $\{A_n\}_{n=1}^{\infty}$ *tvoří úplný systém jevů na* $(\Omega, \mathcal{A}, P)$ *takový, že* $P(A_i) > 0$ *pro* $i = 1, 2, \ldots$. *Pak platí*

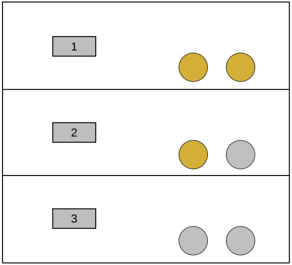
$$P(B) = \sum_{i=1}^{\infty} P(A_i) P(B|A_i).$$



Příklad

Příklad 5 (Poklad)

Trezor má 3 zásuvky se vzácnými mincemi. V první zásuvce jsou 2 zlaté mince, ve druhé je 1 zlatá a 1 stříbrná a ve třetí jsou 2 stříbrné mince. Náhodně zvolíme zásuvku a vytáhneme minci. Jaká je pravděpodobnost, že jsme vytáhli zlatou?



$Z_i \dots$ „otevřeme i-tou zásuvku", $i = 1, 2, 3$

$P(Z_1) = P(Z_2) = P(Z_3) = \frac{1}{3}$

$A \dots$ „vytáhneme zlatou minci", $\boxed{P(A) = ?}$

$A|Z_1 \dots$ „vytáhneme zlatou z 1. zásuvky"

$P(A|Z_1) = 1$

$A|Z_2 \dots$ „vytáhneme zlatou ze 2. zásuvky"

$P(A|Z_2) = \frac{1}{2}$

$A|Z_3 \dots$ „vytáhneme zlatou ze 3. zásuvky"

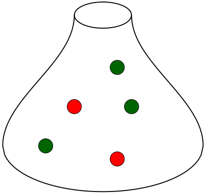
$P(A|Z_3) = 0$

$$\boxed{P(A)} = \underbrace{P(A|Z_1)}_{1}\underbrace{P(Z_1)}_{\frac{1}{3}} + \underbrace{P(A|Z_2)}_{\frac{1}{2}}\underbrace{P(Z_2)}_{\frac{1}{3}} + \underbrace{P(A|Z_3)}_{0}\underbrace{P(Z_3)}_{\frac{1}{3}} = \boxed{\frac{1}{2}}$$

Příklad

Ad **Příklad 1**

Podmínky hry jsou stejné. Kamarád „vylepší“ hru: „Místo tebe budu tahat já a prozradím ti barvu druhé kuličky. Za to mi zaplatíš navíc dalších 20 Kč. Pokud budou obě vytažené kuličky zelené, zaplatím ti výhru 100 Kč.“ Mám si s ním zahrát?



Chceme $P(1.{\color{green}\bullet}|2.{\color{green}\bullet})$, připomeňme $P(1.{\color{green}\bullet} \wedge 2.{\color{green}\bullet}) = 0,3$

Podle definice

$$P(1.{\color{green}\bullet}|2.{\color{green}\bullet}) = \frac{P(1.{\color{green}\bullet} \wedge 2.{\color{green}\bullet})}{P(2.{\color{green}\bullet})}$$

Podle Věty 5

$$P(2.{\color{green}\bullet}) = \underbrace{P(2.{\color{green}\bullet}|1.{\color{green}\bullet})}_{\frac{2}{4}}\underbrace{P(1.{\color{green}\bullet})}_{\frac{3}{5}} + \underbrace{P(2.{\color{green}\bullet}|1.{\color{red}\bullet})}_{\frac{3}{4}}\underbrace{P(1.{\color{red}\bullet})}_{\frac{2}{5}} = 0,6$$

A tedy $P(1.{\color{green}\bullet}|2.{\color{green}\bullet}) = \frac{0,3}{0,6} = \boxed{0,5} \Rightarrow$ „Očekávaná“ výhra: $0,5 \cdot 100 = 50$ Kč

Příklad

Shrnutí

$$P(1.{\color{green}\bullet}|2.{\color{green}\bullet}) = \frac{P(1.{\color{green}\bullet} \wedge 2.{\color{green}\bullet})}{P(2.{\color{green}\bullet})} = \frac{P(2.{\color{green}\bullet}|1.{\color{green}\bullet})P(1.{\color{green}\bullet})}{P(2.{\color{green}\bullet}|1.{\color{green}\bullet})P(1.{\color{green}\bullet}) + P(2.{\color{green}\bullet}|1.{\color{red}\bullet})P(1.{\color{red}\bullet})}$$

