5. cvičení

Střední hodnota

Úloha 1 (První setkání) a) Nechť $P(X = 100) = p$, $P(X = 0) = 1 - p$. Určete $\mathbb{E}(X)$. (Přímo nebo pomocí některé vlastnosti střední hodnoty.)

b) Nechť $P(Y = 100) = p$, $P(Y = 99) = 1 - p$. Určete $\mathbb{E}(Y)$.

Řešení

- $\mathbb{E}(X) = 100p$.
- $\mathbb{E}(Y) = 99 + p$.

Úloha 2 („Průměrná“ rychlost)
Předpokládejme, že vyřešení jednoho příkladu trvá X minut, kde $X = 1, 2, \ldots,$ nebo 5. Doba trvání je náhodná (závislá na počasí) a pravděpodobnostní funkce je $p_X(1) = p_X(2) = 0.1$, $p_X(3) = p_X(4) = 0.2$, $p_X(5) = 0.4$. Spočtěte $\mathbb{E}(X)$.

Řešení
$\mathbb{E}[X] = 0.1 + 0.1 \cdot 2 + 0.2 \cdot 3 + 0.2 \cdot 4 + 0.4 \cdot 5 = 3.7$

Úloha 3 (Lovci ponožek)
Máme neomezený počet černých a červených ponožek. Ponožky vytahujeme poslepu, obě barvy jsou stejně pravděpodobné.

a) Kolik ponožek ve střední hodnotě vytáhneme, než budeme mít dvě stejné barvy?

b) Řešte totéž pro tři různé barvy.

Řešení a) S pravděpodobností 1/2 nám stačí 2 tahy, a za 3 tahy máme dvě barvy určitě (tj. pravděpodobnost je zbývající polovina), tedy $\mathbb{E}[X] = 2.5$

b) $2 \cdot 3 \cdot (1/3)^2 + 3 \cdot 6 \cdot (1/3)^2 \cdot (2/3) + 4 \cdot (1 - 1/3 - 4/9) = 2 \cdot 1/3 + 3 \cdot 4/9 + 4 \cdot 2/9 = 26/9$

Úloha 4 (Know when to walk away)
Nové kasino nabízí následující hru: vsadíme x korun, s pravděpodobností 1/2 o ně přijdeme, ale s pravděpodobností 1/2 vyhrajeme $2x$ (navíc k našim x korunám).

a) Začínáme s k korunami. Pokud chceme maximalizovat střední hodnotu našich zisků po n kolech, jak to udělat (a kolik ta střední hodnota je)?

b) Jaká je pravděpodobnost, že s takovou strategií přijdeme o všechny peníze?

c) Jakou strategii byste zvolili?

Řešení a) Když vsadíme x, máme střední hodnotu $3x/2$. Indukcí pak máme, že po ℓ kolech bychom měli $(\frac{3}{2})^\ell x$, a tedy je nejlepší vždycky sázet celou částku pro $(\frac{3}{2})^n k$

b) Musíme aspoň jednou prohrát, což se stane s pravděpodobností $1 - 2^{-n}$.

c) Nějak sázet méně. Optimum je tzv. Kellyho kritérium, které říká, že máme vsadit f^*-zlomek celkového jmění, kde $f^* = p - \frac{1-p}{v}$, kde p je pravděpodobnost výhry a v je poměrový zisk (tj. kolik jsme nově získali děleno tím, kolik jsme vsadili). V našem případě je $p = 1/2, v = 2$, a tedy $f^* = 1/4$ – měli bychom sázet vždycky čtvrtinu

Linearita

Úloha 5 (Hody mincí)
Hodíme n-krát korunou, která má pravděpodobnost, že padne Panna, rovnou p.

a) Označme X počet po sobě jdoucích hodů PO. (Např. pokud $n = 6$ a padlo postupně POOPOO, tak $X = 2$.) Určete $\mathbb{E}(X)$.

b) Rozmyslete si, proč se nejedná o binomické rozdělení.

c) Označme teď Y počet opakování hodů POP, jaká je $\mathbb{E}(Y)$?

