1. seriálová série

Téma: Pravděpodobnost a matematická statistika

Termín odeslání: 5. ledna 2004

1. úloha (5 bodů)

Nalezněte nebo dokažte, že neexistují:

i) Jevy A a B takové, že $\mathrm{P}(A \cup B) > \mathrm{P}(A) + \mathrm{P}(B)$.

ii) Jevy A, B, C takové, že A není nezávislý s B, B není nezávislý s C, ale A je nezávislý s C.

iii) Jevy A, B, C takové, že A je nezávislý s B, B je nezávislý s C, ale A není nezávislý s C.

iv) Jevy A, B, C takové, že A je nezávislý s B, B je nezávislý s C, A je nezávislý s C, ale

$$\mathrm{P}(A \cap B \cap C) \neq \mathrm{P}(A)\,\mathrm{P}(B)\,\mathrm{P}(C).$$

v) Jev A nezávislý sám se sebou.

vi) Nezávislé jevy A a B takové, že A^c a B^c nejsou nezávislé jevy.

vii) Jevy A, B a rozklady $C_1, \ldots, C_m$ a $D_1, \ldots, D_m$ takové, že

$$\mathrm{P}(A|C_i) < \mathrm{P}(B|D_i), \qquad i = 1, \ldots, m,$$

ale zároveň $\mathrm{P}(A) > \mathrm{P}(B)$.

2. úloha (5 bodů)

Házíme n kostkami. Označme $p(n, a, b, c)$ pravděpodobnost, že padne právě a šestek, b pětek a c čtyřek. Nechť a, b, c jsou libovolná daná přirozená čísla tak, že $a + b + c = 6$. Nalezněte n takové, že $p(n, a, b, c)$ je maximální možné.

3. úloha (5 bodů)

Při souboji se tři soupeři Alexandr, Bořivoj a Celestýn postaví tak, aby každý mohl střílet na každého. Střílí se v pořadí podle abecedy[1], dokud na kolbišti nezůstane jediný živý vítěz. Alexandr se trefí na 50%, Bořivoj na 75% a Celestýn vždy. Každý ze soupeřů použije nejvýhodnější strategii (ale jinak to jsou gentlemani a nebudou schválně špatně mířit, domlouvat se, ani jinak podvádět). Jakou mají jednotliví účastníci souboje naději na přežití?
A jak se situace změní, když Alexandr napoprvé (v rozporu s pravidly) vystřelí do vzduchu?
A jak tomu bude, pokud se souboje zúčastní i nepříliš dobrý střelec Dobromysl, který zasáhne cíl na 25%?

[1] Tedy každý střelec, když na něj přijde řada, si vybere některého ze soků a pokusí se ho zastřelit. Poté čeká, zda se dožije toho, až na něj zase přijde řada.

2. seriálová série

Téma:	Pravděpodobnost a matematická statistika
Termín odeslání:	8. BŘEZNA 2004

4. ÚLOHA (5 BODŮ)

Znáte hru Šťastných deset? Ne? Ale to byste měli, je o ní totiž tento příklad. Určete střední hodnotu a rozptyl výhry člověka, který vsadí deset korun na tip pěti čísel a nehraje Šanci milión. Víme, že hru Šťastných deset hraje v ČR každý týden přes milión hráčů (zdroj Sazka, `http://www.sazka.cz`). Jaká by byla střední hodnota tržby Sazky na této hře, když budeme počítat s pouhým miliónem hráčů sázejících tak, jak je výše popsáno? Dokážete spočítat i rozptyl této tržby?

5. ÚLOHA (5 BODŮ)

Tato úloha bude mít několik samostatných částí (první je za jeden bod, druhá a třetí za dva body):
i) Vyplývá z toho, že A a B jsou podmíněně nezávislé dáno C, i to, že A a B jsou podmíněně nezávislé dáno C^c?
ii) Dokažte, že náhodná veličina X nabývající hodnot 0, 1, 2, ... má geometrické rozdělení (s nějakým parametrem p) právě tehdy, když pro libovolná x, y celá nezáporná čísla platí

$$\mathrm{P}(X = x + y) = \mathrm{P}(X \geq x)\,\mathrm{P}(X = y).$$

iii) Dokažte vzorec pro střední hodnotu náhodné veličiny s geometrickým rozdělením. Prosím, nepoužívejte derivace, ani integrály (jde to i bez nich, pouze s použitím vzorce $\sum_{i=0}^{\infty} q^i = \frac{1}{1-q}$ pro libovolné q mezi nulou a jedničkou).

6. ÚLOHA (6 BODŮ)

Permutací čísel 1, 2, ... , n rozumíme posloupnost těchto čísel napsanou v libovolném pořadí. Například všechny permutace čísel 1, 2, 3 jsou $(1, 2, 3)$, $(1, 3, 2)$, $(2, 1, 3)$, $(2, 3, 1)$, $(3, 1, 2)$ a $(3, 2, 1)$. Počátečním rostoucím úsekem permutace $(q_1, q_2, \dots, q_n)$ myslíme prvních k čísel této permutace $q_1, \dots, q_k$ takových, že $q_1 < q_2 < \dots < q_k$ a buďto $k = n$ nebo $q_k > q_{k+1}$. Tedy kupříkladu počáteční rostoucí úseky výše uvedených permutací jsou po řadě $(1, 2, 3)$, $(1, 3)$, (2), $(2, 3)$, (3) a (3). Délkou počátečního rostoucího úseku míníme počet jeho prvků.
Mějme náhodnou permutaci čísel 1, 2, ... , n, tj. každá permutace nastává se stejnou pravděpodobností (a tato pravděpodobnost je $\frac{1}{n!}$). Jaká je střední hodnota délky počátečního rostoucího úseku této náhodné permutace?

