Jevy, nezávislost, Bayesova věta

17. března 2015

Instrukce:

Projděte si všechny příklady. Každý příklad se snažte pochopit. Pak vymyslete a vyřešte příklad podobný. Tím se ujistíte, že příkladu rozumíte. Další příklady najdete na stránkách Ivana Nagye.

Učivo:

Zápis jevů pomocí množin.
Pravděpodobnost průniku/sjednocení dvou obecných/neslučitelných/nezávislých jevů.
Závislost/nezávislost jevů.
Řešení příkladů pomocí pravděpodobnostního stromu.
Úplná pravděpodobnost.
Bayesův vzorec (bez úplné pravděpodobnosti).

Jevy zapsané množinou

Hrací kostka. Zapište množinově jev $J =$ „Padne sudé číslo." Jaká je pravděpodobnost tohoto jevu?

$J = \{2, 4, 6\}$
Základní prostor:
$\Omega = \{1, 2, 3, 4, 5, 6\}$
Všechny elementární jevy jsou stejně pravděpodobné.
$P = \frac{3}{6} = 0,5$

Dvě hrací kostky. Zapište množinově jev $J =$ „Padne součet menší než 5." Jaká je pravděpodobnost tohoto jevu?

$J = \{[1, 1], [1, 2], [1, 3], [2, 1], [2, 2], [3, 1]\}$
Základní prostor má 6x6 prvků.
Všechny elementární jevy jsou stejně pravděpodobné.
$P = \frac{6}{36} = \frac{1}{6}$

Podmíněná pravděpodobnost

Dvě hrací kostky. Jaká je pravděpodobnost, že padne součet menší než 5, pokud na první kostce padla 2?

> $J = \{[2,1], [2,2]\}$
> Základní prostor má 6x6 prvků. Nás ale zajímají jen ty, kdy na první kostce padne 2. Těch je šest. A z těchto šesti odpovídají našemu jevu dvě možnosti.
> Všechny elementární jevy jsou stejně pravděpodobné.
> $P = \frac{2}{6} = \frac{1}{3}$

Dvě hrací kostky. Jaká je pravděpodobnost, že padne součet větší než 8, pokud na první kostce padlo liché číslo?

> $J = \{[3,6], [5,4], [5,5], [5,6]\}$
> Základní prostor má 6x6 prvků. Nás ale zajímají jen ty začínající lichým číslem. těch je 18. A z těchto 18 odpovídají našemu jevu 4 možnosti.
> Všechny elementární jevy jsou stejně pravděpodobné.
> $P = \frac{4}{18} = \frac{2}{9}$

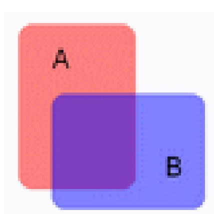
Průnik a sjednocení jevů

A a B jsou neslučitelné jevy. $P(A) = 0,2$. $P(B) = 0,3$. Jaká je pravděpodobnost jejich průniku a sjednocení?

> $P(A \cap B) = 0$
> $P(A \cup B) = P(A) + P(B) = 0,2 + 0,3 = 0,5$

A a B jsou jevy. $P(A) = 0,2$. $P(B) = 0,3$. Pokud nastal jev A, je pravděpodobnost jevu B rovna: $P(B|A) = 0.5$. Jaká je pravděpodobnost jejich průniku a sjednocení?

> $P(A \cap B) = P(B|A) . P(A) = 0,5 . 0,2 = 0,1$
> $P(A \cup B) = P(A) + P(B) - P(A \cap B) = 0,2 + 0,3 - 0,1 = 0,4$
> Vzorec pro průnik je jasný. Pravděpodobnost, že nastanou oba jevy, je pravděpodobnost, že nastane jeden jev krát pravděpodobnost, že nastane druhý jev, ovšem už za podmínky, že nastal ten první jev.
> Vzorec pro sjednocení je krásně vidět z následujícího obrázku. Při sečtení pravděpodobností bychom průnik započítali dvakrát, proto ho musíme jednou odečíst.



A a B jsou nezávislé jevy. $P(A) = 0,2$. $P(B) = 0,3$. Jaká je pravděpodobnost jejich průniku a sjednocení?

> $P\left(A \cap B\right) = P\left(A\right).P\left(B\right) = 0,2.0,3 = 0,06$
> $P\left(A \cup B\right) = P\left(A\right) + P\left(B\right) - P\left(A \cap B\right) = P\left(A\right) + P\left(B\right) - P\left(A\right).P\left(B\right) = 0,2 + 0,3 - 0,2.0,3 = 0,44$

Závislost a nezávislost jevů

Hrací kostka.
Jev A: Padne jedno z čísel 1,2,5,6.
Jev B: Padne jedno z čísel 3,4,6.
Jsou tyto jevy závislé nebo nezávislé?

