Poznámky k předmětu Aplikovaná statistika, 1. téma

Motivace

Na otázku, při jaké teplotě vře voda, nejspíš neodpovíte. Budete chtít znát podmínky, které máte uvažovat. Víme, že za normálního tlaku, tj. $1013,25\,\mathrm{hPa}$, je vždy teplota varu vody $100\,^\circ\mathrm{C}$. Výsledek takového pokusu je stále stejný.

Nyní si představte, že všichni vaši kolegové změří teplotu varu vody v místě svého bydliště ve stejný čas. Tyto hodnoty se budou lišit v závislosti na tlaku v dané lokalitě. Pokud bude sběr dat anonymní, můžete výsledky těchto měření považovat za náhodné.

1. Náhodné jevy

- můžeme na ně nahlížet jako na výsledky pokusů (konaných za stejných podmínek). Dopředu nemůžeme určit, který z více možných výsledků nastane.
- **Deterministický jev** (nenáhodný): za stejných podmínek vychází stále stejný výsledek.
- **Elementární náhodný jev** ω je takový náhodný jev, který už se nedá rozdělit na podjevy. $\Omega\ldots$ množina všech elementárních jevů neboli množina všech možných výsledků pokusu. $\Omega = \{\omega;\ \omega \text{ je elementární jev}\}$.
  Náhodné jevy jsou pak podmnožiny množiny Ω. Značíme je obvykle velkými písmeny ze začátku abecedy.

  Příklad 1. *Házíme spravedlivou šestistěnnou kostkou (pozn.: „spravedlivá“ znamená, že všechny možnosti, co mohou padnout, jsou stejně pravděpodobné). Potom*

  $$\Omega = \{\omega_1, \omega_2, \ldots, \omega_6\},$$

  kde ω_i je elementární jev, že na kostce padlo číslo i. Označme dále A jev, že na kostce padlo liché číslo a B jev, že číslo, které padlo je vyšší než 3. Jevy A a B již nejsou elementární jevy, ale dají se pomocí nich vyjádřit:

  $$\begin{aligned} A &= \{\omega_1, \omega_3, \omega_5\} \\ B &= \{\omega_4, \omega_5, \omega_6\}. \end{aligned}$$

- **Symbolika náhodných jevů**

  $\Omega \ldots$ jev jistý

  $\emptyset \ldots$ jev nemožný

  $A \subset B$ nastane-li jev A, nastane i jev B

  $C = A \cup B \ldots$ jev C nastane, nastane-li alespoň jeden z jevů A, B

  $C = A \cap B \ldots$ jev C nastane, nastanou-li oba jevy A, B zároveň

  $C = A \setminus B \ldots$ jev C nastane, nastane-li jev A, ale nenastane jev B

  $A^c = \Omega \setminus A \ldots$ jev doplňkový k jevu A (jiné možné značení $\bar{A}$, A')

  pokud $A \cap B = \emptyset$, nazýváme jevy A, B **disjunktní (neslučitelné)**

2. Některé definice pravděpodobnosti

(i) **Klasická definice pravděpodobnosti**
Nechť Ω má jen konečně mnoho prvků ($\Omega = \{\omega_1, \dots, \omega_n\}$) a všechny elementární jevy jsou stejně pravděpodobné ($\mathbb{P}(\omega_i) = \frac{1}{n}, \forall i$), potom pravděpodobnost náhodného jevu $A = \bigcup_{j=1}^{m} \{\omega_j\}$ je
$$\mathbb{P}(A) = \frac{m}{n},$$
neboli „počet příznivých možností ku počtu všech možností".

Poznámka 1. *Důležitým předpokladem, na který se zapomíná, je právě stejná pravděpodobnosti elementárních jevů. Je třeba si správně definovat množinu Ω. Například házíme-li dvěma kostkami, pak jevy „padnou dvě jedničky" a „padne jedna jednička a jedna dvojka" nejsou stejně pravděpodobné, ačkoliv se může zdát, že ano. Musíme si uvědomit, že dvě jedničky mohou padnou jen jedním způsobem, kdežto jednička a dvojka způsoby dvěma. Pokud se nám zdá, že tyto dvě možnosti jsou nerozlišitelné, stačí si představit, že házíme dvěma různobarevnými kostkami nebo jednou kostkou dvakrát za sebou. Pak je zřejmé, že ke stejnému výsledku jedna jednička a jedna dvojka můžeme dospět dvěma různými způsoby. Množina Ω je potom množina uspořádaných dvojic čísel od jedné do šesti, z nichž první udává počet ok na první kostce (v prvním hodu) a druhé počet ok na druhé kostce (v druhém hodu). Množina Ω má 36 prvků a pravděpodobnost jednotlivých dvojic je $\frac{1}{36}$. Tedy i pravděpodobnost dvou jedniček je $\frac{1}{36}$, ale pravděpodobnost jedničky a dvojky je $2 \cdot \frac{1}{36} = \frac{1}{18}$.*