3. seriálová série

Téma: Pravděpodobnost a matematická statistika

Termín odeslání: 17. května 2004

7. úloha (5 bodů)

Dokažte, že v textu použité odhady střední hodnoty a rozptylu jsou správné, neboli ověřte, že skutečně platí:

i) $\mathrm{E}\,\overline{X}_n = \mathrm{E}\,X$,
ii) $\mathrm{Var}\,\overline{X}_n = \frac{1}{n}\,\mathrm{Var}\,X$,
iii) $\mathrm{E}\,S^2 = \mathrm{Var}\,X$.

8. úloha (5 bodů)

Král Kulatého království měl ukrutně dlouhý nos a poddaní se mu kvůli němu smáli a nebrali ho vážně. Králův rádce králi poradil, že když nechá děti svých poddaných místo jablek jíst frgálovník[2], naroste i jim delší nos a za pár let už se mu nikdo smát nebude.

Král byl statisticky vzdělán, a tak se rozhodl před masovou výsadbou frgálovníků po celém království rádcovu radu nejprve ověřit. Nejprve nahlédl do knihovny a zjistil, že rozptyl délky nosu pětiletých dětí byl mnoha lety praxe určen jako 0,5 cm^2. Poté vybral dvě skupiny malých dětí, každou o 100 jedincích. Děti z první skupiny dostávaly klasická jablíčka, děti z druhé skupiny jedly místo nich frgálovníky[3]. V pěti letech byla změřena délka jejich nosu. Zatímco v první skupině byl průměr délky nosů roven 3,5 cm, v druhé byl o něco vyšší, totiž 3,9 cm.

Skeptikové však tvrdili, že se jedná o náhodu a že frgálovník není účinný. Co jim král na to odpověděl?

9. úloha (5 bodů)

Nechť X je náhodná veličina nabývající hodnot 1 a -1, obou s pravděpodobností $\frac{1}{2}$. Precizně ověřte, že pro její rozdělení platí námi předpokládaný (ZVČ). Tedy dokažte, že pro libovolně malé $\epsilon > 0$ platí, že pokud vezmeme n dostatečně vysoké (tedy větší nebo rovno nějaké mezi n_ϵ), tak potom

$$\mathrm{P}(|\overline{X}_n| \geq \epsilon) \leq \epsilon.$$

[2] Frgálovník je ovoce k nerozeznání podobné jablkům až na drobné vedlejší účinky při konzumaci.
[3] Král byl i předsedou a jediným členem Etické komise Kulatého království.

Řešení seriálové série

1. úloha
Nalezněte nebo dokažte, že neexistují:
i) Jevy A a B takové, že $\mathrm{P}(A \cup B) > \mathrm{P}(A) + \mathrm{P}(B)$.
ii) Jevy A, B, C takové, že A není nezávislý s B, B není nezávislý s C, ale A je nezávislý s C.
iii) Jevy A, B, C takové, že A je nezávislý s B, B je nezávislý s C, ale A není nezávislý s C.
iv) Jevy A, B, C takové, že A je nezávislý s B, B je nezávislý s C, A je nezávislý s C, ale

$$\mathrm{P}(A \cap B \cap C) \neq \mathrm{P}(A)\,\mathrm{P}(B)\,\mathrm{P}(C).$$

v) Jev A nezávislý sám se sebou.
vi) Nezávislé jevy A a B takové, že A^c a B^c nejsou nezávislé jevy.
vii) Jevy A, B a rozklady $C_1, \ldots, C_m$ a $D_1, \ldots, D_m$ takové, že

$$\mathrm{P}(A|C_i) < \mathrm{P}(B|D_i), \qquad i = 1, \ldots, m,$$

ale zároveň $\mathrm{P}(A) > \mathrm{P}(B)$.

i) Takové jevy neexistují, neboť zjevně $|A \cup B| \leq |A| + |B|$, tudíž nutně

$$\mathrm{P}(A \cup B) = \frac{|A \cup B|}{|\Omega|} \leq \frac{|A|}{|\Omega|} + \frac{|B|}{|\Omega|} = \mathrm{P}(A) + \mathrm{P}(B).$$

ii) Kupříkladu uvažme libovolné nezávislé jevy A a C s pravděpodobností větší než nula a menší než jedna a položme $B = A \cap C$.
iii) Zde je řešení ještě jednodušší. Nechť A a B jsou nezávislé jevy a pravděpodobnost jevu A je větší než nula a menší než jedna. Pak stačí položit $C = A$.
iv) Položme $\Omega = \{1, 2, 3, 4\}$, $A = \{1, 4\}$, $B = \{2, 4\}$, $C = \{3, 4\}$. Potom $\mathrm{P}(A) = \mathrm{P}(B) = \mathrm{P}(C) = \frac{1}{2}$, $\mathrm{P}(A \cap B) = \mathrm{P}(B \cap C) = \mathrm{P}(C \cap A) = \frac{1}{4}$, jevy A, B, C jsou tedy po dvou nezávislé, ovšem

$$P(A \cap B \cap C) = \frac{1}{4} \neq \frac{1}{8} = \mathrm{P}(A)\,\mathrm{P}(B)\,\mathrm{P}(C).$$

v) Tuto vlastnost má libovolný jistý či nemožný jev.
vi) Takové jevy neexistují, neboť pokud A a B jsou nezávislé, potom

$$\mathrm{P}(A \cap B^c) = \mathrm{P}(A) - \mathrm{P}(A \cap B) = \mathrm{P}(A) - \mathrm{P}(A)\,\mathrm{P}(B) = \mathrm{P}(A)\,\mathrm{P}(B^c),$$

což znamená, že nutně A a B^c jsou nezávislé jevy, a zopakováním tohoto postupu dokážeme, že pokud A a B^c jsou nezávislé jevy, pak A^c a B^c jsou nezávislé jevy.
vii) Jedno z mnoha možných řešení je toto: Položme $\Omega = \{1, 2, 3, 4, 5\}$, $m = 2$, $C_1 = D_2 = \{1\}$, $C_2 = D_1 = \{2, 3, 4, 5\}$, $A = \{2, 3, 4\}$, $B = \{1, 5\}$.

