6.1 Príklady na Bayesov vzorec a vzájomnú nezávislosť

Nasleduje úloha, ktorú vyriešime najprv pomocou zdravého 'sedliackeho rozumu' a následne aj pomocou Bayesovho vzorca.

Príklad 6.1 *V červenej a modrej miske boli cukríky. Cukríky boli zelené a biele. V červenej miske boli 2 zelené a 8 bielych cukríkov, v modrej miske boli 4 zelené a 1 biely cukrík. Cukríky sme vysypali na tanier a ponúkli hosťom. Následne sa zistilo, že cukríky v modrej miske olízal pes. Aká je pravdepodobnosť, že hosť si vybral cukrík olízaný psom, ak vieme, že hosť zjedol biely cukrík?*

Riešenie: Úlohu vyriešime ľahko, bez pomoci špeciálnych vzorcov. Bielych cukríkov bolo 9, z toho jeden bol v modrej miske. Teda

$$P(M|B) = \frac{P(M \cap B)}{P(B)} = \frac{1}{9}$$

Nie vždy sú však známe tie pravdepodobnosti, ktoré potrebujeme k takémuto jednoduchému výpočtu. Vyriešime príklad ešte raz, aby sme videli, že na výpočet môžeme použiť aj iné vstupné informácie.

Označme udalosti postupne C (cukrík z červenej misky), M (cukrík z modrej misky), B (biely cukrík)a Z (zelený cukrík). Jednotlivé pravdepodobnosti potom budú

$$P(C) = \frac{2}{3}, \quad P(M) = \frac{1}{3}, \quad P(B|C) = \frac{8}{10}, \quad P(B|M) = \frac{1}{5}$$

Použitím Bayesovho vzorca dostávame:

$$P(M|B) = \frac{P(B|M) \cdot P(M)}{P(B|M) \cdot P(M) + P(B|C) \cdot P(C)} = \frac{\frac{1}{5} \cdot \frac{1}{3}}{\frac{1}{5} \cdot \frac{1}{3} + \frac{8}{10} \cdot \frac{2}{3}} = \frac{1}{9} = 0.11$$

□

Príklad 6.2 *Vo výrobe je* 0.3% *zlých čipov. Test vyradí čip s* $p = 0.98$ *ak bol zlý, a vyradí s* $p = 0.001$ *ak bol dobrý.*

- *Aká je pravdepodobnosť, že testom vyradený čip je naozaj zlý?*
- *Vyrobíme 10000 čipov. Koľko bude (v priemere) vyradených?*

Riešenie:

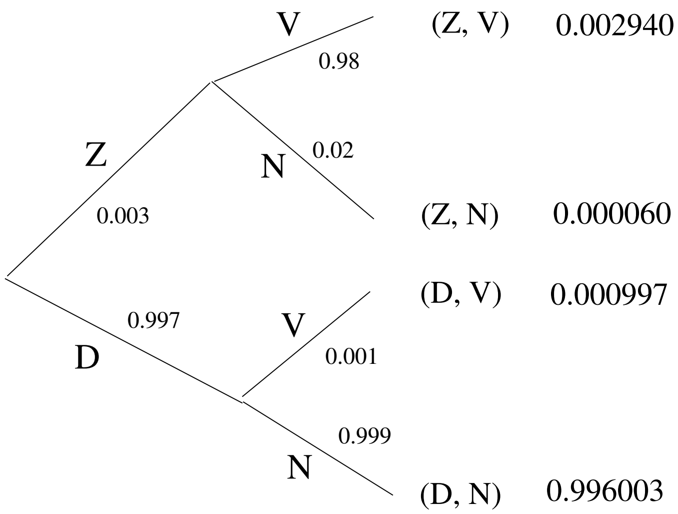
$$Z/D - \text{ zlý|dobrý čip }, \quad \Pr(Z) = 0.003, \quad \Pr(D) = 0.997$$
$$V|Z - \text{ test vyradí|nevyradí čip }, \quad \Pr(V|Z) = 0.98, \quad \Pr(V|D) = 0.001$$

