Poznámky k cvičeniam pre informatikov z predmetu Základy pravdepodobnosti a štatistiky 2023/24

Askar Gafurov

27. novembra 2023

Úvod

Daný materiál slúži ako doplnok k cvičeniam z predmetu Základy pravdepodobnosti a štatistiky v zimnom semestri akademického roku 2023/24. Cieľom materiálu je poskytnúť prehľad úloh z cvičení a techník na ich riešenie, nie však kompletné riešenia. Tak isto budem tu priebežne dopĺňať odkazy na Google Colab so simuláciami a ďalšími programátorskými dodatkami k jednotlivým príkladom (tam, kde to dáva zmysel). V rámci textu sa budú objavovať odkazy na tvrdenia zo skrípt a na zbierky úloh (PaS01, C18). Tento text je priebežne aktualizovaný, najnovšiu verziu nájdete na webe.

Tento text nie je náhradou absolvovania cvičení ani prednášok.

Všeobecné rady do života

- Vyvážená diéta a pravidelná fyzická aktivita (najmä počas skúškového obdobia)
- Robte veci priebežne, nie na poslednú chvíľu (Eisenhowerova matica)
- Ak niečo neviete, opýtajte sa spolužiaka
- Ak to nevie ani spolužiak, opýtajte sa cvičiaceho. Konzultačné hodiny sú vaše právo a naša povinnosť, stačí napísať e-mail a dohodnúť si termín.
- Ak Vám niečo nevyhovuje, treba sa ozvať. Neodznevšie[1] očakávania sú Váš a iba Váš problém, preto ich vyslovte, nech je to problém spoločný (anonymná spätná väzba k týmto cvičeniam).
- Skúste Zettelkasten / Obsidian, Anki cards.
- Nikdy nejedzte žltý sneh.

[1] činné príčastie minulého času od slovesa *neodznieť*.

1 Cvičenie 1

Na tomto cvičení sa preberali príklady **C18.1, C18.3, C18.5, C18.6, PaS01.1.1, PaS01.1.2, PaS01.1.3, PaS01.1.5**. Simulácie k príkladom z tohto cvičenia nájdete tu: Google Colab.

Úlohy na samostatné precvičenie (a potenciálne aj nakódenie): **C18.{2, 4, 7}, PaS01.1.{4, 9 – 16}**. Taktiež si pozrite video *Grading is a Scam*.

Úloha 1 (C18.1): Nech $\Omega = \{a, b, c\}$. Nech ďalej $S_1 = \{\Omega, \varnothing, \{a\}, \{b\}, \{c\}\}$, $S_2 = \{\Omega, \varnothing, \{a, b\}, \{b, c\}\}$, $S_3 = \{\Omega, \varnothing, \{a\}, \{b, c\}\}$. Rozhodnite, ktoré z týchto množín tvoria spolu so základnou množinou Ω σ-algebru[2][3]

Odpoveď: Iba (Ω, S_3) je σ-algebrou.

***Postup*:** Treba ručne overiť, či sú splnené všetky podmienky z definície σ-algebry (definícia 1.1).

Úloha 2 (C18.3): TODO

Odpoveď: a) $S = \{\varnothing, \{a\}, \{b\}, \{a, b\}, \{c, d\}, \{a, c, d\}, \{b, c, d\}, \Omega\}$; b) S je systém všetkých spočítateľných podmnožín $\mathbb{R}$ a ich komplementov.

Postup: Kým nejaká z podmienok z definície σ-algebry je porušená, treba chybajúcu množinu doplniť.

Úloha 3 (C18.5): Nech (Ω, S, P) je pravdepodobnostný priestor[4] a nech A, B, C sú udalosti[5] na tomto priestore. Nech $B \subseteq A$, $P(A) = P(C) = \frac{1}{2}$, $P(B) = P(A \cap C) = \frac{1}{4}$ a $P(B \cap C) = \frac{1}{8}$. Nájdite pravdepodobnosť udalostí $A - (B \cup C)$ a $\Omega - (A \cup B \cup C)$!

Odpoveď: $P(A - (B \cup C)) = 1/8$, $P(\Omega - (A \cup B \cup C)) = 1/4$

Postup: Pomocou Vennovho[6] diagramu (Obrázok 1) vieme informácie o pravdepodobnostnej miere jednotlivých udalostí prepísať ako informácie o súčtoch pravdepodobnostných mier jednotlivých "atomických"[7] udalostí (označených na obrázku 1 malými písmenami[8]):

$$P(A) = \frac{1}{2} \rightsquigarrow \mathbf{a} + \mathbf{b} + \mathbf{d} + \mathbf{e} = \frac{\mathbf{1}}{\mathbf{2}}$$

$$B \subseteq A \Longrightarrow B - A = \varnothing \overset{\text{def } 1.25}{\Longrightarrow} P(B - A) = 0 \rightsquigarrow c + f = 0 \overset{(*)}{\Rightarrow} \mathbf{c} = \mathbf{0} \wedge \mathbf{f} = \mathbf{0}$$

$$\ldots$$

[2] definícia 1.1

[3] Ak si kladiete v duchu otázku "*načo si to všetko komplikujeme σ-algebrami*?", tak v tom nie ste sami. Jeden z dôvodov je zaistenie, že sa "objem" množín správa rozumne. Napríklad, že si množina bodov v 3D priestore nezväčší objem tým, že si ju otočíme o 90 stupňov. Príklady množín, ktoré sú toho schopné, si vypracovali páni Banach a Tarski. Môžete si pozrieť o tom video od Vsauce.

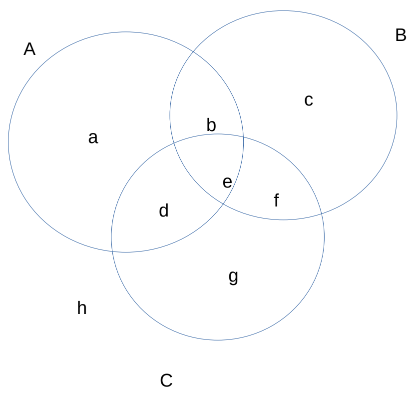
[4] definícia 1.27

[5] definícia 1.9

[6] https://en.wikipedia.org/wiki/Venn_diagram

[7] dá sa to vnímať ako minimálnu σ-algebru indukovanú udalosťami A, B, C nad základnou množinou Ω, v zmysle definície 1.15

[8] pre puristov: používaním malých písmen na označovanie samotných "atomických" udalostí aj ich pravdepodobnostných mier mierne zneužívame zavedené značenie. Pre informatikov by však takéto preťažovanie (angl. *overloading*) významu nemalo robiť žiaden problém :)



Obr. 1: Vennov diagram pre množiny A, B a C.