Poznámky opravovatele: Ad *ii), iii), iv), vii)* U některých řešení se mi nelíbilo, že řešitelé napsali příklady jevů, ale nespočítali pravděpodobnosti.

Ad *iv)* Několik řešitelů napsalo

$$P((A \cap B) \cap C) = P(A \cap B) \cdot P(C) = P(A) \cdot P(B) \cdot P(C)$$

a považovalo to za správné řešení. Neuvědomili si, že potřebují nezávislost jevů $A \cap B$ a C. Další to považovali za samozřejmé z nezávislosti dvojic jevů A, C a B, C. Bohužel toto neplatí vždy. Mějme např.

$$\Omega = \{1, 2, 3, 4\}, \quad A = \{1, 2\}, \quad B = \{2, 3\}, \quad C = \{1, 3\}.$$

Pak $P(A) = P(B) = P(C) = \frac{1}{2}$, $P(A \cap B) = P(A \cap C) = P(B \cap C) = \frac{1}{4}$, ale $P((A \cap B) \cap C) = 0 \neq \frac{1}{8} = P(A \cap B) \cdot P(C)$. Jevy $A \cap B$ a C nejsou nezávislé.

Ad *v)* Někteří řešitelé z rovnosti $P(A) = P(A)^2$ odvodili, že $P(A)$ musí být 1, a zapomněli na možnost $P(A) = 0$. To je klasická chyba.

Nemálo řešení vypadalo tak, že řešitel napsal zadání, zamlžil je a výsledek prohlásil za zřejmý. Někdy se trefil a někdy ne. Fakt, že se náhodou trefil do správné odpovědi, ale není důvod k tomu, abych toto řešení uznal, i kdyby mlha vzdáleně připomínala autorské řešení. Např. „Jsou-li jevy A a B nezávislé, pak nastání či nenastání jevu A neovlivní pravděpodobnost jevu B, podobně pro nezávislé jevy B a C. Proto nastání či nenastání jevu A neovlivní pravděpodobnost jevu B, což neovlivní jev C. Proto jevy A a C musí být nezávislé. Jevy A, B a C vyhovující zadání neexistují.“ K takovýmto mlžícím řešením nedokáži napsat více než „Toto není důkaz.“, „Ale existují.“ a podobně.

2. úloha

Házíme n kostkami. Označme $p(n, a, b, c)$ pravděpodobnost, že padne právě a šestek, b pětek a c čtyřek. Nechť a, b, c jsou libovolná daná přirozená čísla tak, že $a + b + c = 6$. Nalezněte n takové, že $p(n, a, b, c)$ je maximální možné.

Pravděpodobnost $p(n, a, b, c)$ získáme tak, že v n hodech vybereme místa na a šestek, b pětek a c čtyřek, přičemž každé toto číslo na daném místě padne s pravděpodobností $\frac{1}{6}$, a zbylá místa budou zaplněná nižšími čísly, každé s pravděpodobností $\frac{3}{6}$. Tuto pravděpodobnost samozřejmě pronásobíme počtem možných výběrů míst a dostáváme, že

$$p(n, a, b, c) = \frac{n!}{a!b!c!(n-6)!} \left(\frac{1}{6}\right)^6 \left(\frac{3}{6}\right)^{(n-6)},$$

což není výraz, u kterého by bylo na první pohled jasné, kde nabývá maxima.

Zavedeme tedy

$$r(n, a, b, c) = \frac{p(n+1, a, b, c)}{p(n, a, b, c)} = \frac{n+1}{n-5} \cdot \frac{1}{2}.$$

Platí, že $p(n+1, a, b, c) \geq p(n, a, b, c)$ právě tehdy, když $r(n, a, b, c) \geq 1$. Zbývá nám tedy rozhodnout, kdy je $r(n, a, b, c) \geq 1$, ale to je již pouhé vyřešení jedné nerovnice. Dostáváme, že $p(n, a, b, c)$ pro (libovolné) pevné hodnoty a, b a c nabývá maxima pro $n = 11$ a $n = 12$ ($r(11, a, b, c) = 1$).

Poznámky opravovatele: Objevilo se překvapivě mnoho chybných řešení. Při počítání pravděpodobnosti, že na n kostkách padne a šestek, b pětek a c čtyřek pro pevně zvolená a, b a c, se řada řešitelů dostala na scestí. Často si neuvědomili, že pokud při počítání možných konfigurací, jaká čísla mohla na n kostkách padnout, nahlížíme na kostky jako na očíslované, dostaneme jiný výsledek než u kostek neočíslovaných (tj. když si je například seřadíme podle hodnot hozených čísel). Další častou chybou byla úvaha typu: je-li $P(A) = p_1$ a $P(B) = p_2$, pak vždy $P(A \cap B) = p_1 p_2$. Pozor, to platí pouze pro nezávislé jevy! Víme, že $P(\text{na kostce padne jednička}) = \frac{1}{6}$, $P(\text{na kostce padne dvojka}) = \frac{1}{6}$, ale $P(\text{na kostce padne zároveň jednička i dvojka}) \neq \frac{1}{36}$!