Vyrobíme 10000 čipov. Koľko bude (v priemere) vyradených?

$$10000 \Pr(V) = 10000(\Pr(ZV) + \Pr(DV)) = 10000 \cdot 0.003937 = 39.37$$

Aká je pravdepodobnosť, že vyradený čip je naozaj zlý?

$$\Pr(Z|V) = \frac{\Pr(Z \cap V)}{\Pr(V)} = \frac{0.002940}{0.002940 + 0.000997} = 0.747$$



Medzi vyradenými čipmi je $39.37 \cdot 0.253 = 9.96$ dobrých. Výsledok:

$$\Pr(A) = \frac{1}{2} \cdot \frac{1}{6} + \frac{1}{2} \cdot \frac{1}{36} = \frac{7}{72}.$$

$\square$

Príklad 6.3 *Uvažujme nad takýmto experimentom. Najskôr hodíme férovou mincou. Ak padne znak, tak hodíme jeden krát férovou kockou. Ak padne hlava, tak hodíme dvakrát férovou kockou a sčítame hodnoty týchto dvoch hodov. Aká je pravdepodobnosť, že dostaneme výsledok* 2*?*

Nasledujúci príklad je z kurzu 6.042J/18.062J, Fall '05 na MIT (Massachusetts Institute of Technology) Mathematics for Computer Science od Prof. Albert R. Meyer and Prof. Ronitt Rubinfeld

Príklad 6.4 (Príklad číslo 13.7.2) *Lekári objavili novú chorobu nazývanú Bifľomor, ktorá nakazila* $1/1000$ *populácie. Jeden zo symptómov je, že sa začnete poctivo učiť včetky predmety v škole. Je to strašná choroba. V poslednom štádiu sa prejavuje tak, že nakazený jedinci zýskajú červený diplom. Dvaja celosvetovo uznávaný doktori tvrdia, že vedia Bifľomor diagnostikovať. Ak sa necháte otestovať doktorom* X*, či máte túto chorobu, tak*

- *ak máte túto chorobu, tak vám s pravdepodobnosťou* 0.99 *povie áno*
- *ak nemáte túto chorobu, tak vám s pravdepodobnosťou* 0.97 *povie nie.*

Ak sa necháte otestovať doktorom Y*, tak keÄŽže vie, že choroba nakazila* $1/1000$ *populácie, tak vám s pravdepodobnosťou* 0.001 *povie áno a s pravdepodobnosťou* 0.999 *povie nie, teda vôbec vás netestuje, iba náhodne generuje svoje odpovede.*

a) Aká je pravdepodobnosť, že sa doktor X *pomýli?*

b) Aká je pravdepodobnosť, že sa doktor Y *pomýli?*

c) Ktorý doktor je má menšiu pravdepodobnosť, že sa pomýli?

d) Aký test by bol ešte spoľahlivejší a nepotrebovali by ste naň vôbec študovať medicínu?

Vzájomná nezávislosť

Udalosti $E_1, E_2, ...E_n$ sú **vzájomné nezávislé** práve vtedy, ak pre každú k-prvkovú podmnožinu platí

$$\Pr\left[\bigcap_{\forall k} E_k\right] = \prod_{\forall k} \Pr(E_k)$$

$$\begin{aligned}
\Pr(E_i \cap E_j) &= \Pr(E_i)\Pr(E_j) \\
\Pr(E_i \cap E_j \cap E_k) &= \Pr(E_i)\Pr(E_j)\Pr(E_k) \\
&\vdots \\
\Pr(E_1 \cdot ... \cdot E_n) &= \Pr(E_1) \cdot ... \cdot \Pr(E_n)
\end{aligned}$$

Príklad 6.5 *Hody troma mincami:*

Udalosť A_1 spočíva v tom, že to čo padne na prvej minci je rovnaké ako to, čo padne na druhej minci.

Udalosť A_2 spočíva v tom, že to čo padne na druhej minci je rovnaké ako to, čo padne na tretej minci.

Udalosť A_3 spočíva v tom, že to čo padne na tretej minci je rovnaké ako to, čo padne na prvej minci.

