Pravděpodobnost II

2. SERIÁLOVÁ SÉRIE — TERMÍN ODESLÁNÍ: 4. ÚNORA 2019

ÚLOHA 1.

Stádečko 250 PraSátek se rozhodlo uspořádat filmovou noc. Začala tím, že vyrobila seznam 100 filmů a z nich nyní chtějí vybrat nějaké, které pak budou promítat. Každé PraSátko má přitom seznam deseti filmů, které se mu líbí, a seznam deseti filmů, které se mu nelíbí (jednotlivé seznamy se mohou libovolně překrývat).

(a) Dokažte, že můžeme vybrat nějaké filmy tak, že každé PraSátko uvidí alespoň jeden film ze seznamu těch, které se mu líbí, ale nejvýše devět filmů ze seznamu těch, které se mu nelíbí. (3 BODY)

(b) Dokažte, že je to možné, i pokud přidáme podmínku, že vybraných filmů musí být nejvýše 50. (2 BODY)

ÚLOHA 2. (5 BODŮ)

Noe vzal na archu n dvojic zvířat. Když po několika dnech začal být hladový, rozhodl se, že vybere k náhodných zvířat a sní je. Jaká je střední hodnota počtu *různých* druhů zvířat, která Noe vybral?[1]

ÚLOHA 3. (5 BODŮ)

Na Matfyzu se sešlo n informatiků a matematiků, přičemž každý matematik zná alespoň jednoho informatika. Ukažte, že můžeme vybrat takovou skupinu matfyzáků o velikosti alespoň $n/2$, aby uvnitř ní každý matematik znal lichý počet informatiků.

[1] Pokud například Noe vybral lva, lvici a jednorožce, jedná se o dva různé druhy.

Pravděpodobnost II

2. SERIÁLOVÁ SÉRIE VZOROVÉ ŘEŠENÍ

Úloha 1.

Stádečko 250 *PraSátek se rozhodlo uspořádat filmovou noc. Začala tím, že vyrobila seznam* 100 *filmů a z nich nyní chtějí vybrat nějaké, které pak budou promítat. Každé PraSátko má přitom seznam deseti filmů, které se mu líbí, a seznam deseti filmů, které se mu nelíbí (jednotlivé seznamy se mohou libovolně překrývat).*

(a) *Dokažte, že můžeme vybrat nějaké filmy tak, že každé PraSátko uvidí alespoň jeden film ze seznamu těch, které se mu líbí, ale nejvýše devět filmů ze seznamu těch, které se mu nelíbí.* *(3 BODY)*

(b) *Dokažte, že je to možné, i pokud přidáme podmínku, že vybraných filmů musí být nejvýše* 50. *(2 BODY)*

(Vašek Rozhoň)

ŘEŠENÍ:

Budeme postupovat pravděpodobnostní metodou jako v řešení první úlohy druhého dílu seriálu. Zvolíme každý film nezávisle na těch ostatních s pravděpodobností jedna polovina. Pro i-té PraSátko si zavedeme jev A_i, který nastane, pokud žádný film z jeho seznamu oblíbených filmů nebyl vybrán. Dále jev B_i nastane, pokud všech deset filmů ze seznamu jeho neoblíbených filmů bylo vybráno. Díky nezávislosti jednotlivých výběrů je $\mathrm{P}(A_i) = \mathrm{P}(B_i) = 2^{-10}$. Máme tedy dohromady $250 \cdot 2 = 500$ jevů, kterým se chceme vyhnout. Použitím faktu, že pravděpodobnost sjednocení je nejvýše rovna součtu jednotlivých pravděpodobností, dostáváme

$$\mathrm{P}(A_1 \cup A_2 \cup \cdots \cup A_{250} \cup B_1 \cup B_2 \cup \cdots \cup B_{250}) \leq 500 \cdot 2^{-10} = \frac{500}{1024} < 1.$$

S nenulovou pravděpodobností se proto stane, že žádný z těchto jevů nenastane. To nutně znamená, že musí existovat seznam filmů, který splňuje dané podmínky. Tím jsme dokázali první část úlohy. Pro vyřešení druhé části si stačí uvědomit, že při naší náhodné volbě vybereme nejvýše 50 filmů s pravděpodobností vyšší než jedna polovina. To proto, že každému výběru více než 50 filmů odpovídá výběr méně než 50 filmů, které jsme v prvním výběru nevybrali. Označíme-li C jev, že vybereme více než 50 filmů, opět dostáváme

$$\mathrm{P}(A_1 \cup A_2 \cup \cdots \cup A_{250} \cup B_1 \cup B_2 \cup \cdots \cup B_{250} \cup C) \leq \frac{500}{1024} + \frac{1}{2} < 1.$$