Všimnite si, že pravdepodobnostná miera sa pohybuje medzi 0 a 1, a teda z informácie $c + f = 0$ automaticky plynie, že obidve prislušné udalostí majú pravdepodobnostnú mieru 0. Takisto treba zakomponovať informáciu o tom, že všetkých osem "atomických" udalostí spolu tvoria rozklad celého pravdepodobnostného priestoru Ω (hint: akú pravdepodobnostnú mieru bude mať zjednotenie všetkých ôsmich "atomických" udalostí?).

Dostaneme tak sústavu lineárnych rovníc[9] s ôsmimi premennými[10]. Jej vyriešením (či už ručne alebo pomocou všeobecnej Gaussovej eliminačnej metódy[11]) dostaneme všetky potrebné informácie na zístenie požadovaných odpovedí (ako vieme vyjadriť udalosti $A - (B \cup C)$ a $\Omega - (A \cup B \cup C)$ pomocou zjednotenia "atomických" udalostí?).

Úloha 4 (C18.6): TODO (symetrickú diferenciu označujeme symbolom $\oplus$ namiesto $\triangle$)

***Odpoveď*:** QED

***Postup*:** Jedna z možností je dokázať silnejšie tvrdenie $A \oplus C \subseteq (A \oplus B) \cup (B \oplus C)$. Alternatívne, vieme použiť Vennov diagram rovnako ako v predošlej úlohe.

Úloha 5 (PaS01.1.1): TODO

Odpoveď: a) $\frac{1}{6}$; b) $\frac{5}{12}$.

Postup: Základný priestor udalostí je množina dvojíc výsledkov na prvej a druhej kocke $\Omega = \{1, \ldots, 6\}^2$.[12] Celkový počet možných udalostí je teda $6^2 = 36$. Je očividné, že všetky základné udalosti sú rovnako pravdepodobné.

[9] https://en.wikipedia.org/wiki/System_of_linear_equations

[10] pre puristov: priamočiary prepis informácií zo zadania by bola sústava lineárnych nerovností, kde by sme mali dopísať 16 nerovnosti typu $a \geq 0$ a $a \leq 1$. Vieme však túto informáciu zakomponovať počas prepisu rovností, prípadne po vyriešení sústavy rovníc.

[11] https://en.wikipedia.org/wiki/Gaussian_elimination

[12] Táto notácia by mala byť povedomá z predmetov *Úvod do diskrétnych štruktúr* a *Formálne jazyky a automaty (1)*.

Pravdepodobnosť udalosti "obidve hodnoty sú rovnaké" tak získame predelením počtu konfigurácií typu (a, a) celkovým počtom možných udalostí. Tento počet vieme spočítať ručne, rovnako ako aj počet konfigurácií pre ktoré plati, že prvé číslo je ostro väčšie než druhé.

Úloha 6 (PaS01.1.2): TODO

Odpoveď: a) $\frac{63}{256}$;[13] b) $\frac{11}{64}$.

Postup: Základný priestor udalostí je množina všetkých možných postupností výsledkov, teda $\Omega = \{\text{“hlava”}, \text{“znak”}\}^{10}$. Pre jednoduchosť to vieme prerobiť na postupnosti núl a jednotiek: $\Omega = \{0, 1\}^{10}$. Je očividné, že všetky základné udalosti sú rovnako pravdepodobné. Pravdepodobnosť práve piatich znakov vieme spočítať ako $\dfrac{\text{počet sekvencií s práve piatimi jednotkami}}{|\Omega|}$. Počet takých sekvencií vieme ľahko dopočítať pomocou kombinačných čísel.

V druhej podúlohe treba sčítať počet sekvencií s práve nula, jednou, dvomi a tromi jednotkami.

Úloha 7 (PaS01.1.3): TODO

Odpoveď: a) $\frac{1}{863\,040}$; b) $\frac{99}{7\,192}$.

Postup: Základný priestor udalostí je množina všetkých permutácií 32 prvkov z množiny

$$\{\text{sedem}, \text{osem}, \text{deväť}, \text{desať}, \text{dolník}, \text{horník}, \text{kráľ}, \text{eso}\} \times \{\text{srdce}, \text{žaluď}, \text{guľa}, \text{list}\}.$$

Celkový počet permutácií je 32!. Je očividné, že všetky základné udalosti sú rovnako pravdepodobné. Treba už len zistiť, koľko sekvencií začína postupnosťou és zo zadania. Inak povedané, koľkými spôsobmi vieme zvyšné (nezafixované) karty premiešať?

V druhej podúlohe nemáme fixnú počiatočnú postupnosť, ale viacero. Ak však zvolíme si jednu z nich fixne, tak počet takých permutácií už poznáme (prvá podúloha). Koľkými spôsobmi vieme vybrať prvé štyri karty? Máme k dispozícii 16 kariet a treba z nich v vybrať usporiadané štvorice. Keďže všetky fixné začiatky vedú k rozličným permutáciám, tak stačí vynásobiť počet spôsobov určiť prvé štyri karty s počtom spôsobov preusporiadať zvyšných 28 kariet.

Úloha 8 (PaS01.1.5): TODO

Odpoveď: a) $\frac{5}{1\,012}$; b) $\frac{5}{23}$.

Postup: Máme dve očividné možnosti ako si zvoliť základný priestor udalostí: a) všetky možné postupnosti výberu guličiek z vrecka od prvej až po dvadsaťštvrtú, b) všetky možné postupnosti výberu prvých troch guličiek z vrecka. Skúsme použiť prvú možnosť, pretože v tomto prípade všetky základné udalosti budú mať rovnakú pravdepodobnosť[14].

Prvý krok je určenie celkového počtu postupností v Ω. Na to sa hodia multinómy[15].