Ukážte, že tieto udalosti sú po dvoch nezávislé, ale nie sú vzájomne nezávislé.

Riešenie: Označme výsledky hodu mincou H a T. Všetky možné výsledky sú:

$$\Omega = \{HHH,\ HHT,\ HTH,\ HTT,\ THH,\ THT,\ TTH,\ TTT\}$$

$$\Pr(A_1) = \frac{4}{8} = \frac{1}{2} = \Pr(A_2) = \Pr(A_3)$$

$$\Pr(A_1 \cap A_2) = \frac{2}{8} = \frac{1}{4} = \Pr(A_2 \cap A_3) = \Pr(A_3 \cap A_1)$$

$$\Pr(A_1 \cap A_2 \cap A_3) = \frac{2}{8} \neq \frac{1}{8} = \Pr(A_1) \cdot \Pr(A_2) \cdot \Pr(A_3)$$

Príklad 6.6 *Zločin v Seatli. Simpson vs. štát Washington, máj 1995. Počet obyvateľov Seatli je a cca 745000. O páchateľovi je známe, že:*

- má čierne BNW, (vyhovuje 20800 ľudí)
- je plešatý (vyhovuje 30% populácie)
- je psychológ (má visačku na saku), (vyhovuje 1 na 100 ľudí)
- má medzi 40-45 rokov, (vyhovuje 7% populácie)
- má cca 100 kg, (vyhovuje 23% populácie)

Riešenie: Poradca žaloby neabsolvoval kurz pravdepodobnosti, preto predpokladal, že všetky znaky sú nezávislé. Vtedy je pravdepodobnosť výskytu páchateľa v Seatli

$$\Pr(V) = \left[\frac{20800}{745000} \cdot 0.3 \cdot 0.01 \cdot 0.07 \cdot 0.23\right] \cdot 745000 = 1.3485 \cdot 10^{-6}$$

Počet obyvateľov Seatlu, ktorí majú všetky vlastnosti a teda by mohli byť páchateľom je

$$\left[\frac{20800}{745000} \cdot 0.3 \cdot 0.01 \cdot 0.07 \cdot 0.23\right] \cdot 745000 = 1.3485 \cdot 10^{-6} \cdot 745000 = 1.0046$$

Obžaloba tvrdí, že je iba jeden taký človek. Ten, má všetky vlastnosti a je to páchateľ. Simpson všetky vlastnosti mal, ale obhajoba vzniesla námietku.

Námietka. Nie všetky znaky sú nezávislé. Pripustíme iba nezávislosť vlastnenia BNW a plešiny:

$$\left[\frac{20800}{745000} \cdot 0.3\right] \cdot 745000 = 0.0084 \cdot 745000 = 6258$$

Obhajoba ukázala, že počet možných páchateľov sa pohybuje od 1 do 6258.

7 Náhodná premenná a pravdepodobnostná funkcia, nezávislosť náhodných premenných, distribučná funkcia, doplnková distribučná funkcia.

7.1 Náhodná premenná a pravdepodobnostná funkcia (PDF)

Matematika má veľa užitočných nástrojov pre prácu s číslami. Preto je rozumné urobiť prevod medzi náhodnými udalosťami a číselnými množinami. Tento prevod (zobrazenie, mapovanie) sa nazýva: **náhodná premenná**. Presnejšie, náhodná premenná je zobrazenie elementárnych náhodných udalostí na reálne čísla. Elementárnym udalostiam (výsledkom pokusu) sú priradené čísla, ktoré ich reprezentujú. **Pravdepodobnostná funkcia, PDF náhodnej premennej** opisuje, ako je rozdelená pravdepodobnosť medzi jednotlivé čísla. Platí pri tom, že hodnota pravdepodobnostnej funkcie pre číslo x, teda $PDF_X(x)$ je rovnaká ako pravdepodobnosť udalosti, zodpovedajúcej tomuto číslu x.