3. úloha

Při souboji se tři soupeři Alexandr, Bořivoj a Celestýn postaví tak, aby každý mohl střílet na každého. Střílí se v pořadí podle abecedy[4], dokud na kolbišti nezůstane jediný živý vítěz. Alexandr se trefí na 50%, Bořivoj na 75% a Celestýn vždy. Každý ze soupeřů použije nejvýhodnější strategii (ale jinak to jsou gentlemani a nebudou schválně špatně mířit, domlouvat se, ani jinak podvádět). Jakou mají jednotliví účastníci souboje naději na přežití?

A jak se situace změní, když Alexandr napoprvé (v rozporu s pravidly) vystřelí do vzduchu?

A jak tomu bude, pokud se souboje zúčastní i nepříliš dobrý střelec Dobromysl, který zasáhne cíl na 25%?

Nejprve uvažme souboj dvou soků, z nichž ten, co střílí první, zasáhne protivníka s pravděpodobností p, a ten, co střílí druhý, s pravděpodobností q. Pravděpodobnost, že v tomto souboji zvítězí první, budeme značit $f_1(p, q)$, pravděpodobnost vítězství druhého $f_2(p, q)$.

Nejprve střílí první a s pravděpodobností p je souboj ukončen a první vítězí, jinak střílí druhý. Pokud zasáhne, zvítězil, jinak se vše opakuje. Tedy šance na vítězství je po dvou neúspěšných výstřelech stejná, jako byla na začátku. Platí tedy

$$f_1(p, q) = p + (1 - p)(1 - q) f_1(p, q).$$

Odtud snadnou úpravou dostáváme

$$f_1(p, q) = \frac{p}{p + q - pq}.$$

A obdobně nebo s využitím $f_1(p, q) + f_2(p, q) = 1$ odvodíme

$$f_2(p, q) = \frac{q - pq}{p + q - pq}.$$

Nyní uvažme, že nám do souboje přibude třetí gentleman. Snadným pozorováním je, že se každému zápasníkovi vyplatí střílet po lepším ze zbývajících dvou soků[5]. V našem případě tedy

[4] Tedy každý střelec, když na něj přijde řada, si vybere některého ze soků a pokusí se ho zastřelit. Poté čeká, zda se dožije toho, až na něj zase přijde řada.

[5] Pozor! U souboje čtyř už to platit nemusí! Viz níže.

Alexandr (dále A) i Bořivoj (dále B) střílí na začátku po Celestýnovi (dále C) a Celestýn po Bořivoji. Označme pravděpodobnosti zásahu jednotlivých střelců podle abecedy $a = 0{,}5$, $b = 0{,}75$, $c = 1$ a pro pozdější použití (pro Dobromysla) $d = 0{,}25$.

Tudíž C přežije, pokud se do něj A ani B netrefí. Poté zastřelí B a pakliže se do něj A znovu netrefí, zastřelí i jeho a zvítězí. Zapsáno vzorcem:

$$\mathrm{P}(\text{C zvítězí}) = (1-a)(1-b)(1-a) \doteq 0{,}06.$$

Pravděpodobnost, že zvítězí B, si rozepíšeme podle toho, zda A zasáhl při prvním výstřelu C. Pokud ano, pak dochází k souboji B a A (B střílí první), tudíž B vítězí s pravděpodobností $f_1(b,a)$. V druhém případě buďto B zastřelí C a dochází k souboji A a B (A střílí první), nebo následně umírá zastřelen C. Tudíž

$$\mathrm{P}(\text{B zvítězí}) = \mathrm{P}(\text{C na začátku zastřelen})\,\mathrm{P}(\text{B zvítězí}|\text{C na začátku zastřelen}) +$$

$$\mathrm{P}(\text{C na začátku není zastřelen})\,\mathrm{P}(\text{B zvítězí}|\text{C na začátku není zastřelen}) =$$

$$= af_1(b,a) + (1-a)bf_2(a,b) \doteq 0{,}59.$$

A protože vždy zvítězí právě jeden, dostáváme

$$\mathrm{P}(\text{A zvítězí}) = 1 - \mathrm{P}(\text{B zvítězí}) - \mathrm{P}(\text{C zvítězí}) \doteq 0{,}35.$$

Tím je první část příkladu vyřešena.

Druhá část je pouhou obdobou první, proto zde uvedu pouze výsledky, komentář si zajisté domyslíte sami:

$$\mathrm{P}(\text{C zvítězí}) = (1-b)(1-a) \doteq 0{,}13,$$
$$\mathrm{P}(\text{B zvítězí}) = bf_2(a,b) \doteq 0{,}32,$$
$$\mathrm{P}(\text{A zvítězí}) = 1 - \mathrm{P}(\text{B zvítězí}) - \mathrm{P}(\text{C zvítězí}) \doteq 0{,}55.$$

Povšimněme si, že výstřel do vzduchu (t.j. neúspěšný výstřel) zvýšil šanci A na vítězství.

Třetí část příkladu je o něco obtížnější v tom, že v případě čtyř soupeřů není tak úplně jasné, kdo má na koho střílet. Je zjevné, že C (pokud na něj dojde řada a jsou ve hře ještě čtyři) by měl střílet po nejsilnějším ze zbývajících, t.j. po B, neboť v souboji tří budou jistě všichni střílet po něm, a tudíž čím slabší soupeři, tím větší šance na přežití. Proto by i B (při souboji čtyř) měl střílet po C, jinak ho ten bez ohledu na výsledek zastřelí.