Označíme jevy
A_1 … „1.${\color{green}\bullet}$“
A_2 … „1.${\color{red}\bullet}$“
B … „2.${\color{green}\bullet}$“
Pak

$$P(A_1|B) = \frac{P(A_1 \cap B)}{P(B)} = \frac{P(B|A_1)P(A_1)}{P(B|A_1)P(A_1) + P(B|A_2)P(A_2)}$$

$$\Downarrow$$

Bayesův vzorec

Bayesův vzorec

Věta 6 (Bayesův vzorec)

Nechť posloupnost $\{A_n\}_{n=1}^{\infty}$ *tvoří úplný systém jevů na* $(\Omega, \mathcal{A}, P)$ *takový, že* $P(A_i) > 0$ *pro* $i = 1, 2, \ldots$ *a* $B \in \mathcal{A}$, *kde* $P(B) > 0$. *Pak*

$$P(A_j|B) = \frac{P(A_j)P(B|A_j)}{\sum\limits_{i=1}^{\infty} P(A_i)P(B|A_i)} \quad \text{pro} \quad j = 1, 2, \ldots.$$

Terminologie

- **Apriorní pravděpodobnost** … $P(A_i)$
- **Aposteriorní pravděpodobnost** … $P(A_i|B)$ (aktualizovaná pravděpodobnost)

Bayesův vzorec

- reverend **Thomas Bayes** (?1701 – 7.4.1761) presbytariánský kněz v Tunbridge Wells
- za života publikoval dvě práce: teologickou *Divine Benevolence (Laskavost Boží)* a anonymně obhajobu diferenciálního počtu
- nejdůležitější dílo: *An essay towards solving a problem in the doctrine of chances (Esej o řešení problému v doktríně o možnostech)* vydal po jeho smrti R. Price jakožto důkaz Boží existence

Příklad

Bayesovský filtr na spam

(Paul Graham, http://www.paulgraham.com/spam.html)

$Spam$ ... „dopis je spam“

$P(Spam)$ ... apriorní pravděpodobnost, globálně nastavená (0,5 – 0,8)

W ... „dopis obsahuje slovo W“

$P(W|Spam)$ ... pst, že spam obsahuje slovo W, stanovuje se z uložené pošty a spamu

$P(W|\overline{Spam})$ ... pst, že nespam obsahuje slovo W

$$P(Spam|W) = \frac{P(W|Spam)P(Spam)}{P(W|Spam)P(Spam) + P(W|\overline{Spam})P(\overline{Spam})}$$

Příklad

Ad **Příklad 4**: Náhodně vybraná žárovka praskla. Jaká je pravděpodobnost, že splňovala normu?

Připomeňme

A_1 ... „žárovka splňuje normu", $P(A_1) = \frac{6}{8} = 0,75$

A_2 ... „žárovka je vadná", $P(A_2) = \frac{2}{8} = 0,25$

$B|A_1$ ... „dobrá žárovka praskne", $P(B|A_1) = 0,2$

$B|A_2$ ... „vadná žárovka praskne", $P(B|A_2) = 0,5$

$A_1|B$ ... „žárovka, která praskne, splňuje normu", $P(A_1|B) = ?$

$$\begin{aligned} P(A_1|B) = \frac{P(A_1 \cap B)}{P(B)} &= \frac{P(B|A_1)P(A_1)}{P(B|A_1)P(A_1) + P(B|A_2)P(A_2)} \\ &= \frac{0,2 \cdot 0,75}{0,2 \cdot 0,75 + 0,5 \cdot 0,25} = 0,54 \end{aligned}$$

Příklad

Ad **Příklad 5**: Náhodně vybraná mince je zlatá. Jaká je pravděpodobnost, že byla vybrána z 1. zásuvky?