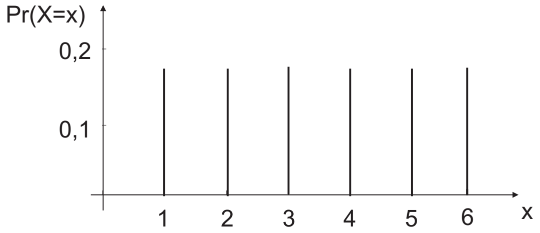
Definícia 7.1 *Náhodná premenná X je zobrazenie z množiny výsledkov pokusu Ω do množiny reálnych čísel, $X : \Omega \to R$. Náhodná premenná priraďuje výsledkom čísla, resp. vytvára matematický model nejakého reálneho procesu. Je jednoznačne popísaná svojou pravdepodobnostnou funkciou (probability distribution function). Pravdepodobnostnú funkciu náhodnej premennej X budeme označovať $PDF_X(.)$.*

$$PDF_X(x_k) = \Pr(X = x_k) = p_k$$

Vlastnosti pravdepodobnostnej funkcie

1. $\forall x_i;\ PDF_X(x_i) = \Pr(X = x_i) \geq 0$
2. $\sum\limits_{\forall x_i} PDF_X(x_i) = \sum\limits_{\forall x_i} \Pr(X = x_i) = 1$

Úloha 7.1 *Nech náhodná premenná $\mathbb{X}$ popisuje hod kockou. Udalosti, že na kocke padla jednotka, teda stena so znakom ⚀ priradí číslo 1; udalosti, že na kocke padlo ⚁ priradí číslo 2, …, udalosti, že na kocke padlo ⚅ priradí číslo 6. Aké rozdelenie pravdepodobnosti má náhodná premenná, ktorá takýmto spôsobom popisuje hod kockou?*



Obr. 1: Rozdelenie pravdepodobnosti náhodnej premennej popisujúcej hod kockou

Riešenie: Každá z uvedených udalostí má rovnakú pravdepodobnosť rovnú 1/6, preto aj každé z čísel $x = 1, 2, \ldots, 6$ sa bude vyskytovať s touto pravdepodobnosťou. Rozdelenie pravdepodobnosti pre jednotlivé hodnoty náhodnej premennej vidíme na obrázku 1.

Predpis pre náhodnú premennú, prevod medzi udalosťami a číslami, je

$$\mathbb{X} : \{\text{⚀}, \text{⚁}, \text{⚂}, \text{⚃}, \text{⚄}, \text{⚅}\} \to \{1, 2, 3, 4, 5, 6\}$$

$$\mathbb{X}\,(\text{na kocke padla hodnota} \quad k) \mapsto k$$

$$P(\mathbb{X} = k) = \frac{1}{6}$$

$$P(\mathbb{X} = \text{⚀}) = \frac{1}{6}, \quad P(\mathbb{X} = \text{⚁}) = \frac{1}{6}, \quad P(\mathbb{X} = \text{⚂}) = \frac{1}{6}, \ldots, \quad P(\mathbb{X} = \text{⚅}) = \frac{1}{6}$$

$$P(\mathbb{X} =) = \frac{1}{6}, \quad P(\mathbb{X} = 2) = \frac{1}{6}, \quad P(\mathbb{X} = 3) = \frac{1}{6}, \ldots, \quad P(\mathbb{X} = 6) = \frac{1}{6}$$

□

Táto úloha bola príliš jednoduchá na to, aby sme pojem v úplnosti pochopili. Ďalšie už budú zložitejšie.

Úloha 7.2 *Náhodná premenná* $\mathbb{X}$ *popisuje hru, spočívajúcu v hode kockou a vyplatení príslušnej výhry. Ak na kocke padne číslo* 1 *alebo číslo* 2*, hráč dostane* 20 *Eur. Ak padne niečo iné, hráč zaplatí* 2 *Eurá. Aké rozdelenie pravdepodobnosti má náhodná premenná, ktorá popisuje túto hazardnú hru?*

Riešenie: Udalosti sú tentokrát dve, hráč dostane 20 Eur, hráč zaplatí 2 Eurá. Náhodná premenná $\mathbb{X}$ teda nadobudne dve hodnoty.