> Pro nezávislé jevy platí dva speciální vztahy:
> $P\left(A|B\right) = P\left(A\right)$ ... nezáleží na podmínce
> $P\left(A \cap B\right) = P\left(A\right).P\left(B\right)$ ... plyne z předchozího vztahu po dosazení do $P\left(A \cap B\right) = P\left(A|B\right).P\left(B\right)$.
> Testujme např. podle druhého vztahu:
> $A \cap B = \{6\}$
> $P\left(A\right) = \frac{4}{6}$
> $P\left(B\right) = \frac{3}{6}$
> $P\left(A \cap B\right) = \frac{1}{6}$
> Dosadíme do vztahu a dostaneme:
> $\frac{1}{6} \neq \frac{4}{6}.\frac{3}{6}$
> Tedy jevy NEjsou NEzávislé, tedy jsou závislé.

Hrací kostka.
Jev A: Padne jedno z čísel 1,2,3,6.
Jev B: Padne jedno z čísel 3,4,6.
Jsou tyto jevy závislé nebo nezávislé?

> Pro nezávislé jevy platí dva speciální vztahy:
> $P\left(A|B\right) = P\left(A\right)$ ... nezáleží na podmínce
> $P\left(A \cap B\right) = P\left(A\right).P\left(B\right)$ ... plyne z předchozího vztahu po dosazení do $P\left(A \cap B\right) = P\left(A|B\right).P\left(B\right)$.
> Testujme např. podle druhého vztahu:
> $A \cap B = \{3, 6\}$
> $P\left(A\right) = \frac{4}{6}$
> $P\left(B\right) = \frac{3}{6}$
> $P\left(A \cap B\right) = \frac{2}{6}$
> Dosadíme do vztahu a dostaneme:
> $\frac{2}{6} = \frac{4}{6}.\frac{3}{6}$
> Tedy jevy jsou NEzávislé.

Chováme myšky. Bílé a černé. S modrýma nebo hnědýma očima. Jejich počty ukazuje následující tabulka:

	modré oči	hnědé oči
bílé	5	15
černé	10	30

Jev A: Myška je černá.
Jev B: Myška má modré oči.
Jsou tyto jevy závislé nebo nezávislé?

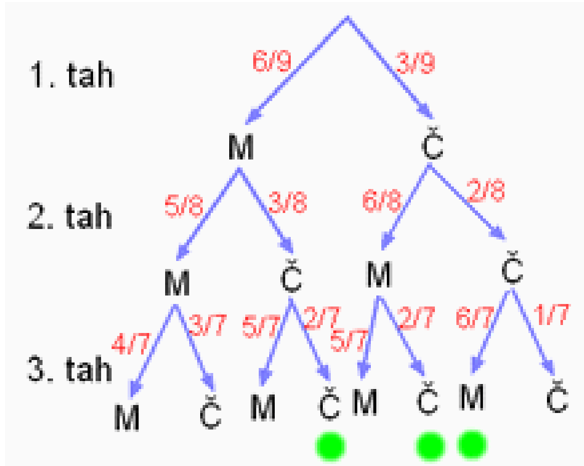
> Pro nezávislé jevy platí dva speciální vztahy:
> $P(A|B) = P(A)$ ... nezáleží na podmínce
> $P(A \cap B) = P(A) . P(B)$ ... plyne z předchozího vztahu po dosazení do $P(A \cap B) = P(A|B) . P(B)$.
> Testujme např. podle druhého vztahu:
> $P(A) = \frac{40}{60}$
> $P(B) = \frac{15}{60}$
> $P(A \cap B) = \frac{10}{60}$
> Dosadíme do vztahu a dostaneme:
> $\frac{10}{60} = \frac{40}{60} \cdot \frac{15}{60}$
> Tedy jevy jsou NEzávislé.

Poznámka: V rámci náhody se samozřejmě může stát, že i když jsou jevy nezávislé, rovnost $P(A \cap B) = P(A) . P(B)$ nebude splněna přesně. Zvládnout tuto situaci se budeme učit ve Statistice.

Strom

V pytlíku je šest kuliček modrých a tři červené. Jaká je pravděpodobnost, že vytáhnu jednu modrou a dvě červené?

> Kuličky po jednotlivých tazích nevracíme. Tím se tedy průběžně mění pravděpodobnost, že vytáhnu takovou či onakou kuličku. Musíme si tedy sestrojit strom všech možností a pak vybrat ty možnosti, které vyhovují zadání.