Druhý krok je určenie počtu postupností, ktoré začínajú tromi bielymi guličkami. Môžeme to spočítať ako počet postupností zo zvyšných guličiek (t.j. 2 bielych, 8 modrých a 11 červených) postupom z prvého kroku. Odpoveď na prvú podúlohu získame predelením tohto čísla celkovým počtom základných udalostí.

[13] Pre správne názvoslovie odporúčam si prečítať článok https://pohodovamatematika.sk/co-je-zlomok-citanie-a-zapisovanie-zlomkov.html

[14] druhá možnosť by viedla k tomu, že by sme museli vyriešiť prvú podúlohu už pri definovaní pravdepodobnostného priestoru.

[15] https://en.wikipedia.org/wiki/Multinomial_theorem#Interpretations

Tretí krok je určenie počtu postupností, ktoré začínajú dvomi červenými a jednou modrou v nejakom poradí. My už vieme spočítať počet postupností s úplne fixným začiatkom, teraz máme tri možné začiatky: "červená, červená, modrá", "červená, modrá, červená" a "modrá, červená, červená". Čiže treba sčítať spolu počty postupností s takýmito začiatkami.

Sumár techník:

- Vennov diagram (vieme ho nakresliť na papieri pre dve, tri a štyri množiny)
- Pravdepodobnostné priestory s rovnomernou pravdepodobnosťou a konečnou základnou množinou Ω prevádzajú úlohy typu "*spočítaj pravdepodobnosť udalosti X*" na "*spočítaj počet objektov v Ω, ktoré majú vlastnosť X*". Vieme potom použiť techniky z kombinatoriky: počet permutácií (faktoriály), počet výberov s návratom (mocniny), počet výberov bez návratu (kombinačné čísla, multinómy), atď.
- Ak počet objektov v Ω je malý, tak si môžete všetky udalosti vypísať na papier.

2 Cvičenie 2

Na tomto cvičení sa preberali príklady **C18.8, C18.11, C18.12, C18.14, a tri príklady mimo zbierok**. Simulácie k príkladom z tohto cvičenia nájdete tu: Google Colab.

Úlohy na samostatné precvičenie (a potenciálne aj nakódenie): **C18.{9, 10, 13, 15, 16, 17, 19}, PaS01.1.{6, }**, **1.{35, 41, 42, 43, 50} zo skrípt**.

Úloha 1 (C18.8): TODO

Odpoveď: QED

Postup: Treba postupne overiť podmienky z definície pravdepodobnostnej miery[16]. Zíde sa vedieť pravidlo distributivity[17].

Uloha 2: Na úseku dĺžky 3 rovnomerne náhodne a nezávisle zvolíme body A, B. Aká je pravdepodobnosť, že ich vzdialenosť bude menej než 1?

Uloha 3: Dvaja kolegovia A a B prídu do spoločnej kancelárie medzi ôsmou a štrnástou hodinou. (Predpokladáme, že A a B prichádzajú nezávisle na sebe a príchod každého z nich je rovnako pravdepodobný počas celého šesťhodinového intervalu.) Vieme, že A sa zdrží v kancelárii 1 hodinu a B sa zdrží 2 hodiny. Určte pravdepodobnosť, že B príde neskôr ako A, a zároveň sa A a B stretnú.

Úloha 4 (C18.11, Buffonova ihla): V rovine sú zakreslené rovnobežné priamky s rozostupmi L. Určte pravdepodobnosť toho, že ak na túto rovinu náhodne hodíme ihlu dĺžky $\ell < L$, tak pretne niektorú z priamok.

Odpoveď: $\dfrac{2\ell}{\pi L}$

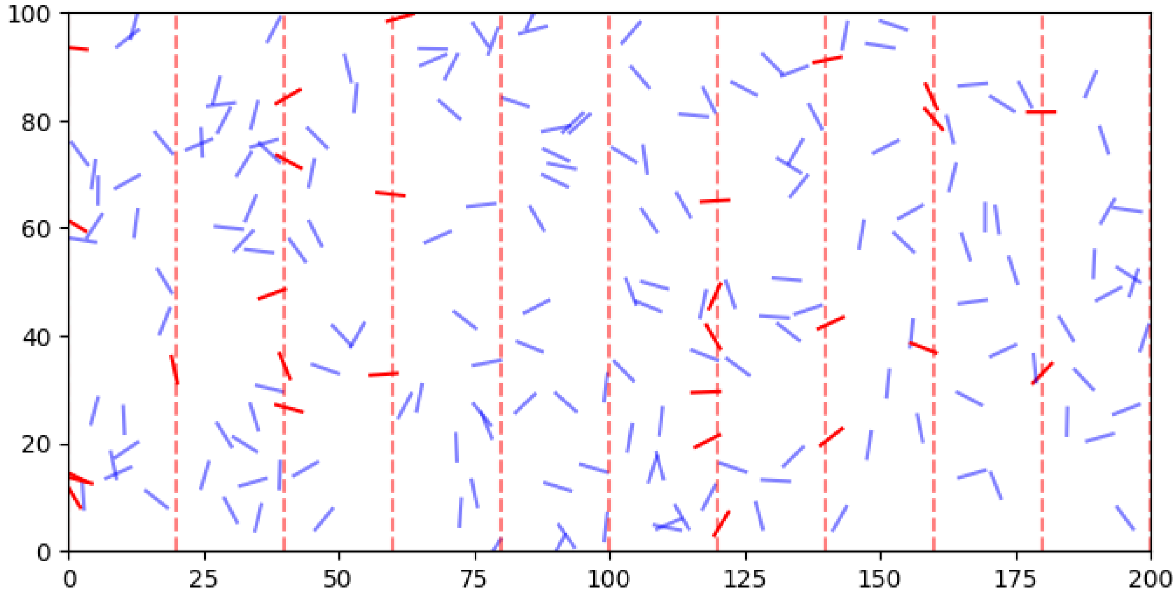
Postup: Dohodneme sa, že priamky sú zvislé. Príklad si môžete pozrieť na obrázku 2.