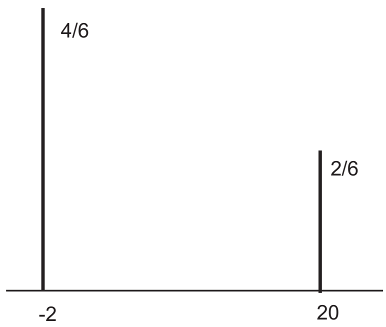
$$\mathbb{X}\,(\text{hráč dostane 20 Eur}) = 20$$

$$\mathbb{X}\,(\text{hráč zaplatí 2 Eurá}) = -2$$

Uvedené udalosti majú pravdepodobnosti rovné 2/6 a 4/6.
Preto sa čísla -2 a 20 budú vyskytovať s týmito pravdepodobnosťami. Rozdelenie pravdepodobnosti vidíme na obrázku 2.

$$P(\mathbb{X} = -2) = \frac{4}{6}, \quad P(\mathbb{X} = 20) = \frac{2}{6}$$

□



Obr. 2: Rozdelenie pravdepodobnosti náhodnej premennej popisujúcej hazardnú hru s kockou

7.2 Nezávislosť náhodných premenných

Nezávislosť dvoch náhodných premenných

Náhodné premenné R_1 a R_2 sú nezávislé, ak pre všetky výsledky r_1 a r_2 platí

$$\Pr(R_1 = r_1 \cap R_2 = r_2) = \Pr(R_1 = r_1) \cdot \Pr(R_2 = r_2)$$

Príklad 7.1 *Nech náhodná premenná S označuje súčet pri hode dvoma kockami a náhodná premenná R označuje ich rozdiel (vždy odčítame menšie od väčšieho, aby bol rozdiel kladný). Sú náhodné premenné S a R nezávislé?*

Riešenie: Ak chceme dokázať, že sú nezávislé, urobíme všetky možné dvojice hodnôt týchto náhodných prememnných a overíme, či platí rovnosť

$$\Pr(R_1 = r_1 \cap R_2 = r_2) = \Pr(R_1 = r_1) \cdot \Pr(R_2 = r_2)$$

Myslíme si, že náhodné premenné S a R sú závislé. Vtedy stačí nájsť aspoň jednu situáciu, pre ktoré rovnosť neplatí. Takou situáciou je

$$\Pr((S = 2) \cap (R = 1)) \neq \Pr(S = 2) \cdot \Pr(R = 1)$$

pretože

$$0 \neq \frac{1}{36}\frac{10}{36} = 0.0077$$

Vzájomná nezávislosť:

Náhodné premenné $R_1, R_2, ..., R_n$ sú nezávislé, ak pre všetky výsledky $r_1, r_2, ..., r_n$ platí

$$\Pr((R_1 = r_1) \cap ... \cap (R_n = r_n)) = \Pr(R_1 = r_1) \cdot ... \cdot \Pr(R_n = r_n)$$

7.3 Distribučná funkcia náhodnej premennej (CDF)

Distribučná funkcia (*cumulative distribution function*, $CDF_X(.)$) náhodnej premennej X je zobrazenie z reálnych čísel do intervalu $\langle 0, 1\rangle$ definované vzťahom:

$$CDF_X(x) = \Pr(X \leq x) = \sum_{x_k \leq x} \Pr(X = x_k) = \sum_{x_k \leq x} PDF_X(x_k)$$

7.4 Doplnková distribučná funkcia náhodnej premennej (TDF)

Doplnková distribučná funkcia (*tail distribution function* $TDF_X(.)$, rozdelenie pravdepodobnosti chvosta) náhodnej premennej X je zobrazenie z reálnych čísel do intervalu $\langle 0, 1\rangle$ definované vzťahom:

$$TDF_X(x) = 1 - CDF_X(x) = \bar{F}_X(x) = \Pr(X > x) = \sum_{x_k > x} \Pr(X = x_k) = \sum_{x_k > x} PDF_X(x_k)$$