Řešení a) $\mathbb{E}(X) = (n-1) \cdot p \cdot (1-p)$

b) Třeba protože nikdy nemůže nabýt hodnoty $n-1$.

c) $\mathbb{E}(Y) = (n-2) \cdot p^2 \cdot (1-p)$

Úloha 6 ($G(n,p)$)
Hodíme $\binom{n}{2}$-krát korunou, na které padne Panna s pravděpodobností p. Přitom tvoříme graf s vrcholy $V = \{1, 2, \ldots, n\}$. Postupně pro všechny dvojice $\{i, j\} \in \binom{V}{2}$ určíme, jestli jsou spojené hranou – to bude tehdy, když příslušným hodem padla Panna. Vzniklému grafu se říká (Erdősův-Rényiho) náhodný graf $G(n,p)$.
Ukažte, že střední hodnota počtu hran v grafu je $p\binom{n}{2}$ a střední hodnota počtu trojúhelníků v grafu je $p^3\binom{n}{3}$.

Řešení
Standardní indikátory: máme $\binom{n}{2}$ hran, každá s pravděpodobností p. Máme $\binom{n}{3}$ trojúhelníků, každý se objeví s pravděpodobností p^3 (všechny hrany musí být vybrané).

Podmíněná střední hodnota

Úloha 7 (Kvíz)
V kvízu je 20 otázek s volbami a,b,c,d. Za správnou odpověď (vždy je jen jedna odpověď správná) je 1 bod, za špatnou $-1/4$ bodu, za nevyplněnou otázku nula. Každá otázka je s pravděpodobností q jednou z těch, co se Kvído naučil a tedy zná správnou odpověď. Pokud správnou odpověď nezná, ví o tom, a může se rozhodnout, zda tipovat.

a) Jaká je střední hodnota počtu bodů, které Kvído získá, pokud bude odpovídat jenom otázky, u kterých zná odpověď?

b) A co když bude tipovat, když nezná správnou odpověď?

c) Jak by se musela změnit penalizace za chybnou odpověď, aby byly odpovědi v částech a, b stejné?

Řešení
Vždycky použijeme linearitu a $\mathbb{E}(X) = P(Z)\mathbb{E}(X|Z) + P(Z^c)\mathbb{E}(X|Z^c)$, kde $P(Z) = q, \mathbb{E}(X|Z) = 1$, Y je pak n.v. pro body z 20 otázek

1. Tady $\mathbb{E}(X|Z^c) = 0$, a tedy $\mathbb{E}(Y) = 20q$
2. Tedy je $\mathbb{E}(X|Z^c) = 1 \cdot 1/4 + (-1/4) \cdot (3/4) = 1/16$, a $\mathbb{E}(Y) = 20\left(q + \frac{1-q}{16}\right)$
3. Měla by být $-1/3$, aby $\mathbb{E}(X|Z^c)$ ve druhém případě vyšlo 0.

Úloha 8 (Trpělivost)
Můj počítač občas zlobí: každý den s pravděpodobností $p > 0$ zamrzne. Když se to stane dva dny po sobě, začnu to řešit. Jaký je střední počet dnů, než se to stane?

Řešení
Označíme si jevy B_1, B_2, B_3 jako jevy, kde B_1 je „první dva dny zamrznul“, B_2 je „první den zamrznul, druhý ne“, a B_3 je „první den nezamrznul“. Pak $\mathbb{E}[X|B_1] = 2$, $\mathbb{E}[X|B_2] = \mathbb{E}[X] + 2, \mathbb{E}[X|B_3] = \mathbb{E}[X] + 1$. Celkem tedy $\mathbb{E}[X] = p^2 \cdot 2 + p(1-p)(\mathbb{E}[X]+2) + (1-p)(\mathbb{E}[X]+1) = 2p^2 + 2p - 2p^2 + p\mathbb{E}[X] - p^2\mathbb{E}[X] + \mathbb{E}[X] + 1 - p\mathbb{E}[X] - p = p + 1 + \mathbb{E}[X](1-p^2) \rightsquigarrow p^2\mathbb{E}[X] = p + 1 \rightsquigarrow \mathbb{E}[X] = \frac{p+1}{p^2}$.

Úloha 9 (Riskuj!)
V televizní soutěži si účastník může vybrat dvě otázky. U otázky A odhaduje, že správně odpoví s pravděpodobností 0.8 (a dostane za to 1 000 Kč). U otázky B je jeho pravděpodobnost úspěchu jen 0.5, zato za správnou odpověď dostane 2 000 Kč. Po špatné odpovědi hra končí, po správné může zkusit druhou otázku (a odměna za už správně odpovězenou otázku mu při špatně odpovězené další nepropadne).

a) Jaká je střední hodnota výhry, pokud začne otázkou A?

b) Co když začne otázkou B?

c) Bonus: pokud jsou pravděpodobností úspěchu p_A, p_B a odměny m_A, m_B, jak se má soutěžící rozhodnout?