Prvotnou úlohou je formovanie pravdepodobnostného priestoru. Začneme pár pozorovaniami. Prvé pozorovanie je, že polohu ihly v rovine vieme jednoznačne reprezentovať súradnicami jej stredu (resp. ťažiska) a uhlom rotácie, napríklad voči osi Y. Druhé pozorovanie je, že nám nezáleží na ypsilonovej súradnici stredu ihly. Tretie pozorovanie je, že nám netreba modelovať celú os X, ale iba priestor medzi dvomi susednými priamkami. Štvrté pozorovanie je, že aj v rámci jedného "pruhu" šírky L nám stačí modelovať iba jeho polovicu, lebo ihla vždy bude bližšia k jednej z priamok a môže pretnúť iba ju (vďaka podmienke $\ell < L$). Piate pozorovanie je, že nám stačí skúmať nie celých 360 stupňov rotácie ihly, ale iba 90, keďže jeden z koncov ihly vždy bude bližší k priamke a to, či leží pod alebo nad kolmicou medzi stredom ihly a priamky, je úplne jedno. Šieste pozorovanie je, že uhol rotácie ihly ako náhodný jav je nezávislý od polohy stredu ihly.

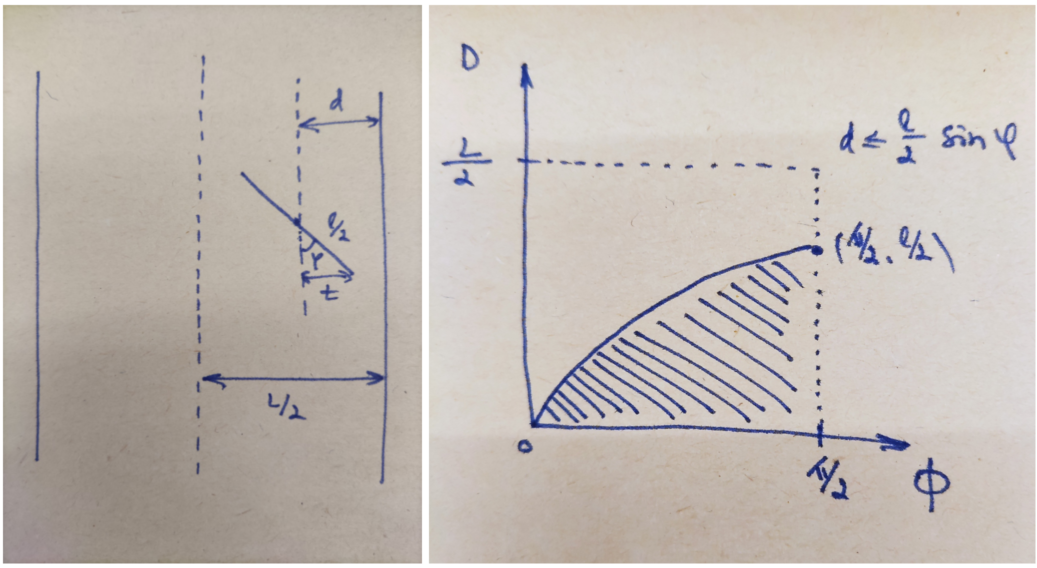
V tomto bode už môžeme skonštruovať pravdepodobnostný priestor. Základnou množinou udalostí bude karteziánsky súčin $\Omega = \left[0, \frac{\pi}{2}\right] \times \left[0, \frac{L}{2}\right]$, kde prvá zložka bude bude uhol rotácie Φ voči osi Y a druhá zložka vzdialenosť D medzi stredom ihly a priamkou. Kľúčový je fakt, že jednotlivé zložky (súradnice) v množine udalostí sú od seba *nezávislé*. Množinu merateľných množín zvolíme ako štandardnú sadu borelovských množín[18], a pravdepodobnostná miera bude rovnomerná, t.j. pravdepodobnosť obdĺžnika vo vnútri priestoru Ω bude pomerom jeho obsahu k celému obsahu Ω:

$$P([a, b] \times [c, d]) = \frac{(b-a)\cdot(d-c)}{\frac{\pi}{2}\cdot\frac{L}{2}},$$

[16] definícia 1.25
[17] https://en.wikipedia.org/wiki/Union_(set_theory)
[18] príklad 1.32 zo skrípt



Obr. 2: Ilustrácia k úlohe Buffonovej ihly. Červeným sú znázornené ihly, ktoré pretínajú zvislé čiary.



Obr. 3: Obrazok k ulohe C18.11 (Buffon's needle)

pre každé $0 \leq a \leq b \leq \frac{\pi}{2}$ a $0 \leq c \leq d \leq \frac{L}{2}$. Takéto konštrukcie sa občas volajú *geometrické*[19], lebo umožňujú počítať pravdepodobnosti udalostí cez výpočet plôch (resp. objemov) geometrických útvarov. Presne takto ideme počítať aj tento príklad.

Ideme teda sa pozrieť na udalosť

$$A := \{(\phi, d) \in \Omega : \text{ihla so stredom vo vzdialenosti } d \text{ a s uhlom } \phi \text{ pretína priamku}\}.$$

Pre ktoré dvojice vzdialenosti d a uhlu ϕ toto tvrdenie platí? Pozrime sa na kolmý priemet bližšej polovice ihly na os X. Ak ihla pretína priamku, tak ju pretína aj jej kolmý priemet. Aká je jeho dĺžka? Vďaka goniometrii vieme, že je to $\frac{\ell}{2} \cdot \sin \phi$. A teda, ak dĺžka priemetu je väčšia ako vzdialenosť stredu ihly od priamky, tak priamka je danou ihlou pretnutá:

$$A = \left\{(\phi, d) \in \Omega : d \leq \frac{\ell}{2} \cdot \sin \phi\right\}.$$

A teda, udalosti A v množine Ω zodpovedá plocha pod krivkou $d = \frac{\ell}{2} \cdot \sin \phi$ na rozsahu $\left[0, \frac{\pi}{2}\right]$, ak na vodorovnú os dáme hodnoty ϕ a na zvislú os dáme hodnoty d (obrázok 3). Obsah plochy pod krivkou vieme spočítať ako určitý integrál $\int_0^{\pi/2} \frac{\ell}{2} \cdot \sin \phi \, d\phi$. A teda, pravdepodobnosť pretnutia je rovná

$$P(A) = \frac{\displaystyle\int_0^{\pi/2} \frac{\ell}{2} \cdot \sin \phi \, d\phi}{\frac{\pi}{2} \cdot \frac{L}{2}} = \frac{\left[-\frac{\ell}{2} \cdot \cos \phi\right]_{\phi=0}^{\pi/2}}{\frac{\pi}{2} \cdot \frac{L}{2}} = \frac{\frac{\ell}{2}}{\frac{\pi}{2} \cdot \frac{L}{2}} = \frac{2\ell}{\pi L}.$$

Pozrite si taktiež materiály na Google Colab, je tam ukázaný aj simulačný spôsob odvodenia tohto vzorca aj počítania tohto určitého integrálu.