Připomeňme

Z_i ... „otevřeme i-tou zásuvku“, $i = 1, 2, 3$, $P(Z_1) = P(Z_2) = P(Z_3) = \frac{1}{3}$

A ... „vytáhneme zlatou minci“, $P(A) = 0,5$

$A|Z_1$ ... „vytáhneme zlatou z 1. zásuvky“ $P(A|Z_1) = 1$

$A|Z_2$ ... „vytáhneme zlatou ze 2. zásuvky“ $P(A|Z_2) = \frac{1}{2}$

$A|Z_3$ ... „vytáhneme zlatou ze 3. zásuvky“ $P(A|Z_3) = 0$

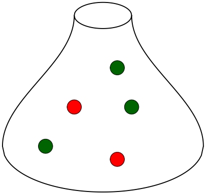
$Z_1|A$ ... „zlatá mince byla vytažena z 1. zásuvky“, $\boxed{P(Z_1|A) = ?}$

$$\boxed{P(Z_1|A)} = \frac{P(A|Z_1)P(Z_1)}{P(A|Z_1)P(Z_1) + P(A|Z_2)P(Z_2) + P(A|Z_3)P(Z_3)}$$

$$= \frac{1 \cdot \frac{1}{3}}{1 \cdot \frac{1}{3} + \frac{1}{2} \cdot \frac{1}{3} + 0 \cdot \frac{1}{3}} = \boxed{\frac{2}{3}}$$

Nezávislost jevů

Ad **Příklad 1**:



Pokud by se kulička **vracela**:

$$P(1.\bullet \wedge 2.\bullet) = \underbrace{P(1.\bullet)}_{\frac{3}{5}} \cdot \underbrace{P(2.\bullet)}_{\frac{3}{5}} = 0,36$$

Obecně

$$P(A \cap B) = P(A)P(B)$$

Intuice: jevy A, B na sobě nezávisí

$$P(A|B) = \frac{P(A \cap B)}{P(B)} = P(A) \text{ a } P(B|A) = \frac{P(A \cap B)}{P(A)} = P(B)$$

Nezávislost jevů

Definice 7

Nechť $(\Omega, \mathcal{A}, P)$ je pravděpodobnostní prostor. Pak řekneme, že jev $A \in \mathcal{A}$ a jev $B \in \mathcal{A}$ jsou **nezávislé** (**independent**) (vzhledem k pravděpodobnosti P), jestliže

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

Věta 8

- *Libovolný náhodný jev $A \in \mathcal{A}$ a jev* **jistý** *jsou* **nezávislé**.
- *Libovolný náhodný jev $A \in \mathcal{A}$ a jev* **nemožný** *jsou* **nezávislé**.
- *Nechť $A \in \mathcal{A}$ a $B \in \mathcal{A}$ jsou* **nezávislé** *jevy. Pak také A a $\bar{B}$, $\bar{A}$ a B, $\bar{A}$ a $\bar{B}$ jsou* **nezávislé**.

Skupinová nezávislost

Definice 9

Nechť $(\Omega, \mathcal{A}, P)$ je pravděpodobnostní prostor a $A_1, \dots, A_n \in \mathcal{A}$. Řekneme, že náhodné jevy $A_1, \dots, A_n$ jsou **skupinově** (**sdruženě**) **nezávislé**, jestliže pro libovolné $k \in \{1, \dots, n\}$ a libovolnou množinu indexů $\{i_1, \dots, i_k\} \subseteq \{1, \dots, n\}$ platí

$$P\left(\bigcap_{j=1}^{k} A_{i_j}\right) = \prod_{j=1}^{k} P(A_{i_j})$$

Nechť $M = \{A_i \in \mathcal{A}, i \in \mathcal{J}\}$, kde $\mathcal{J}$ je daná indexová množina (i nekonečná). Řekneme, že náhodné jevy systému M jsou **nezávislé**, jestliže pro každou konečnou množinu indexů $\{i_1, \dots, i_n\}$, kde $i_j \in \mathcal{J}, j = 1, \dots, n$ platí

$$P\left(\bigcap_{j=1}^{n} A_{i_j}\right) = \prod_{j=1}^{n} P(A_{i_j})$$

Příklad

Příklad 6

Dvakrát házíme kostkou.
Uvažujme následující jevy $A \ldots$ *v 1. hodu padne sudé číslo*
$B \ldots$ *v 2. hodu padne liché číslo*
$C \ldots$ *v obou hodech padne číslo stejné parity*