(ii) **Statistická definice pravděpodobnosti**
Provedeme n pokusů, z nichž v n_A případech nastal jev A. Potom odhad pravděpodobnosti jevu A je přibližně
$$\frac{n_A}{n} \approx \mathbb{P}(A) \qquad \text{a} \qquad \mathbb{P}(A) = \lim_{n\to\infty} \frac{n_A}{n}.$$

(iii) **Geometrická pravděpodobnost**
Je jakousi spojitou analogií klasické pravděpodobnosti. Možností již není konečně mnoho, ale spadají do nějaké množiny $\Omega \subset \mathbb{R}^d$. Jevy nejsou ani stejně pravděpodobné, nicméně pravděpodobnost, že výsledek pokusu padne do nějaké množiny, je úměrná velikosti (délce, ploše, objemu) té množiny vzhledem k velikosti množiny Ω. Geometrická pravděpodobnost, že náhodný jev A padne do množiny S je potom
$$\mathbb{P}(A) = \frac{\mu(S)}{\mu(\Omega)},$$
kde μ označuje velikost množiny (délka, obsah, objem,...). Jak je vidno, na pravděpodobnost můžeme nahlížet různými způsoby.

Cvičení 1. Do stejného přístavu mají během následujících dvanácti hodin přijet dvě lodi. Jejich posádky jsou dohodnuté, že na sebe navzájem v přístavu hodinu počkají. Pokud během té doby nepřijede druhá loď, loď přístav opouští. Obě lodi přijíždí do přístavu náhodně. Jaká je pravděpodobnost, že se posádky setkají?

Cvičení 2. Zasažení terče tvaru kruhu o poloměru R je jistý jev. Určete pravděpodobnost, že střela zasáhne vnitřek rovnostranného trojúhelníku vepsaného hraniční kružnici.

- **Vlastnosti pravděpodobnosti**
  - $\mathbb{P}(\Omega) = 1$, $\mathbb{P}(\emptyset) = 0$
  - $0 \leq \mathbb{P}(A) \leq 1, \forall A$
  - Jsou-li jevy $A_1, A_2, \ldots, A_n$ po dvou disjunktní, potom $\mathbb{P}\left(\bigcup_{i=1}^n A_i\right) = \sum_{i=1}^n \mathbb{P}(A_i)$
  - $\mathbb{P}(A \cup B) = \mathbb{P}(A) + \mathbb{P}(B) - \mathbb{P}(A \cap B)$
    spec. A, B disjunktní $\Rightarrow \mathbb{P}(A \cup B) = \mathbb{P}(A) + \mathbb{P}(B)$
  - $A \subset B \Rightarrow \mathbb{P}(A) \leq \mathbb{P}(B)$
  - $\mathbb{P}(A \setminus B) = \mathbb{P}(A) - \mathbb{P}(A \cap B)$
    spec. $B \subset A \Rightarrow \mathbb{P}(A \setminus B) = \mathbb{P}(A) - \mathbb{P}(B)$
  - $\mathbb{P}(A^c) = 1 - \mathbb{P}(A)$

Příklad 2. *V osudí je deset losů, dva z nich jsou výherní a osm nevýherních. Vytáhneme náhodně dvojici losů. Jaká je pravděpodobnost, že z nich alespoň jeden bude výherní?*

Řešení. *Počet všech možností, jak vytáhnout dvojici losů, je roven kombinačnímu číslu* $\binom{10}{2} = 45$. *Počet možností, jak vytáhnout jeden výherní a jeden nevýherní los je 16, neboť ke každé možnosti výherního (celkem dvě) můžeme vytáhnout libovolný nevýherní (celkem 8). Možnost, jak vytáhnout dva výherní je jen jedna. Počet příznivých jevů je tedy 17. Potom pravděpodobnost vytáhnutí alespoň jednoho výherního losu je tedy rovna*

$$\frac{17}{45}.$$

Cvičení 3. V šuplíku je 6 černých, 4 bílé a 2 modré ponožky. Ráno (ještě potmě) vytáhneme 3 ponožky najednou. Určete pravděpodobnost jevu, že všechny ponožky jsou stejné barvy a pravděpodobnost jevu, že aspoň 2 jsou stejné barvy.