ALTERNATIVNÍ ŘEŠENÍ:

Vyřešíme rovnou druhou podúlohu. Vybereme náhodnou podmnožinu právě 50 filmů; pracujeme tedy s pravděpodobnostním prostorem se $\binom{100}{50}$ prvky. Zavedeme stejné jevy A_i a B_i. Jev A_i nastane právě tehdy, když ze zbylých 90 filmů, jež nejsou na seznamu oblíbených filmů i-tého PraSátka, vybereme všech 50 filmů. To můžeme udělat $\binom{90}{50}$ možnostmi, takže $\mathrm{P}(A_i) = \binom{90}{50}/\binom{100}{50}$. Obdobně jev B_i nastane právě tehdy, když ze zbylých 90 filmů vybereme právě 40, a proto $\mathrm{P}(B_i) = \binom{90}{40}/\binom{100}{50}$. Nyní postupujeme stejně jako v předchozím řešení a spočítáme, že

$$\mathrm{P}(A_1 \cup \dots A_{250} \cup B_1 \cup \dots B_{250}) \leq 250 \cdot \frac{\frac{90!}{40! \cdot 50!} + \frac{90!}{50! \cdot 40!}}{\frac{100!}{50! \cdot 50!}} = 250 \cdot 2 \cdot \frac{90! \cdot 50! \cdot 50!}{100! \cdot 40! \cdot 50!} =$$
$$= 500 \cdot \frac{50 \cdot 49 \cdots 41}{100 \cdot 99 \cdots 91} = 500 \cdot \frac{50}{100} \cdot \frac{49}{99} \cdots \frac{41}{91} < 500 \cdot \left(\frac{1}{2}\right)^{10} = \frac{500}{1024} < 1.$$

Opět dostáváme vhodný seznam filmů s nenulovou pravděpodobností.

Poznámky:
Pěknou interpretaci prvního řešení měla *Magdaléna Mišinová*: rozdělíme filmy na dvě skupiny a stejně jako ve vzorovém řešení dokážeme, že s nenulovou pravděpodobností nebude žádná desetice celá v jedné skupině. Pak na filmové noci budeme pouštět tu menší skupinu filmů.

Většina řešitelů nicméně volila druhý postup (podobný poslední částí druhé úlohy ve studijním textu), který sice vede na složitější počítání, ale zato v něm není potřeba trik s interpretací podmínky na počet filmů jako dalšího jevu. *(Vašek Rozhoň)*

Úloha 2.

Noe vzal na archu n dvojic zvířat. Když po několika dnech začal být hladový, rozhodl se, že vybere k náhodných zvířat a sní je. Jaká je střední hodnota počtu různých druhů zvířat, která Noe vybral?[1]

(Vašek Rozhoň)

Řešení (volně podle Dominika Stejskala):
Druhy zvířat si očíslujeme čísly od jedné do n. Pak A_i pro i od 1 do n bude jev „i-tý druh byl ochutnán“. Spočítejme jeho pravděpodobnost.

Všech možných výběrů zvířat k sežrání je $\binom{2n}{k}$, pokud se chceme vyhnout jednomu konkrétnímu druhu, máme na výběr pouze z $2n-2$ zvířat, tedy k z nich k sežrání můžeme vybrat $\binom{2n-2}{k}$ způsoby. Pravděpodobnost, že daný druh byl ochutnán je pravděpodobnost, že jsme se mu nevyhnuli, tedy:

$$\mathrm{P}(A_i) = 1 - \frac{\binom{2n-2}{k}}{\binom{2n}{k}} = 1 - \frac{\frac{(2n-2)!}{k! \cdot (2n-k-2)!}}{\frac{(2n)!}{k! \cdot (2n-k)!}} = 1 - \frac{(2n-2)! \cdot (2n-k)!}{(2n)! \cdot (2n-k-2)!} =$$
$$= 1 - \frac{(2n-k) \cdot (2n-k-1)}{2n \cdot (2n-1)} = \frac{4n^2 - 2n - (4n^2 - 2nk - 2n - 2nk + k^2 + k)}{n \cdot (4n-2)} =$$
$$= \frac{4nk - k^2 - k}{n \cdot (4n-2)},$$

což se současně rovná střední hodnotě příslušného indikátoru.

Nyní již stačí využít linearity střední hodnoty, protože hledaná náhodná veličina je součtem indikátorů přes všechny druhy, tedy její střední hodnota je součtem příslušných středních hodnot, a proto je rovna

$$n \cdot \frac{4nk - k^2 - k}{n \cdot (4n-2)} = \frac{4nk - k^2 - k}{4n - 2}.$$

[1] Pokud například Noe vybral lva, lvici a jednorožce, jedná se o dva různé druhy.