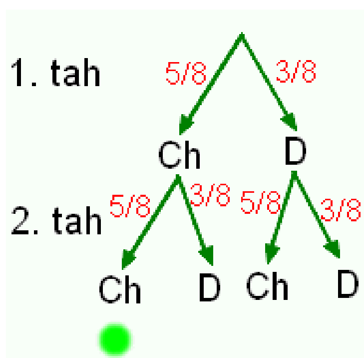


Zadání vyhovují tři zeleně označené možnosti.
$\frac{6}{9} \cdot \frac{3}{8} \cdot \frac{2}{7} + \frac{3}{9} \cdot \frac{6}{8} \cdot \frac{2}{7} + \frac{3}{9} \cdot \frac{2}{8} \cdot \frac{6}{7} = \frac{3}{14}$

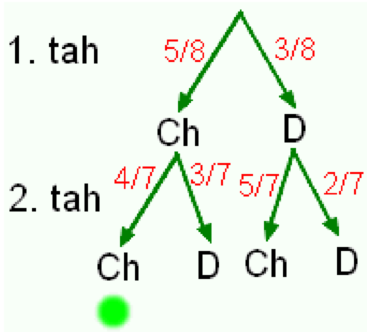
V klobouku je 5 jmen chlapeckých a 3 jména dívčí. Losuji dvě. Jaká je pravděpodobnost, že obě jména budou chlapecká, pokud

1. po vylosování lístek vrátím do klobouku.
2. po vylosování lístek nevrátím do klobouku.

1) Vrátím do klobouku



$\frac{5}{8} \cdot \frac{5}{8} = \frac{25}{64}$
2) Nevrátím do klobouku



$\frac{5}{8} \cdot \frac{4}{7} = \frac{5}{14}$

Bayesova věta

Odvození Bayesovy věty:

$$\begin{aligned} P(A \cap B) &= P(A|B).P(B) \\ &= P(B|A).P(A) \end{aligned}$$

Tedy: $P(A|B).P(B) = P(B|A).P(A)$
Tedy: $P(A|B) = \frac{P(B|A).P(A)}{P(B)}$

90% lidí s rakovinou plic jsou kuřáci.
28% lidí kouří.
8% lidí má rakovinu plic.
Jaká je pravděpodobnost, že onemocním rakovinou plic, pokud budu kouřit?

K ... Kouří.
R ... Rakovina plic.
$P(K) = 0,28$
$P(R) = 0,08$
Na těchto příkladech je nejtěžší určit, co je podmínka.
První větu můžu přepsat: Když vím, že někdo má rakovinu plic, tak je to na 90% kuřák. Nebo taky: Za podmínky, že má rakovinu plic, tak je to na 90% kuřák. Tedy: $P(K|R) = 0,9$.
Bayesova věta:
$P(R|K) = \frac{P(K|R).P(R)}{P(K)} = \frac{0,9.0,08}{0,28} = 0,26 = 26\%$
Když budu kouřit, na 26% dostanu rakovinu plic.

Když sedneme za volant, máme šanci asi 0,01%, že během této jízdy způsobíme vážnou dopravní nehodu. Asi půl procenta řidičů na silnicích řídí pod vlivem alkoholu. Asi 14% vážných dopravních nehod způsobili řidiči pod vlivem alkoholu. Když se napiji alkoholu a pak jdu řídit, kolikrát vzroste pravděpodobnost, že během této jízdy způsobím vážnou dopravní nehodu?

A ... Alkohol.
N ... Nehoda.
$P(A) = 0,005$
$P(N) = 0,0001$
... Z těch řidičů, kteří způsobili nehodu, jich bylo 14% pod vlivem. = Za podmínky, že způsobili nehodu, je pravděpodobnost, že byli pod vlivem 14%. Tedy: $P(A|N) = 0,14$.
Bayesova věta:
$P(N|A) = \frac{P(A|N).P(N)}{P(A)} = \frac{0{,}14.0{,}0001}{0{,}005} = 0,0028$
Pravděpodobnost nehody je u opilého řidiče $\frac{P(N|A)}{P(N)} = 28$ krát větší, než když o pití řidiče nevíme nic.