Po kom by však měl střílet A? Odpověď na tuto otázku rozhodně není triviální! Pravděpodobnost vítězství A za předpokladu, že střílí na počátku po X, je rovna

$$\mathrm{P}(\text{A zvítězí}) = a\,\mathrm{P}(\text{A zvítězí}|\text{X na začátku zastřelen}) +$$

$$+ (1-a)\,\mathrm{P}(\text{A zvítězí}|\text{X na začátku není zastřelen}).$$

Povšimněme si, že člen $\mathrm{P}(\text{A zvítězí}|\text{X na začátku není zastřelen})$ nezávisí na X, stačí tedy maximalizovat $\mathrm{P}(\text{A zvítězí}|\text{X na začátku zastřelen})$ přes všechna možná X.

Pokud položíme X=B, potom P(A zvítězí|B na začátku zastřelen) = 0, neboť A je za tohoto předpokladu zastřelen C.

Pokud položíme X=D, potom dostáváme úlohu obdobnou prvním dvěma částem našeho příkladu. Daná pravděpodobnost je zjevně stejná jako v druhé části příkladu, kdy A střílel do vzduchu, neboť ve hře v obou případech ve hře zůstávají B, C, A (první střílí B):

$$\mathrm{P}(\text{A zvítězí}|\text{D na začátku zastřelen}) \doteq 0{,}55.$$

Pokud položíme X=C, je všechno složitější. Ve hře zůstává B, D, A, první střílí B na A, pak D na B a poté A na B. Pokud se někdo z nich strefí, pokračujeme soubojem nějaké dvojice. Pokud ne, vracíme se po třech neúspěšných výstřelech do výchozí situace a šance A na přežití je tudíž stejná jako na začátku, odtud

$$\mathrm{P}(\text{A zvítězí}|\text{C na začátku zastřelen}) = (1-b)df_1(a,b) + (1-b)(1-d)af_2(d,a) +$$

$$+(1-b)(1-d)(1-a)\,\mathrm{P}(\text{A zvítězí}|\text{C na začátku zastřelen}).$$

Odtud snadnou úpravou

$$\mathrm{P}(\text{A zvítězí}|\text{C na začátku zastřelen}) = \frac{(1-b)df_1(a,b) + (1-b)(1-d)af_2(d,a)}{1-(1-b)(1-d)(1-a)} \doteq 0{,}10.$$

Můžeme tedy uzavřít – A bude střílet po D.

Zbytek příkladu je spíše pracný nežli obtížný. Úlohu postupně rozepíšeme podle toho, zda A a B byli při svých prvních výstřelech úspěšní. Poté nám již ve hře zůstávají pouze tři střelci. Jediný obtížnější případ (pokud zůstanou D, A a B) vyřešíme analogicky jako výše (doporučuji si namalovat obrázek – vývojový diagram).

$$\mathrm{P}(\text{A zvítězí}) = abf_1(a,b) + a(1-b)a + (1-a)b\cdot\frac{df_1(a,d)+(1-d)af_2(d,a)}{1-(1-a)(1-b)(1-d)} +$$

$$+(1-a)(1-b)[df_1(a,d)+(1-d)af_2(d,a)] \doteq 0{,}51,$$

$$\mathrm{P}(\text{B zvítězí}) = abf_2(a,b) + (1-a)b\cdot\frac{(1-d)(1-a)bf_2(d,b)}{1-(1-a)(1-b)(1-d)} \doteq 0{,}24,$$

$$\mathrm{P}(\text{C zvítězí}) = a(1-b)(1-a) + (1-a)(1-b)(1-d)(1-a)(1-d) \doteq 0{,}10,$$

$$\mathrm{P}(\text{D zvítězí}) = (1-a)b\cdot\frac{df_2(a,d)+(1-d)af_1(d,a)+(1-d)(1-a)bf_1(d,b)}{1-(1-a)(1-b)(1-d)} +$$

$$+(1-a)(1-b)[df_2(a,d)+(1-d)af_1(d,a)+(1-d)(1-a)d] \doteq 0{,}16.$$

Poznámky opravovatele: S první částí většina řešitelů problémy neměla. S druhou, jež byla ve skutečnosti nápovědou ke třetí, teprve ne. Potíže nastaly s částí poslední, kdy někteří neodhalili, že Alexandrovi se vyplatí střílet spíš na Dobromysla než na Celestýna (neboť když zabije Celestýna, začnou všichni střílet po něm). Některým řešitelům také činilo obtíže upočítat souboj Alexandra, Bořivoje a Dobromysla.

4. úloha

Znáte hru Šťastných deset? Ne? Ale to byste měli, je o ní totiž tento příklad. Určete střední hodnotu a rozptyl výhry člověka, který vsadí deset korun na tip pěti čísel a nehraje Šanci milión. Víme, že hru Šťastných deset hraje v ČR každý týden přes milión hráčů (zdroj Sazka, `http://www.sazka.cz`). Jaká by byla střední hodnota tržby Sazky na této hře, když budeme počítat s pouhým miliónem hráčů sázejících tak, jak je výše popsáno? Dokážete spočítat i rozptyl této tržby?

Pravděpodobnost, že vylosují právě i námi tipovaných čísel, je rovna

$$\frac{\binom{5}{i}\binom{75}{20-i}}{\binom{80}{20}},$$

neboť počet všech možností je dán počtem všech výběrů 20 čísel z 80 a i uhodnutých čísel dosáhneme, pokud vybereme i z pěti čísel a $20-i$ ze zbylých 75 čísel.