Úloha 5 (C18.12): TODO

Odpoveď: a) $\frac{1}{4}$; b) $\frac{15}{16}$; c) $\frac{3}{4}$.

Postup: Geometrickou konštrukciou. Jednotlivými súradnicami budú polohy prvej a druhej bodky na úsečke. V podúlohe a) je potrebné si uvedomiť, aké podmienky na polohy bodov vyplývajú z trojuholníkovej nerovnosti. Ďalej si je potrebné uvedomiť, že podmienka "*minimum z množiny je väčšie ako X*" je ekvivalentná s podmienkou "*každé číslo z množiny je väčšie ako X*".

Úloha 4 (o stoličkách): Za okrúhlym stolom je 6 stoličiek. Náhodne za tento stôl rozsadíme 6 ľudí, ktorí tvoria 3 manželské páry. Aká je pravdepodobnosť, že aspoň jeden manželský pár bude sedieť oproti sebe?

Odpoveď: $\frac{7}{15}$

Postup: Princípom zapojenia-vypojenia[20]. Nech udalosť "*i-tý pár sedí oproti sebe*" označíme ako A_i, potom skúmaná udalosť "*aspoň jeden pár sedí oproti sebe*" je vyjadriteľná ako $A_1 \cup A_2 \cup A_3$. Zostáva už len spočítať potrebné prieniky zo vzorca.

Úloha 6 (C18.14): TODO

[19] príklad 1.34 zo skrípt
[20] veta 1.40 a príklad 1.41 zo skrípt

Odpoveď: $4q^3 - 6q^5 + 3q^6$, kde $q := 1 - p$.

Postup: Krkolomným rozborom prípadov[21]. Alebo princípom zapojenia-vypojenia. Ak označíme udalosť "*i-tý vrchol je izolovaný*" ako A_i, tak skúmanú udalosť "*aspoň jeden vrchol je izolovaný*" je vyjadriteľná ako $A_1 \cup A_2 \cup A_3 \cup A_4$.

Uloha 7 (nebola): Predpokladajme, že koeficienty kvadratickej rovnice $x^2 + px + q = 0$ sú náhodné, vyhovujú podmienke $|\,p\,| \leq 1$ a $|\,q\,| \leq 1$ a sú na celých prípustných intervaloch rovnako pravdepodobné. Určte pravdepodobnosť, že korene kvadratickej rovnice sú reálne čísla.

Sumár techník:

- Vlastnosti množín (disributivita, de Morgan, atď.).
- Riešenie cez geometrické konštrukcie. Vlastné určité integrály.
- Prevod výrokov typu "*minimum (maximum) je väčšie (menšie) ako X*" na výrok typu "*každé číslo je väčšie (menšie) ako X*". V budúcnosti sa táto technika rozšíri na takzvané *poriadkové štatistiky*.
- Princíp zapojenia-vypojenia.
- Nezávislé udalosti, výpočet ich združenej pravdepodobnosti.

[21] Takto danú úlohu riešil váš pokorný sluha, keď sa s ňou prvýkrát stretol :)

3 Cvičenie 3

Na tomto cvičení sa preberali príklady **C18.18, C18.20, C18.21, C18.23, C18.24, C18.25 a jedna úloha mimo zbierok**. Simulácie k príkladom z tohto cvičenia nájdete tu: Google Colab. Navyše, je tam aj ukážka počítania aposteriórnych pravdepodobností.

Úlohy na samostatné precvičenie (a potenciálne aj nakódenie): **C18.{26, 27, 29,}**.

Úloha 1 (C18.18): TODO

Odpoveď: 0.524

Postup: Treba si uvedomiť, že krvná skupina prvého a druhého človeka sú pre všetky účely a potreby nezávislé[22] od seba. A teda, vieme združenú pravdepodobnosť dvojice krvných skupín spočítať ako súčin jednotlivých pravdepodobností. Ďalej, drevorubačským sčítaním pravdepodobností všetkých vyhovujúcich kombinácií typov krvi dostaneme hľadanú odpoveď. Jediný rozdiel oproti príkladom ako PaS01.1.1 je, že jednotlivé elementy množiny Ω majú rozličné pravdepodobnosti.

TODO prerobit riesenie tak, aby pouzivalo vetu o uplnej pravdepodobnosti

Úloha 2 (C18.20): TODO

Odpoveď: TODO

Postup: Cez vzorec úplnej pravdepodobnosti[23]. Ak si definujeme pravdepodobnostný priestor so základnou množinou $\Omega = \{\mathbf{W}, \mathbf{B}, \mathbf{R}\}^2$, kde prvá zložka určuje výsledok prvého potahnutia, a druhá zložka — druhého, tak vieme urobiť *rozklad množiny* Ω podľa hodnoty prvej zložky, a následne aplikovať vzorec z vety o úplnej pravdepodobnosti:

$$\Pr[(*, \mathbf{B})] \overset{\text{veta 2.8}}{=} \sum_{c \in \{\mathbf{W}, \mathbf{B}, \mathbf{R}\}} \Pr[(*, \mathbf{B}) \mid (c, *)] \cdot \Pr[(c, *)],$$

respektíve v ľudskej reči:

$$\Pr[\text{druhá gulička je biela}] \overset{\text{veta 2.8}}{=} \sum_{c \in \{\mathbf{W}, \mathbf{B}, \mathbf{R}\}} \Pr[\text{druhá gulička je biela} \mid \text{prvá gulička je farby } c] \cdot \Pr[\text{prvá gulička je farby } c],$$

Hodnoty $\Pr[(c, *)]$ pre jednotlivé farby c máme zo zadania, a hodnoty $\Pr[(*, \mathbf{B}) \mid (c, *)]$ vieme dopočítať pre každú farbu c osobitne.