Protože platí

$$
\begin{aligned}
P(A) &= \tfrac{3\cdot 6}{36} = \tfrac{1}{2} & P(A\cap B) &= \tfrac{3\cdot 3}{36} = \tfrac{1}{4} = P(A)P(B) \\
P(B) &= \tfrac{6\cdot 3}{36} = \tfrac{1}{2} & P(A\cap C) &= \tfrac{3\cdot 3}{36} = \tfrac{1}{4} = P(A)P(C) \\
P(C) &= \tfrac{3\cdot 3+3\cdot 3}{36} = \tfrac{1}{2} & P(B\cap C) &= \tfrac{3\cdot 3}{36} = \tfrac{1}{4} = P(B)P(C)
\end{aligned}
$$

$$P(A\cap B\cap C) = \frac{0}{36} = 0 \neq P(A)P(B)P(C),$$

jsou jevy A, B a C **po dvou nezávislé**, ale **ne skupinově nezávislé**.

Poznámka

Frekventistická škola

vs.

Bayesovská škola

Příklad

Příklad 7

Pravděpodobnost rakoviny prsu u žen ve věku 40 – 50 let je 0,014. Podle studií naznačí mamograf u žen, které rakovinu nemají, nesprávně přítomnost nemoci pouze asi v 10 % případů. Pokud naopak rakovinu mají, odhalí ji asi v 75 % případů. Jaká je pravděpodobnost nemoci, jestliže výsledek vyšetření na mamografu byl pozitivní?

R ... „žena má rakovinu“, $P(R) = 0{,}014$, **apriorní** pst

MP ... „mamograf ukázal pozitivní výsledek“

$MP|R$ ... „pozitivní mamograf pro nemocnou ženu“, $P(MP|R) = 0{,}75$

$MP|\overline{R}$ ... „pozitivní mamograf pro zdravou ženu“, $P(MP|\overline{R}) = 0{,}1$

$P(R|MP) = ?$

Příklad

Příklad 8

Pravděpodobnost rakoviny prsu u žen ve věku 40 – 50 let je 0,014. Podle studií naznačí mamograf u žen, které rakovinu nemají, nesprávně přítomnost nemoci pouze asi v 10 % případů. Pokud naopak rakovinu mají, odhalí ji asi v 75 % případů. Jaká je pravděpodobnost nemoci, jestliže výsledek vyšetření na mamografu byl pozitivní?

R ... „žena má rakovinu", $P(R) = 0,014$, **apriorní** pst

MP ... „mamograf ukázal pozitivní výsledek"

$MP|R$ ... „pozitivní mamograf pro nemocnou ženu", $P(MP|R) = 0,75$

$MP|\overline{R}$ ... „pozitivní mamograf pro zdravou ženu", $P(MP|\overline{R}) = 0,1$

$\boxed{P(R|MP) = ?}$

$$\boxed{P(R|MP)} = \frac{P(MP|R)P(R)}{P(MP|R)P(R) + P(MP|\overline{R})P(\overline{R})} = \frac{0,75 \cdot 0,014}{0,75 \cdot 0,014 + 0,1 \cdot (1 - 0,014)} = \boxed{0,096}$$

Srovnání přístupů

Příklad 9

Máme chorobu, která bez léčby zabije 50 % nemocných, a zajímá nás, zda nově vyvinutý lék je účinný. Je podán pěti pacientům a všichni přežijí. Vyplývá z toho, že lék alespoň nějak funguje (na určité hladině pravděpodobnosti)?

Frekventistický přístup:

A . . . „všech pět pacientů přežije“

Lék není účinný $\Rightarrow$ pst přežití je pořád $\frac{1}{2} \Rightarrow P(A|N) = \left(\frac{1}{2}\right)^5 = \frac{1}{32} = \boxed{0{,}03125}$

Bayesovský přístup:

N . . . „lék je NEúčinný“, $P(N) = 0{,}5$, **apriorní** pst

$A|N$ . . . „všech pět pacientů přežije, přestože dostali neúčinný lék“, $P(A|N) = \frac{1}{32}$

$A|\overline{N}$ . . . „všech pět pacientů přežije, protože dostali účinný lék“, $P(A|\overline{N}) = 1$

$$P(N|A) = \frac{P(A|N)P(N)}{P(A|N)P(N) + P(A|\overline{N})P(\overline{N})} = \frac{\frac{1}{32}\cdot\frac{1}{2}}{\frac{1}{32}\cdot\frac{1}{2} + 1\cdot\frac{1}{2}} = \frac{1}{33} = \boxed{0{,}0303}$$