Střední hodnota výhry v korunách je tedy rovna

$$10\cdot\left[200\cdot\frac{\binom{5}{5}\binom{75}{15}}{\binom{80}{20}}+16\cdot\frac{\binom{5}{4}\binom{75}{16}}{\binom{80}{20}}+2\cdot\frac{\binom{5}{3}\binom{75}{17}}{\binom{80}{20}}\right]\doteq 4{,}9.$$

Analogicky spočítáme střední hodnotu druhé mocniny výhry a pomocí ní z definice rozptyl, jenž se přibližně rovná 2897. Jelikož tržba Sazky je rovna 10 Kč mínus výhra hráče, dostáváme, že střední tržba na daném hráči je 5,1 a rozptyl je tentýž.

Víme, že střední hodnota součtu náhodných veličin je součet jejich středních hodnot. Tedy střední hodnota tržby na miliónu (stejných) hráčů je rovna miliónkrát střední hodnota tržby na jednom hráči, tedy zhruba 5096675 Kč.

A jak to bude s rozptylem? Tady je třeba trochu popřemýšlet. Kdyby sázky hráčů byly nezávislé, jednoduše bychom rozptyly sečetli. Je tomu ale opravdu tak? Spíše ne, lidé sázejí častěji na známá čísla jako 3, 7 nebo 13. Rozptyl tržby tedy nemůžeme bez dalších informací určit.

Poznámky opravovatele: Úloha nebyla obtížná a v došlých řešeních mnoho chyb (až na ty numerické) nebylo. Jediná obtíž spočívala v poslední části, kdy se mělo rozhodnout, zda lze spočítat rozptyl tržby Sazky.

Potíž je v tom, že není jasné, zda lze brát čísla sázená jednotlivými hráči (a tudíž i výhry jim vyplácené) jako nezávislé veličiny a zda tedy můžeme použít, že rozptyl součtu je roven součtu rozptylů. Dle mého názoru ne (šťastná čísla jako 3, 5, 7, 13 se mezi sázkami určitě vyskytnou častěji), každopádně se slušelo provést diskuzi. To udělal pouze jediný řešitel, ostatní to stálo jeden bod.

5. úloha

Tato úloha bude mít několik samostatných částí (první je za jeden bod, druhá a třetí za dva body):

i) Vyplývá z toho, že A a B jsou podmíněně nezávislé dáno C, i to, že A a B jsou podmíněně nezávislé dáno C^c?

ii) Dokažte, že náhodná veličina X nabývající hodnot 0, 1, 2, ... má geometrické rozdělení (s nějakým parametrem p) právě tehdy, když pro libovolná x, y celá nezáporná čísla platí

$$\mathrm{P}(X = x + y) = \mathrm{P}(X \geq x)\,\mathrm{P}(X = y).$$

iii) Dokažte vzorec pro střední hodnotu náhodné veličiny s geometrickým rozdělením. Prosím, nepoužívejte derivace, ani integrály (jde to i bez nich, pouze s použitím vzorce $\sum_{i=0}^{\infty} q^i = \frac{1}{1-q}$ pro libovolné q mezi nulou a jedničkou).

i) Evidentně nevyplývá. Protipříkladem je třeba $\Omega = \{1, 2, 3\}$, $A = B = \{1, 2\}$, $C = \{1\}$.

ii) Pakliže $X \sim \mathit{Geom}(p)$, platí

$$\mathrm{P}(X = x + y) = p(1-p)^{x+y} = (1-p)^x \cdot p(1-p)^y = \mathrm{P}(X \geq x) \cdot \mathrm{P}(X = y).$$

Opačnou implikaci dokážeme indukcí. Označme $p = \mathrm{P}(X = 0)$. Nechť $0 < p < 1$, potom ukážeme, že $\mathrm{P}(X = k) = p(1-p)^k$: Pro $k = 0$ toto tvrzení zjevně platí. A pokud platí pro všechna k od 0 do $n-1$, pak dostáváme za využití indukčního předpokladu

$$\mathrm{P}(X = n + 0) = \mathrm{P}(X \geq n)\,\mathrm{P}(X = 0) = (1 - \mathrm{P}(X < n))\,\mathrm{P}(X = 0) =$$

$$= p\left(1 - \sum_{i=0}^{n-1} p(1-p)^i\right) = p\left(1 - p\frac{1-(1-p)^n}{p}\right) = p(1-p)^n,$$

což je přesně to, co jsme chtěli.

Zbývá dořešit okrajové případy: Pro $p = 0$ dostáváme, že $P(X = x) = 0$ pro každé x, tedy spor. Pro $p = 1$ dostáváme $X \sim \mathit{Geom}(1)$.

iii) Uvažujme $X \sim \mathit{Geom}(p)$, označme $q = 1 - p$ a upravujme:

$$\mathrm{E}\,X = \sum_{k=0}^{\infty} kpq^k = p\sum_{k=1}^{\infty} kq^k = p\sum_{k=1}^{\infty}\sum_{l=k}^{\infty} q^l = p\sum_{k=1}^{\infty} q^k \sum_{m=0}^{\infty} q^m =$$

$$= p\sum_{k=1}^{\infty} q^k \frac{1}{1-q} = \sum_{k=1}^{\infty} q^k = \frac{1}{p} - 1 = \frac{1-p}{p}.$$

6. úloha

Permutací čísel 1, 2, ... , n rozumíme posloupnost těchto čísel napsanou v libovolném pořadí. Například všechny permutace čísel 1, 2, 3 jsou $(1, 2, 3)$, $(1, 3, 2)$, $(2, 1, 3)$, $(2, 3, 1)$, $(3, 1, 2)$ a $(3, 2, 1)$. Počátečním rostoucím úsekem permutace $(q_1, q_2, \ldots, q_n)$ myslíme prvních k čísel této permutace $q_1, \ldots, q_k$ takových, že $q_1 < q_2 < \cdots < q_k$ a buďto $k = n$ nebo $q_k > q_{k+1}$. Tedy kupříkladu počáteční rostoucí úseky výše uvedených permutací jsou po řadě $(1, 2, 3)$, $(1, 3)$, (2), $(2, 3)$, (3) a (3). Délkou počátečního rostoucího úseku míníme počet jeho prvků.