Cvičení 4. Házíme třikrát hrací kostkou. Označme jevem A, že součet čísel na kostkách je sudý a jevem B, že v prvním hodu padne jednička. Spočítejte $\mathbb{P}(A)$, $\mathbb{P}(B)$, $\mathbb{P}(A \cap B)$ a $\mathbb{P}(A \cup B)$.

Cvičení 5. (vznikl na základě ankety zadané na cvičení vašich kolegů) Nechť pravděpodobnost jevu K, že student VŠCHT pije kofolu, je $\mathbb{P}(K) = 3/5$, jevu P, že student VŠCHT pije pivo, je $\mathbb{P}(P) = 13/20$. Dále víme, že $\mathbb{P}(K \cap P) = 7/20$. Stanovte pravděpodobnosti jevů, že student VŠCHT

(a) pije aspoň jeden nápoj,

(b) pije pouze kofolu,

(c) nepije ani pivo, ani kofolu,

(d) nepije pivo,

(e) pije buď jen kofolu, anebo pouze pivo,

(f) pije pivo, ale nepije kofolu,

(g) pije nejvýše jeden z uvedených nápojů.

3. Podmíněná pravděpodobnost

Uvažujme jevy A a B, kde $\mathbb{P}(B) > 0$. *Podmíněnou pravděpodobností* jevu A za podmínky, že nastal jev B (značíme $\mathbb{P}(A|B)$), nazýváme podíl

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

Podmíněná pravděpodobnost je v podstatě opět pravděpodobností jen „žije ve světě", kde už nastal jev B. Má tedy analogické vlastnosti:

- $\mathbb{P}(\Omega|B) = 1$,
- $\mathbb{P}(A|B) \geq 0, \ \forall A$,
- $\mathbb{P}(A|B) = 1 - \mathbb{P}(A^c|B)$. Zde pozor, NEPLATÍ, že $\mathbb{P}(A|B) = 1 - \mathbb{P}(A|B^c)$!!!

Poznámka 2. *Pokud navíc je $\mathbb{P}(A) > 0$, pak ze symetrie průniku plyne i*

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

a vyjádříme-li z obou rovností průnik, dostáváme vztah

$$\mathbb{P}(A \cap B) = \mathbb{P}(A|B)\,\mathbb{P}(B) = \mathbb{P}(B|A)\,\mathbb{P}(A).$$

Nezávislost jevů

Jevy A, B jsou **nezávislé**, jestliže platí:

$$\mathbb{P}(A \cap B) = \mathbb{P}(A) \cdot \mathbb{P}(B).$$

Skupina jevů $A_1, \ldots, A_n$ je nezávislá, jestliže pro každou $\{n_1, \ldots n_k\} \subseteq \{1, \ldots, n\}$ platí

$$\mathbb{P}\left(\bigcap_{i=1}^{k} A_{n_i}\right) = \prod_{i=1}^{k} \mathbb{P}(A_{n_i}) = \mathbb{P}(A_{n_1}) \cdot \ldots \cdot \mathbb{P}(A_{n_k}).$$

Poznámka 3. *Nepleťme si pojmy nezávislosti a disjunktnosti jevů.*

Poznámka 4. *Ve světle definice podmíněné pravděpodobnosti by pro jevy A, B s kladnou pravděpodobností šlo nezávislost jevů definovat následovně:*
Nechť $\mathbb{P}(A) > 0, P(B) > 0$. Potom jsou jevy A, B nezávislé jestliže

$$\mathbb{P}(A|B) = \mathbb{P}(A) \qquad \textit{nebo} \qquad \mathbb{P}(B|A) = \mathbb{P}(B)$$

Cvičení 6. Házíme dvěma kostkami. Označme jevem A, že padne šestka a jevem B, že součet na obou kostkách je 8. Určete $\mathbb{P}(A|B)$ a zjistěte, zdali jsou jevy A, B nezávislé.