Poznámky:

Řešitelé, kteří alespoň něco poslali, většinou úlohu vyřešili, avšak nepočítali $\mathrm{P}(A_i)$ pomocí pravděpodobnosti doplňku. Pak bylo potřeba zvlášť rozebrat situaci, kdy Noe snědl pouze jedno zvíře daného druhu, a kdy snědl obě, čímž přibylo trochu úprav navíc. *Dominik Stejskal* a *Kačka Panešová*, kteří postupovali podle vzoráku, si vysloužili jeden kladný imaginární bod.

Naproti tomu někteří řešitelé počítali střední hodnotu přímočaře bez použití linearity střední hodnoty. Těm pak vyšla střední hodnota jako

$$\frac{1}{\binom{2n}{k}} \sum_{i=\lceil \frac{k}{2} \rceil}^{k} i \cdot \binom{n}{i} \cdot \binom{i}{k-i} \cdot 2^{2i-k}.$$

To je sice správná hodnota, ale výsledek vždy chceme dostat v co nejjednodušším možném tvaru – jistě uznáte, že podíl dvou polynomů stupně jedna a dva dá mnohem lepší představu o tom, kolik to tedy je pro konkrétní k a n, než takováhle suma. Taková řešení byla tedy hodnocena dvěma body.

(Viki Němeček)

Úloha 3.

Na Matfyzu se sešlo n informatiků a matematiků, přičemž každý matematik zná alespoň jednoho informatika. Ukažte, že můžeme vybrat takovou skupinu matfyzáků o velikosti alespoň $n/2$, aby uvnitř ní každý matematik znal lichý počet informatiků.

(Danil Koževnikov)

Řešení:

V řešení použijeme kombinaci pravděpodobnostní metody a linearity střední hodnoty. Budeme náhodně volit skupinu, kde každý matematik zná lichý počet informatiků, a ukážeme, že střední hodnota její velikosti je rovna $n/2$, což zaručí existenci alespoň jedné skupiny s alespoň polovičním počtem matfyzáků.

Označme si počty matematiků a informatiků na Matfyzu postupně m, i. Zavedeme si n náhodných veličin $X_1, \ldots, X_m$ a $Y_1, \ldots, Y_i$, které pro každého matematika, respektive informatika, indikují, zda je ve vybrané skupině. Celkový počet lidí ve skupině je potom náhodná veličina $S = \sum_{j=1}^{m} X_j + \sum_{j=1}^{i} Y_j$.

Abychom si co nejvíce usnadnili výpočet $\mathrm{E}(S)$, zvolíme naši skupinu specifickým způsobem: nejprve do ní přidámě každého informatika s pravděpodobností $\frac{1}{2}$ a poté tam doplníme všechny matematiky, kteří znají lichý počet zvolených informatiků. V tomto případě zjevně platí $\mathrm{E}(Y_j) = \frac{1}{2}$ pro j od 1 do i.

Řekněme, že daný matematik zná $k \geq 1$ informatiků, které jsou nějakým (jakýmkoliv) způsobem seřazeni. Označme si jako A jev to, že byl zvolen sudý počet z prvních $k-1$ známých informatiků, a jako B jev to, že byl zvolen k-tý z nich. Potom pravděpodobnost, že zná lichý počet informatiků ze zvolené skupiny, můžeme vyjádřit jako $P(A \cap B) + P(\overline{A} \cap \overline{B}) = \frac{1}{2}(P(A) + P(\overline{A})) = \frac{1}{2}$. Využili jsme toho, že volba posledního známého je nezávislá na všech ostatních, a rovněž toho, že $k-1 \geq 0$, aby byly pravděpodobnosti z posledního kroku dobře definované. Takže každý matematik bude ve zvolené skupině rovněž s pravděpodobností jedna polovina, pročež $\mathrm{E}(X_j) = \frac{1}{2}$ a z linearity střední hodnoty $\mathrm{E}(S) = \frac{m}{2} + \frac{i}{2} = \frac{n}{2}$, což jsme chtěli dokázat.

Poznámky:

Všechna správná řešení se myšlenkou volby největší možné skupiny z náhodné skupiny informatiků podobala tomu vzorovému. Místo použití pravděpodobnostních pojmů jste však často používali ne vždy zcela jasné argumenty se symetrií nebo naopak přímočařejší přístup s počítáním střední hodnoty z definice přes sumu. Ačkoliv oba přístupy u této úlohy byly celkem schůdné (například poslední část řešení se dala nahradit trikovým použitím binomické věty), tak obecně spíš platí, že když už řešení používá pravděpodobnostní metodu, pak půjde elegantně zformulovat v jazyce pravděpodobnosti.

(Danil Koževnikov)