Mějme náhodnou permutaci čísel 1, 2, ... , n, tj. každá permutace nastává se stejnou pravděpodobností (a tato pravděpodobnost je $\frac{1}{n!}$). Jaká je střední hodnota délky počátečního rostoucího úseku této náhodné permutace?

Označme náhodnou veličinu udávající délku počátečního úseku náhodné permutace $(Q_1, \dots, Q_n)$ písmenem X. Označme dále I_k indikátor jevu (tj. náhodnou veličinu, jež je rovna jedné, pokud jev nastal, a nule, pokud nenastal), že prvních k čísel v permutaci je v rostoucím pořadí, neboli $Q_1 < Q_2 < \cdots < Q_k$. Povšimněme si, že $I_j \sim Alt(\frac{1}{j!})$ (jediná z $j!$ permutací je rostoucí). Zjevně platí

$$X = I_1 + I_2 + \cdots + I_n,$$

tudíž

$$\mathrm{E}\, X = \mathrm{E}\, I_1 + \mathrm{E}\, I_2 + \cdots + \mathrm{E}\, I_n.$$

Ovšem střední hodnoty indikátorů (alternativně rozdělených veličin) spočteme snadno, tedy

$$\mathrm{E}\, X = 1 + \frac{1}{2} + \cdots + \frac{1}{n!} = \sum_{i=1}^{n} \frac{1}{i!}.$$

A je to!

7. úloha

Dokažte, že v textu použité odhady střední hodnoty a rozptylu jsou správné, neboli ověřte, že skutečně platí:

i) $\mathrm{E}\, \overline{X}_n = \mathrm{E}\, X$,

ii) $\operatorname{Var} \overline{X}_n = \frac{1}{n} \operatorname{Var} X$,

iii) $\mathrm{E}\, S^2 = \operatorname{Var} X$.

i) Střední hodnota součtu je součet středních hodnot, proto

$$\mathrm{E}\, \overline{X}_n = \mathrm{E}\, \frac{1}{n} \sum_{i=1}^{n} X_i = \frac{1}{n} \sum_{i=1}^{n} \mathrm{E}\, X_i = \frac{n}{n} \mathrm{E}\, X = \mathrm{E}\, X.$$

ii) Použijeme vytýkání konstanty z rozptylu a fakt, že rozptyl součtu nezávislých veličin je roven součtu rozptylů těchto veličin:

$$\operatorname{Var} \overline{X}_n = \operatorname{Var} \frac{1}{n} \sum_{i=1}^{n} X_i = \frac{1}{n^2} \operatorname{Var} \sum_{i=1}^{n} X_i = \frac{1}{n^2} \sum_{i=1}^{n} \operatorname{Var} X_i = \frac{n}{n^2} \operatorname{Var} X = \frac{1}{n} \operatorname{Var} X.$$

iii) Tato úloha je o něco obtížnější. Postupně využijeme toho, že součet středních hodnot je střední hodnota součtu, že rozptyl součtu nezávislých veličin je součet rozptylů a že střední hodnota součinu nezávislých veličin je součin středních hodnot:

$$\mathrm{E}\sum_{i=1}^{n}(X_i-\overline{X}_n)^2=\mathrm{E}\sum_{i=1}^{n}((X_i-\mathrm{E}\,X)-(\overline{X}_n-\mathrm{E}\,X))^2=$$

$$=\mathrm{E}\sum_{i=1}^{n}(X_i-\mathrm{E}\,X)^2+\mathrm{E}\sum_{i=1}^{n}(\overline{X}_n-\mathrm{E}\,X)^2-\frac{2}{n}\,\mathrm{E}\sum_{i=1}^{n}\sum_{j=1}^{n}(X_i-\mathrm{E}\,X)(X_j-\mathrm{E}\,X)=$$

$$=n\operatorname{Var}X+\frac{n}{n}\operatorname{Var}X-\frac{2}{n}\,\mathrm{E}\sum_{i=j}(X_i-\mathrm{E}\,X)(X_j-\mathrm{E}\,X)-\frac{2}{n}\,\mathrm{E}\sum_{i\neq j}(X_i-\mathrm{E}\,X)(X_j-\mathrm{E}\,X)=$$

$$=(n+1)\operatorname{Var}X-\frac{2n}{n}\operatorname{Var}X-\frac{2(n^2-n)}{n}\,0=(n-1)\operatorname{Var}X,$$

odtud již snadno

$$\mathrm{E}\,S^2=\mathrm{E}\,\frac{1}{n-1}\sum_{i=1}^{n}(X_i-\overline{X}_n)^2=\operatorname{Var}X.$$

8. úloha

Král Kulatého království měl ukrutně dlouhý nos a poddaní se mu kvůli němu smáli a nebrali ho vážně. Králův rádce králi poradil, že když nechá děti svých poddaných místo jablek jíst frgálovník[6], naroste i jim delší nos a za pár let už se mu nikdo smát nebude.