Cvičení 7. Házíme dvěma kostkami – červenou a zelenou. Označme jevy

$$\begin{aligned} A &= \{\text{na červené kostce padne sudé číslo}\}, \\ B &= \{\text{na zelené kostce padne liché číslo}\}, \\ C &= \{\text{součet čísel na obou kostkách je lichý}\}. \end{aligned}$$

Jsou jevy A, B, C nezávislé? Jsou po dvou nezávislé?

Věta 1. ***(O úplné pravděpodobnosti)*** *Nechť* $\Omega = \bigcup_{i=1}^{n} E_i$*, kde* E_i *jsou jevy navzájem disjunktní a* $\mathbb{P}(E_i) > 0, \forall i$*. Potom pravděpodobnost náhodného jevu* A *je rovna*

$$\mathbb{P}(A) = \sum_{i=1}^{n} \mathbb{P}(A|E_i)\,\mathbb{P}(E_i).$$

Cvičení 8. V osudí je osm bílých a dvě červené koule. V prvním tahu vytáhneme jednu kouli. Kouli vrátíme zpět a ještě do osudí přidáme jednu kouli stejné barvy jako byla vytažená koule. Táhneme znovu z osudí. Určete pravděpodobnost, že

(a) ve druhém tahu vytáhneme červenou kouli,
(b) ve druhém tahu vytáhneme bílou kouli,
(c) vytáhneme dvakrát červenou kouli,
(d) první tažená koule byla červená, jestliže jsme v druhém tahu vytáhli červenou kouli.

Bayesova věta

Z definice podmíněné pravděpodobnosti dostáváme nejjednodušší verzi Bayesovy věty.

Věta 2. *Nechť* $\mathbb{P}(A) > 0$ *a* $\mathbb{P}(B) > 0$*. Potom*

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

O něco obecnější verzi získáme, rozepíšeme-li pravděpodobnost jevu B pomocí věty o úplné pravděpodobnosti:

Věta 3. *Nechť* $\Omega = \bigcup_{i=1}^{n} E_i$*, kde* E_i *jsou jevy navzájem disjunktní,* $\mathbb{P}(E_i) > 0, \forall i$ *a* $\mathbb{P}(B) > 0$*. Potom*

$$\mathbb{P}(E_j|B) = \frac{\mathbb{P}(B|E_j)\mathbb{P}(E_j)}{\sum_{i=1}^{n} \mathbb{P}(B|E_i)\,\mathbb{P}(E_i)}.$$

Příklad 3. *Je známo, že nemocí* N *trpí* $2\,\%$ *populace. Nemoc je diagnostikována na základě vyšetření, jež odhalí u nemocného nemoc s pravděpodobností* $0{,}9$ *a u zdravého ji vyloučí v* $70\,\%$ *případů. Byla vyšetřena náhodně vybraná osoba. Určíme pravděpodobnosti, že diagnóza byla správná,*

(a) byl-li výsledek vyšetření pozitivní,
(b) byl-li výsledek vyšetření negativní.

Řešení. *Označme* Z *jev, že pacient netrpí nemocí* N*. Jev, že nemocí trpí označme* Z^c*. Dále* D *značí jev, že nemoc byla diagnostikována, jev, že nebyla diagnostikována, označme* D^c*. Potom ze zadání známe následující pravděpodobnosti:*

$\mathbb{P}(Z^c) = 0{,}02\ldots$ *pravděpodobnost, že náhodně vybraný člověk trpí nemocí* N*,*

$\mathbb{P}(D|Z^c) = 0{,}9\ldots$ *pravděpodobnost odhalení nemoci u nemocného,*

$\mathbb{P}(D^c|Z) = 0{,}7\ldots$ *pravděpodobnost vyloučení nemoci* N *u zdravého.*

V (a) odpovídáme na otázku, jaká je pravděpodobnost, že pacient nemocí skutečně trpí, byl-li výsledek vyšetření pozitivní, tedy počítáme pravděpodobnost jevu $Z^c|D$*. Dle Bayesovy věty je*

$$\mathbb{P}(Z^c|D) = \frac{\mathbb{P}(D|Z^c)\mathbb{P}(Z^c)}{\mathbb{P}(D)} = \frac{\mathbb{P}(D|Z^c)\mathbb{P}(Z^c)}{\mathbb{P}(D|Z^c)\mathbb{P}(Z^c) + \mathbb{P}(D|Z)\mathbb{P}(Z)}.$$