Řešení

1. $0.8 \cdot 0.5 \cdot 3000 + 0.8 \cdot 0.5 \cdot 1000 = 1600$
2. $0.8 \cdot 0.5 \cdot 3000 + 0.2 \cdot 0.5 \cdot 2000 = 1400$
3. Začít s první: $p_1(p_2 \cdot (m_1 + m_2) + (1 - p_2)m_1)$, začít s druhou: $p_2(p_1 \cdot (m_1 + m_2) + (1 - p_1)m_2)$ – po úpravách je výhodnější začít s druhou otázkou, pokud $m_1 > m_2 \cdot \frac{(1-p_1)p_2}{p_1(1-p_2)}$.

Bonusové úlohy

Úloha 10
Připomeňte si definici indikátorové náhodné veličiny I_A.

a) Jaká je $\mathbb{E}(I_A)$?

b) Nechť $A = A_1 \cup A_2 \cup \cdots \cup A_n$. Ověřte rovnost $1 - I_A = \prod_{i=1}^{n}(1 - I_{A_i})$.

c) Roznásobte a použijte větu o linearitě střední hodnoty. Dostanete princip inkluze a exkluze, který znáte z diskrétní matematiky.

Řešení

1. $P(I_A)$?
2. Pokud $I_A = 1$, pak nějaké $I_{A_i} = 1$ a naopak.
3. TODO

Úloha 11 (Sběratel Hot Wheels)
Ke každému nákupu dostaneme jako dárek autíčko – náhodně vybrané z n typů. Kolik průměrně musíme udělat nákupů, než dostaneme všechny typy autíček?

Řešení
Nemůžeme udělat náhodné veličiny typu "počet jiných autíček do prvního správného autíčka", takže uděláme následující: celkový počet nákupů budiž $T = T_1 + \ldots + T_n$, kde T_i je počet nákupů mezi získáním $(i-1)$-ního a i-tého nového autíčka. Jistě vidíme, že $T_i \sim \mathrm{Geo}(\frac{n-(i-1)}{n})$, což má střední hodnotu $\mathbb{E}[T_i] = \frac{n}{n-(i-1)}$. Pak $\mathbb{E}[T] = \sum \mathbb{E}[T_i] = \sum_{i=1}^{n} \frac{n}{n-(i-1)} = n \cdot H_n = n \log n + \gamma n + \frac{1}{2} + \mathcal{O}(\frac{1}{n})$, kde γ je Eulerova-Mascheroniho konstanta, nebo obecně slaběji $\mathcal{O}(n \log n)$.

Na procvičení

Úloha 12 (Ludvík kam se podíváš)
Král Ludvík chce mít mužského potomka, aby ho mohl opět pojmenovat Ludvík. V každém roce mu jeho manželka porodí právě jedno dítě, které je stejně pravděpodobně chlapec i děvče, nezávisle na předchozích pokusech. Všechny narozené děti přežijí. Pokud se narodí chlapec, tak další potomky už Ludvík mít nebude. Označme S počet narozených synů a D počet narozených dcer. Určete $\mathbb{E}(S)$ a $\mathbb{E}(D)$.

Řešení

a) $\mathbb{E}(S) = 1$ (deterministicky)

b) $\mathbb{E}(D) = 1$ (je potřeba spočítat, ale je to vlastně $\mathbb{E}[\mathrm{Geo}(1/2)] - 1$)

Užitečné pojmy

- Střední hodnota: $\mathbb{E}(X) = \sum_{x \in Im(X)} x \cdot P(X = x)$
- Indikátor jevu A je náhodná veličina $I_A = 1[A]$, která nabývá hodnoty 0 když A nenastane a hodnoty 1 když A nastane
- Podmíněná střední hodnota:

$$\mathbb{E}(X|B) = \sum_{x \in Im(X)} x \cdot P(X = x|B)$$

$$\mathbb{E}(X) = \sum_{i} P(B_i) \cdot \mathbb{E}(X|B_i) \qquad B_1, B_2, \ldots \text{ je rozklad } \Omega$$