Pokud chceme spočíst, kolikrát vzroste pravděpodobnost proti řidiči, který nepil, musíme spočíst zlomek : $\frac{P(N|A)}{P(N|A')}$.
Použijeme rozpis pro úplnou pravděpodobnost: $P(N) = P(N \wedge A) + P(N \wedge A') = P(N|A).P(A) + P(N|A').P(A')$.
Vyjádříme to, co nás zajímá: $P(N|A') = \frac{P(N)-P(N|A).P(A)}{P(A')}$.
Ještě potřebujeme vyjádřit pravděpodobnost, že náhodný řidič nepil alkohol: $P(A') = 1 - P(A) = 0,995$
Spočteme pravděpodobnost, že řidič, který nepil, způsobí vážnou dopravní nehodu: $P(N|A') = \frac{P(N)-P(N|A).P(A)}{P(A')} = \frac{0{,}0001-0{,}0028.0{,}005}{0{,}995} = 0,000086$
Dosadíme: $\frac{P(N|A)}{P(N|A')} = \frac{0{,}0028}{0{,}000086} = 32,4$.
Pravděpodobnost vážné nehody u řidiče, který pil alkohol je 32-krát větší než u řidiče, který alkohol nepil.

Bayesův vzorec s úplnou pravděpodobností

V anglickém městečku Longsweek ukradli z muzea vzácnou sochu Černého prince. Přihlásil se svědek, který uvedl, že u muzea viděl dodávku značky Bedford CA modré barvy. To by mohlo vyšetřujícímu detektivovi velmi pomoci, protože zjistil, že funkčních tmavých Bedfordů CA je v Británii už pouze 432 a z nich je modrých pouze patnáct. Ostatní jsou černé. Chtěl si ověřit, že se svědek nespletl v barvě a proto druhý den nechal na stejné místo za stejných okolností předjíždět několik tmavých aut. Svědek se spletl v jednom pokusu z osmi. Jaká je za těchto okolností pravděpodobnost, že viděné auto bylo skutečně modré?

Jev M ... Bedford CA je modrý.
Jev V ... Svědek vyhodnotí barvu jako modrou.
$P(M) = \frac{15}{432} = 0,0347$
Pokud je auto modré, pak jej svědek vyhodnotí jako modré s pravděpodobností:
$P(V|M) = \frac{7}{8} = 0,875$.
Pokud auto není modré, pak jej svědek vyhodnotí jako modré s pravděpodobností:
$P(V|M') = \frac{1}{8} = 0,125$.
Jaká je pravděpodobnost $P(M|V)$?

Vyjdeme z Bayesova vzorce: $P(M|V) = \frac{P(V|M).P(M)}{P(V)}$.

Neznáme však $P(V)$. Spočteme je úplnou pravděpodobností:
$P(V) = P(V \cap M) + P(V \cap M') = P(V|M).P(M) + P(V|M').P(M')$
Známe vše až na $P(M')$, které je ale triviálně: $P(M') = 1 - P(M) = 0,9653$.
Dosadíme a spočteme: $P(V) = 0,875 \cdot 0,0347 + 0,125 \cdot 0,9653 = 0,151$.
Dosadíme do Bayesova vzorce: $P(M|V) = \frac{0,875 \cdot 0,0347}{0,151} = 0,201$.
Pravděpodobnost, že Bedford CA byl opravdu modrý je pouhých dvacet procent.

Těhotná maminka může absolvovat invazivní genetický test miminka. Při něm se zjišťuje kromě jiného Downův syndrom. Spolehlivost testu je cca 99,5%. Downův syndrom má cca jedno z 900 počatých dětí. Pokud byl test u miminka pozitivní, jaká je pravděpodobnost, že má skutečně Downův syndrom?

Jev D ... Miminko má Downův syndrom.
Jev T ... Test je pozitivní.
$P(D) = \frac{1}{900} = 0,0011$
Pokud je miminko nemocné, pak test bude pozitivní na 99,5%: $P(T|D) = 0,995$.
Pokud je miminko zdravé, bude test pozitivní na 0,5%: $P(T|D') = 0,005$.
Jaká je pravděpodobnost $P(D|T)$?
Vyjdeme z Bayesova vzorce: $P(D|T) = \frac{P(T|D).P(D)}{P(T)}$.
Neznáme však $P(T)$. Spočteme je úplnou pravděpodobností:
$P(T) = P(T \cap D) + P(T \cap D') = P(T|D).P(D) + P(T|D').P(D')$
Známe vše až na $P(D')$, které je ale triviálně: $P(D') = 1 - P(D) = 0,9989$.
Dosadíme a spočteme: $P(T) = 0,995 \cdot 0,0011 + 0,005 \cdot 0,9989 = 0,006089$.
Dosadíme do Bayesova vzorce: $P(D|T) = \frac{0,995 \cdot 0,0011}{0,006089} = 0,18$.
Pravděpodobnost, že miminko má skutečně Downův syndrom je cca 18%.

Poznámka 1: Ve skutečnosti je situace s testováním Downova syndromu o něco složitější.
Poznámka 2: Invazivní vyšetření (odběr vzorku z placenty nebo plodové vody) se někdy nepovede. Asi v jednom procentu případů vyšetření způsobí smrt miminka.