Úloha 3 (C18.21): TODO

Odpoveď: a) $\approx 8.7\%$; b) $\approx 90\%$.

[22] tu narážame na úlohu teórie pravdepodobnosti modelovať reálny svet. Samozrejme, výberom jedného človeka z populácie s krvnou skupinou X *de iure* zmenšujeme pravdepodobnosť, že druhý bude mať tú istú krvnú skupinu. Pri populácii v niekoľko miliónov je tento rozdiel ale zanedbateľný. Ako vravel veľký britský štatistik George Box: "*Všetky modely sú zlé, niektoré sú užitočné*".

[23] veta 2.8

Postup: Cez Bayesov vzorec[24]. Označme si udalosť "*pacient je infikovaný*" ako I, udalosť "*prvý test je pozitívny*" ako P_1 a udalosť "*druhý test je pozitívny*" ako P_2. Preveďme informácie o teste zo zadania do reči teórie pravdepodobnosti:

- "*podozrenie (pred testom) na chorobu je* 0.1%" $\rightsquigarrow \Pr[I] = 0.001$
- "*test dáva pozitívny výsledok v* 95% *prípadoch pre infikovaných*" $\rightsquigarrow \Pr[P_1 \mid I] = 0.95$
- "*test dáva negatívny výsledok v* 99% *prípadoch pre neinfikovaných*" $\rightsquigarrow \Pr[\neg P_1 \mid \neg I] = 0.99$

Prvá otázka z úlohy sa nás pýta na pravdepodobnosť udalosti "*pacient je infikovaný **za predpokladu, že prvý test je pozitívny***", teda na hodnotu $\Pr[I \mid P_1]$. Aplikáciou Bayesovho vzorca na túto pravdepodobnosť dostávame:

$$\Pr[I \mid P_1] \overset{\text{veta 2.9}}{=} \frac{\Pr[P_1 \mid I] \cdot \Pr[I]}{\Pr[P_1 \mid I] \cdot \Pr[I] + \Pr[P_1 \mid \neg I] \cdot \Pr[\neg I]}.$$

S prekvapením zistíme, že všetky hodnoty na pravej strane vieme vyčítať zo zadania.

Pri druhej podúlohe potrebujeme spočítať pravdepodobnosť $\Pr[I \mid P_1 \cap P_2]$. Aplikáciou Bayesovho vzorca dostaneme:

$$\Pr[I \mid P_1 \cap P_2] \overset{\text{veta 2.9}}{=} \frac{\Pr[P_1 \cap P_2 \mid I] \cdot \Pr[I]}{\Pr[P_1 \cap P_2 \mid I] \cdot \Pr[I] + \Pr[P_1 \cap P_2 \mid \neg I] \cdot \Pr[\neg I]}.$$

Ostáva už len uvedomiť si, že počítanie $\Pr[P_1 \cap P_2 \mid I]$ je jednoduché, keďže podľa zadania sú udalosti P_1 a P_2 nezávislé.

Morálne ponaučenie tohto príkladu je, že presnosť medicínskeho testu negarantuje jeho predikčnú schopnosť[25]. Odporúčam si pozrieť aj YouTube video od 3Blue1Brown k tomu.

Tento príklad je ukážkou *bayesovskej štatistiky*. Cieľom je kombinovanie predchádzajúcich (*apriórnych*) predpokladov s pozorovanými dátami. V danom príklade, podozrenie na chorobu $\Pr[I]$ sa volá *apriórnou pravdepodobnosťou* udalosti I ("*pacient je infikovaný*"). Výsledky testov P_1 a P_2 sú *pozorovania*. Podmienená pravdepodobnosť udalosti I vzhľadom na pozorovania P_1 a P_2 $\Pr[I \mid P_1, P_2]$ sa volá *aposteriórnou pravdepodobnosťou* udalosti I.

Úloha 4 (C18.23): TODO

Odpoveď: a) ≈ 0.61, b) ≈ 0.017.

Postup: Potrebujeme spočítať pravdepodobnosť udalosti "*Trafil strelec A **za predpokladu,** že trafil práve jeden strelec*". Označme si udalosť "*trafil práve jeden strelec*" ako T_1 a udalosť "*Trafil strelec X*" ako T_X. Potom skúmaná udalosť "*Trafil strelec A **za predpokladu,** že trafil práve jeden strelec*" je vyjadriteľná ako $T_A \mid T_1$. Pravdepodobnosť tejto podmienenej udalosti vieme spočítať z definície podmienenej pravdepodobnosti:

$$\Pr[T_A \mid T_1] \overset{\text{def 2.1}}{=} \frac{\Pr[T_A \cap T_1]}{\Pr[T_1]},$$

ak poznáme hodnoty $\Pr[T_A \cap T_1]$ a $\Pr[T_1]$. Poďme ich spočítať!

Udalosť $T_A \cap T_1$ ("*strelec A trafil **a** trafil práve jeden strelec*") je ekvivalentná s udalosťou "*strelec A trafil **a** strelec B netrafil **a** strelec C netrafil **a** strelec D netrafil*", čiže $T_A \cap T_1 = T_A \cap \neg T_B \cap \neg T_C \cap \neg T_D$. Môžeme si všimnúť, že jednotlivé udalosti v rámci tohto prieniku sú združene nezávislé. A teda, vieme spočítať ten prienik ako súčin už známych zo zadania hodnôt:

$$\Pr[T_A \cap T_1] = \Pr[T_A] \cdot \Pr[\neg T_B] \cdot \Pr[\neg T_C] \cdot \Pr[\neg T_D].$$

[24] veta 2.9

[25] Spomeňme si na celoslovenské plošné testovania na COVID-19 na jeseň roku 2020

Všimneme si, že množina udalostí $\{T_A \cap T_1, T_B \cap T_1, T_C \cap T_1, T_D \cap T_1\}$ tvorí rozklad udalosti T_1 (niekto z nich musel trafiť, a nemohli to byť viacerí naraz). A teda (vďaka definícii pravdepodobnostnej miery), pravdepodobnosť udalosti T_1 vieme napísať ako súčet ich pravdepodobností:

$$\Pr[T] = \sum_{X \in \{A,B,C,D\}} \Pr[T_X \cap T_1].$$

Úloha 5 (C18.24): TODO

Odpoveď: QED.