Král byl statisticky vzdělán, a tak se rozhodl před masovou výsadbou frgálovníků po celém království rádcovu radu nejprve ověřit. Nejprve nahlédl do knihovny a zjistil, že rozptyl délky nosu pětiletých dětí byl mnoha lety praxe určen jako 0,5 cm^2. Poté vybral dvě skupiny malých dětí, každou o 100 jedincích. Děti z první skupiny dostávaly klasická jablíčka, děti z druhé skupiny jedly místo nich frgálovníky[7]. V pěti letech byla změřena délka jejich nosu. Zatímco v první skupině byl průměr délky nosů roven 3,5 cm, v druhé byl o něco vyšší, totiž 3,9 cm.

Skeptikové však tvrdili, že se jedná o náhodu a že frgálovník není účinný. Co jim král na to odpověděl?

Král řekl: No jo, vyloučit to nemůžu (to bych nemohl, ani kdyby nosy frgálovníkových dětí byly třikrát tak dlouhé). Na druhou stranu, pokud by zde skutečně žádný vliv frgálovníku nebyl, pak by rozdíl nosů dvou dětí (první frgálovníkové, druhé ne) byla náhodná veličina se střední hodnotou 0 a rozptylem 1[8].

---

[6] Frgálovník je ovoce k nerozeznání podobné jablkům až na drobné vedlejší účinky při konzumaci.

[7] Král byl i předsedou a jediným členem Etické komise Kulatého království.

[8] Využíváme faktu, že střední hodnota resp. rozptyl součtu nezávislých veličin je roven součtu středních hodnot resp. rozptylů.

Na průměr sta těchto rozdílů (označme jej $\overline{X}$) tedy mohu aplikovat (CLV) a dostávám, že rozdělení $\sqrt{100}\frac{\overline{X}}{1} \approx Norm$. Pokud se mi tento rozdíl odchýlí daleko od nuly (nad kritickou mez k), pak prohlásím, že frgálovník zjevně nosy prodlužuje. Jak daleko od nuly? To záleží na tom, jakou pravděpodobnost omylu (prohlášení frgálovníku za účinný, i když není) jsme ochotni přijmout. Řekněme, že akceptujeme pravděpodobnost omylu 1%, neboli chceme, aby za předpokladu neúčinnosti

$$\mathrm{P}(\overline{X} \le k) = 0{,}99.$$

My však víme, že za dané hypotézy $10\overline{X} \approx Norm$, tudíž po dosazení z tabulek

$$k = \frac{1}{10} \cdot 2{,}326348 \doteq 0{,}23.$$

Zároveň si uvědomme, že realizace $\overline{X}$ je v našem případě $3{,}9 - 3{,}5 = 0{,}4$, tedy číslo zdaleka překračující kritickou mez. Rozdíl mezi nosy frgálovníkových a nefrgálovníkových dětí tedy můžeme považovat za statisticky průkazný.

Poznámky opravovatele: Překvapilo mne, kolik rozmanitých řešení se řešitelům podařilo vymyslet. A přestože ne všechna byla zcela statisticky košér, uznávám, že to bylo způsobeno především mým intuitivním, neprecizním výkladem látky, a po zásluze jsem je ocenil.

9. úloha

Nechť X je náhodná veličina nabývající hodnot 1 a -1, obou s pravděpodobností $\frac{1}{2}$. Precizně ověřte, že pro její rozdělení platí námi předpokládaný (ZVČ). Tedy dokažte, že pro libovolně malé $\epsilon > 0$ platí, že pokud vezmeme n dostatečně vysoké (tedy větší nebo rovno nějaké mezi n_ϵ), tak potom

$$\mathrm{P}(|\overline{X}_n| \ge \epsilon) \le \epsilon.$$

Náhodná veličina $\overline{X}_n$ nabývá jistě konečného počtu hodnot, označme množinu jejích hodnot M a navíc zaveďme

$$M_1 = \{m \in M : m \ge \epsilon\},$$
$$M_0 = \{m \in M : m < \epsilon\},$$

potom za použití drobného triku dostáváme

$$\mathrm{P}(|\overline{X}_m| \ge \epsilon) = \sum_{m \in M_1} 1 \cdot \mathrm{P}(|\overline{X}_m| = m) + \sum_{m \in M_0} 0 \cdot \mathrm{P}(|\overline{X}_m| = m) =$$

$$= \frac{1}{\epsilon^2} \left[\sum_{m \in M_1} \epsilon^2 \cdot \mathrm{P}(|\overline{X}_m| = m) + \sum_{m \in M_0} 0 \cdot \mathrm{P}(|\overline{X}_m| = m) \right] \le$$

$$\le \frac{1}{\epsilon^2} \left[\sum_{m \in M_1} m^2 \cdot \mathrm{P}(|\overline{X}_m| = m) + \sum_{m \in M_0} m^2 \cdot \mathrm{P}(|\overline{X}_m| = m) \right] =$$

$$= \frac{1}{\epsilon^2} \left[\sum_{m \in M} m^2 \cdot \mathrm{P}(|\overline{X}_m| = m) \right] = \frac{1}{\epsilon^2} \,\mathrm{E}\, |\overline{X}_n|^2 = \frac{1}{\epsilon^2} \operatorname{Var} \overline{X}_n = \frac{1}{\epsilon^2} \cdot \frac{1}{n},$$

pokud tedy vezmeme $n \geq n_\epsilon = \frac{1}{\epsilon^3}$, potom nám bude platit v zadání požadovaná nerovnost.

Úlohu by snad mělo být možné dokázat i přímo (bez triku), ovšem mnohem pracnějším způsobem.