Z vlastností pravděpodobnosti a podmíněné pravděpodobnosti je $\mathbb{P}(Z) = 1 - \mathbb{P}(Z^c) = 0{,}98$ *a* $\mathbb{P}(D|Z) = 1 - \mathbb{P}(D^c|Z) = 0{,}3$. *Potom*

$$\mathbb{P}(Z^c|D) = \frac{0{,}9 \cdot 0{,}02}{0{,}9 \cdot 0{,}02 + 0{,}3 \cdot 0{,}98}.$$

V (b) odpovídáme na otázku, jaká je pravděpodobnost, že pacient nemocí netrpí, byl-li výsledek vyšetření negativní, tedy počítáme pravděpodobnost jevu $Z|D^c$. *Ta je rovna*

$$\mathbb{P}(Z|D^c) = \frac{\mathbb{P}(D^c|Z)\mathbb{P}(Z)}{\mathbb{P}(D^c)} = \frac{\mathbb{P}(D^c|Z)\mathbb{P}(Z)}{\mathbb{P}(D^c|Z)\mathbb{P}(Z) + \mathbb{P}(D^c|Z^c)\mathbb{P}(Z^c)}.$$

Chybějící pravděpodobnosti opět snadno dopočteme: $\mathbb{P}(D^c|Z^c) = 1 - \mathbb{P}(D|Z^c) = 0{,}1$. *Pak*

$$\mathbb{P}(Z|D^c) = \frac{0{,}7 \cdot 0{,}98}{0{,}7 \cdot 0{,}98 + 0{,}1 \cdot 0{,}02}.$$

Do kalkulačky si čtenář snadno dosadí. Všimněte si, že jevy E_i *z Bayesovy věty jsou zde pouze dva,* $E_1 = Z, E_2 = Z^c$.

Cvičení 9. Ve studijním kruhu je 30 studentů, z toho 20 chlapců. Pravděpodobnost jevu, že chlapec má krátké vlasy, je $0,7$. Pravděpodobnost, že dívka bude mít dlouhé vlasy, je 0,9. Spočítejte

(a) pravděpodobnost, že náhodně vybraný chlapec bude dlouhovlasý,
(b) pravděpodobnost, že náhodně vybraná dívka bude krátkovlasá,
(c) náhodně vybraný student bude dlouhovlasý chlapec,
(d) náhodně vybraný student bude krátkovlasá dívka,
(e) náhodně vybraný student bude mít dlouhé vlasy.
(f) Se zavázanýma očima jsme náhodně vybrali studenta a dotekem zjistili, že má krátké vlasy. Jaká je pravděpodobnost, že je to dívka?
(g) Opět jsme naslepo vybrali studenta a dotekem zjistili, že je dlouhovlasý. Určete pravděpodobnost, že vybraný student je
   - chlapec,
   - dívka.

Domácí cvičení 1. Jaká je pravděpodobnost, že v rodině se třemi dětmi jsou dvě dívky a jeden chlapec, víte-li, že pravděpodobnost narození chlapce je 3/5 a dívky je 2/5.

Domácí cvičení 2. (inspirován ženským finále Wimbledonu 2015)
Pravděpodobnost, že tenistka dá první servis, je 0,6. Pravděpodobnost (podmíněná!), že udělá dvojchybu, je 0,2. Spočtěte, jaká je pravděpodobnost, že tenistka

(i) dá druhý servis (nepodmíněná pravděpodobnost),

(ii) nezkazila první servis, pokud víme, že neudělala dvojchybu.

Domácí cvičení 3. Manžel nepřišel včas ze zaměstnání. Manželka ze zkušenosti ví, že s pravděpodobností 0,3 (resp. 0,6; 0,1) pracuje přesčas (resp. odpočívá v hospodě, zdržel se z jiné příčiny). Pravděpodobnosti, že manžel bude ve 20 hodin doma jsou, podle toho, kde se zdržel, 0,7; 0,2; 0,9.

(i) Jaká je pravděpodobnost, že manžel nebude ve 20 hodin doma?

(ii) Manžel nakonec ve 20 hodin doma byl. Jaká je pravděpodobnost, že se zdržel v hospodě?