Postup: Priame dôkazy. Oplatí sa pamätať, že pole reálnych čísel (ako aj každé iné pole) je oborom integrity[26].

Úloha 6 (C18.25): TODO

Odpoveď: QED.

Postup: Priame dôkazy.

Úloha 7*: Generujeme reťazec nasledovným spôsobom: so šancou $\frac{1}{3}$ pridáme na koniec písmeno a, so šancou $\frac{1}{3}$ pridáme na koniec písmeno b, a so šancou $\frac{1}{3}$ ukončíme generovanie reťazca. S akou pravdepodobnosťou po ukončení generovania dostaneme palindróm (reťazec rovný svojmu reverzu)?

Odpoveď: $\dfrac{5}{7}$.

Postup: Vzorcom z vety o úplnej pravdepodobnosti.

Sumár techník:

- Podmienené pravdepodobnosti.
- Rozklad udalosti.
- Veta o úplnej pravdepodobnosti.
- Bayesov vzorec.

[26] Obor integrity je (komutatívny) netriviálny okruh *bez deliteľov nuly*, t.j. $\forall a, b \in \mathbb{R} : ab = 0 \Rightarrow (a = 0 \vee b = 0)$

4 Cvičenie 4

Na tomto cvičení sa preberali príklady **C18.28, C18.30, C18.32, C18.33, C18.34, C18.36, C18.37**. Simulácie k príkladom z tohto cvičenia nájdete tu: Google Colab.

Úlohy na samostatné precvičenie (a potenciálne aj nakódenie): **C18.{31, 35, 38, 39}**.

Úloha 1 (C18.28): TODO

Odpoveď: a) ≈ 0.0563; b) ≈ 0.99958; c) ≈ 0.6778.

Postup: Začneme budovaním pravdepodobnostného priestoru. Základná množina udalosti môžu byť jednotlivé *usporiadané* desatice farieb potomkov, čiže $\Omega = \{\mathbf{W}, \mathbf{P}\}^{10}$. Keďže Ω je konečná, tak môžeme nastaviť množinu S na 2^{Ω}. Táto konštrukcia má výhodu, že jednotlivé zložky tých desatíc sú *združene nezávislé*, a teda výpočet pravdepodobnosti jednotlivej desatice je len súčin pravdepodobností pre jednotlivých potomkov:

$$\Pr[(a_1, \ldots, a_{10})] \overset{\text{nezávislosť}}{=} \prod_{i=1}^{10} \Pr[\text{potomok } i \text{ má farbu } a_i] = \prod_{i=1}^{10} p_{a_i},$$

kde $p_{\mathbf{P}} = \frac{3}{4}$ a $p_{\mathbf{W}} = \frac{1}{4}$ (zo zadania).

Môžeme si ďalej všimnúť, že pri výpočte pravdepodobnosti jednej desatice nám nezáleží na poradí, ale len na počte fialových a bielych potomkov:

$$\Pr[(a_1, \ldots, a_{10})] = p_{\mathbf{P}}^{\#_{\mathbf{P}}} \cdot p_{\mathbf{W}}^{\#_W} = p_{\mathbf{P}}^{\#_{\mathbf{P}}} \cdot (1 - p_{\mathbf{P}})^{10 - \#_{\mathbf{P}}},$$

kde $\#_{\mathbf{P}}$ a $\#_{\mathbf{W}}$ sú počty fialových a bielych potomkov v desatici $(a_1, \ldots, a_{10})$. To znamená, že všetky desatice s rovnakým počtom fialových potomkov majú rovnakú pravdepodobnosť. Týmto je konštrukcia pravdepodobnostného priestoru ukončená.

Zadanú úlohu vieme vyriešiť, ak by sme vedeli počítať pravdepodobnosť udalosti "*práve k potomkov má fialovú farbu*". Označme si ju ako A_k. Do množiny A_k patria práve všetky desatice s práve k fialovými potomkami. My už ale vieme, že každý z nich má rovnakú pravdepodobnosť $p_{\mathbf{P}}^k \cdot (1 - p_{\mathbf{P}})^{10-k}$. Čiže zostáva už len spočítať *počet desatíc s práve k fialovými potomkami* a vynásobiť ten počet pravdepodobnosťou konkrétnej desatice. Ten počet desatíc ale už vieme počítať (viď príklad PaS01.1.2 z prvého cvičenia). Dostávame tak vzorec

$$\Pr[A_k] = \binom{10}{k} p_{\mathbf{P}}^k (1 - p_{\mathbf{P}})^{10-k}.$$

Úloha 2 (C18.30): TODO

Odpoveď: QED.

Postup: Obidve implikácie vieme dokázať priamo.

Úloha 3 (C18.32): TODO

Odpoveď: a) $F_X(x) = \begin{cases} 0 & x \le 1 \\ \frac{1}{6} & 1 < x \le 2 \\ \frac{2}{6} & 2 < x \le 3 \\ \frac{3}{6} & 3 < x \le 4 \\ \frac{4}{6} & 4 < x \le 5 \\ \frac{5}{6} & 5 < x \le 6 \\ 1 & 6 < x \end{cases}$, $\mathrm{E}[X] = 3.5$; b) $F_X(x) = \begin{cases} 0 & x \le 0 \\ \frac{1}{6} & 0 < x \le 1 \\ \frac{3}{6} & 1 < x \le 2 \\ \frac{5}{6} & 2 < x \le 3 \\ 1 & 3 < x \end{cases}$, $\mathrm{E}[X] = 1.5$.

Postup: Pozorným čítaním definícií náhodnej premennej (definícia 3.1), distribučnej funkcie (definícia 3.13) a strednej hodnoty (definícia 4.6). Ako prvý krok odporúčam vypísať všetky nenulové body[27] funkcie $\Pr[X = k]$ (vzhľadom na k).

Úloha 4 (C18.33): TODO

Odpoveď: $\mathrm{E}[X] = 2$, $\mathrm{Var}[X] = \frac{2}{3}$.

Postup: Nech c_k je výsledný počet kociek s k zafarbenými stenami. Potom

$$\Pr[X = k] = \frac{c_k}{27}.$$

Zostáva už len dopočítať strednú hodnotu a varianciu[28] (definícia 4.14).

Úloha 5 (C18.34): TODO

Odpoveď: $\Pr[X = k] = \begin{cases} \frac{2(n-k)}{n(n-1)} & 1 \le k \le n-1 \\ 0 & inak \end{cases}$, $\mathrm{E}[X] = \dfrac{n+1}{3}$.

Postup: Keď skúmame minimum, oplatí sa najprv počítať hodnotu $\Pr[X \ge k]$. Z nej už vieme dopočítať $\Pr[X = k] = \Pr[X \ge k] - \Pr[X \ge k-1]$, a z toho už vieme dopočítať aj strednú hodnotu z jej definície.

Úloha 6 (C18.36): TODO

Odpoveď: $e - 1$.

Postup: Spočítame hodnotu $\Pr[X = k]$. Z nej spočítame strednú hodnotu.

Úloha 7 (C18.37): TODO

Odpoveď: QED.

Postup: Výmenou poradia sumácie:

$$\mathrm{E}[X] = \sum_{k=1}^{\infty} k \cdot \Pr[X = k] = \sum_{k=1}^{\infty} \sum_{j=1}^{k} \Pr[X = k] \overset{\text{výmena súm}}{=} \sum_{j=1}^{\infty} \sum_{k=j}^{\infty} \Pr[X = k] \overset{\text{X je diskretná}}{=} \sum_{j=1}^{\infty} \Pr[X \ge j].$$

[27] množina všetkých nenulových bodov funkcie sa vo všeobecnosti volá *nosič* funkcie (angl. *support*)

[28] učebnicový názov je *rozptyl*, a namiesto značenia $\mathrm{Var}[X]$ sa používa $D[X]$. Táto hodnota je známa v slovenčine aj pod názvami *disperzia*, *stredná kvadratická odchýlka*, *stredná kvadratická fluktuácia*, či *druhý centrálny moment*.

Úloha 8* (príklad 5.4 z Anděla): Predpokladajme, že sa konajú nezávislé pokusy. V každom z nich nastane s rovnakou pravdepodobnosťou $\frac{1}{m}$ práve jeden z javov $A_1, \ldots, A_m$. Vypočítajte strednú hodnotu počtu pokusov, ktoré treba uskutočniť na to, aby sa niektorý z javov vyskytol prvýkrát neprerušovane k-krát za sebou!

Odpoveď: $\dfrac{m^k - 1}{m - 1}$.

Postup: Označme si X_i ako počet pokusov potrebných na to, aby sme uvideli k rovnakých výsledkov za sebou, ak predošlých i pokusov bolo rovnakých. Potom platia nasledovné (rekurentné) rovnice:

$$
\begin{aligned}
\mathrm{E}[X_k] &= 0 \\
\mathrm{E}[X_i] &= \frac{1}{m} \cdot (1 + \mathrm{E}[X_{i+1}]) + \frac{m-1}{m} \cdot (1 + \mathrm{E}[X_1]) \text{ pre } \forall k \in \{0, \ldots k-1\}
\end{aligned}
$$

Potrebujeme spočítať teda $\mathrm{E}[X_0]$. Túto rekurentnú rovnicu vieme vyriešiť opatrným rozpisovaním od $\mathrm{E}[X_0]$ po $\mathrm{E}[X_{k-1}]$.

5 Cvičenie 5

Na tomto cvičení sa preberali príklady **C18.38, C18.42, C18.43, C18.44, C18.46, C18.47, C18.48**.

Sumár techník:

- Výpočet súm pomocou integrovania-derivovania: $f(x) = \frac{d}{dx}\left(\int f(x)\,dx\right)$. Binomická formula. Súčet konečného a nekonečného geometrického radu.
- Metóda indikátorov na výpočet stredných hodnôt a rozptylu.

6 Cvičenie 6

Na tomto cvičení sa preberali príklady **C18.38, C18.42, C18.43, C18.44, C18.46, C18.47, C18.48** a vybrané príklady z minuloročného midtermu.

Sumár techník:

- Výpočet strednej hodnoty a variancie pomocou štandardných diskrétnych distribúcií: rovnomerné diskretné rozdelenie, Bernoulliho rozdelenie, binomické rozdelenie, geometrické rozdelenie, Poissonové rozdelenie, hypergeometrické rozdelenie, multinomické rozdelenie

- Aproximácia binomického rozdelenia Poissonovym rozdelením

7 Cvičenie 7

Na tomto cvičení sa preberali príklady **C18.53, C18.54, C18.56, C18.57**.

Sumár techník:

- Konvergentné nevlastné integrály s nekonečnými hranicami. Vlastnosti priamych a inverzných goniometrických funkcií. Derivácia inverznej funkcie.
- Výpočet hustoty z distribučnej funkcie derivovaním
- Výpočet distribučnej funkcie z hustoty integrovaním
- Výpočet strednej hodnoty a variancie pomocou hustoty
- Výpočet distribučnej funkcie, hustoty, strednej hodnoty a variancie prostej transformácie absolútne spojitej náhodnej premennej

8 Cvičenie 8

Na tomto cvičení sa preberali príklady **C18.58, C18.60-C18.64**.

Sumár techník:

- Limitné vlastnosti distribučnej funkcie
- Prevod výpočtov o normálnom rozdelení na štandardizované normálne rozdelenie, použitie kvantilovej funkcie.

9 Cvičenie 10

Na tomto cvičení sa preberali príklady **C18.69, C18.71, C18.77, C18.80** a jeden príklad mimo zbierky.

Sumár techník:

- Výpočet pravdepodobností náhodných vektorov pomocou viacrozmerných integrálov. Prevod viacrozmerných integrálov na viacnásobné integrály. Počítanie viacnásobných integrálov.
- Distribučná funkcia minima a maxima nezávislých náhodných premenných. Poriadkové štatistiky.