

5. cvičení

Podmíněná střední hodnota

Úloha 1 (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.

- Jaká je střední hodnota počtu bodů, které Kvído získá, pokud bude odpovídat jenom otázky, u kterých zná odpověď?
- A co když bude tipovat, když nezná správnou odpověď?
- 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

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

Úloha 2 (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 zamrznu“, B_2 je „první den zamrznu, druhý ne“, a B_3 je „první den nezamrznu“. 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 3 (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).

- Jaká je střední hodnota výhry, pokud začne otázkou A?
- Co když začne otázkou B?
- 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 $m_1 > m_2 \cdot \frac{(1-p_1)p_2}{p_1(1-p_2)}$.

Název	Pravděpodobnostní funkce	Rozsah (Im X)	Střední hodnota	Rozptyl
Bernoulliho $Ber(p)$	$p_X(1) = p, p_X(0) = 1 - p$	$\{0, 1\}$	p	$p(1 - p)$
Binomické $Bin(n, p)$	$p_X(k) = \binom{n}{k} p^k (1 - p)^{n-k}$	$\{0, 1, \dots, n\}$	np	$np(1 - p)$
Geometrické $Geo(p)$	$p_X(k) = (1 - p)^{k-1} p$	$\{1, 2, \dots\}$	$\frac{1}{p}$	$\frac{1-p}{p^2}$
Poissonovo $Poi(\lambda)$	$p_X(k) = e^{-\lambda} \frac{\lambda^k}{k!}$	$\{0, 1, \dots\}$	λ	λ
Uniformní $Uni(a, b)$	$p_X(k) = \frac{1}{b-a+1}$	$\{a, a+1, \dots, b\}$	$\frac{a+b}{2}$	$\frac{(b-a+1)^2 - 1}{12}$

Poznávka náhodných veličin

Úloha 4 (Hacker)

Pravděpodobnost, že do našeho serveru pronikne hacker je během každého dne 0.01, nezávisle pro každý den. Označme T počet dnů do prvního průniku. Jaké je rozdělení T , $\mathbb{E}(T)$, $\text{var}(T)$? Jaká je pravděpodobnost, že server zůstane bezpečný po celý rok?

Řešení

$T \sim \text{Geom}(0.01)$, $\mathbb{E}(T) = 100$, $\text{var}(T) = \frac{1-p}{p^2} = \frac{0.99}{0.0001} = 9900$, $P[365 \text{ dní bez incidentu}] = 0.99^{365} \approx 0.025518$

Úloha 5 (Test-driven probability)

Každý test programu může skončit buď nalezením chyby (úspěch) nebo ne (neúspěch). Předpokládáme, že pravděpodobnost nalezení chyby při jednom testu je 0.05 a vývojář provede 20 nezávislých testů, označíme X počet nalezených chyb. Jaké je rozdělení X , $\mathbb{E}(X)$, $\text{var}(X)$? Jaká je pravděpodobnost, že najde právě tři chyby?

Řešení

$X \sim \text{Bin}(20, 0.05)$, $\mathbb{E}[X] = 1$, $\text{var}(X) = np(1 - p) = 0.95$, $P[X = 3] = \binom{20}{3} 0.05^3 \cdot 0.95^{17} \approx 0.059582$

Úloha 6 (Vyřizování žádostí)

Historická data ukazují, že náš server obdrží průměrně 30 žádostí za minutu. Použijte Poissonovo rozdělení k určení pravděpodobnosti, že server obdrží přesně 40 žádostí v následující minutě.

Řešení

Použijeme tedy $Z \sim \text{Poi}(30) \rightsquigarrow P(Z = 40) = \frac{30^{40}}{40!} \cdot e^{-30} = \frac{584517784584313631057739257812500}{3922767865085986845929e^{30}} \approx 0.0139435$

Rozptyl

- **Definice** rozptyl X je $\text{var}(X) = \mathbb{E}((X - \mathbb{E}(X))^2)$
- **Věta:** $\text{var}(X) = \mathbb{E}(X^2) - \mathbb{E}(X)^2$
- **Věta:** $\text{var}(aX + b) = a^2 \text{var}(X)$
- **Věta:** pokud $X \perp Y$, máme $\text{var}(X + Y) = \text{var}(X) + \text{var}(Y)$
- **Definice:** směrodatná odchylka X je $\sigma_X = \sqrt{\text{var}(X)}$
- **Definice:** variační koeficient X je $CV_X = \sigma_X / \mathbb{E}(X)$ (pokud $\mathbb{E}(X) > 0$)

Úloha 7 (Rozptyl z definice)

Spočtete přímo z definice rozptyl $\text{Unif}(a, b)$ – stačí pro $a = -2$, $b = 2$. Srovnajte s tabulkou na minulé straně.

Řešení

$\text{var}(X) = \frac{4+1+0+1+4}{5} = 2$.

Úloha 8 (Slide to the right)

Dokažte, že $\text{var}(X + b) = \text{var}(X)$.

Řešení

$\mathbb{E}[(X + b)^2] - \mathbb{E}[X + b]^2 = \mathbb{E}[X^2 + 2bX + b^2] - (\mathbb{E}[X] + b)^2 = \mathbb{E}[X^2] + 2b\mathbb{E}[X] + b^2 - \mathbb{E}[X]^2 - 2b\mathbb{E}[X] - b^2 = \mathbb{E}[X^2] - \mathbb{E}[X]^2 = \text{var}(X)$

Úloha 9 (Geometrické rozdělení a škálování)

Bud' $X \sim \text{Geo}(0.1)$ a $Y \sim \text{Geo}(0.01)$. Zvolte konstantu c tak, aby veličiny X a $Z = cY$ měly stejnou střední hodnotu. Porovnejte rozptyl, směrodatnou odchylku a variační koeficient pro X , Y , Z .

Řešení

$\mathbb{E}[X] = 10, \mathbb{E}[Y] = 100$, tedy je třeba nastavit $c := 1/10$. Pak z tabulky $\text{var}(X) = \frac{1-p}{p^2} = \frac{9/10}{1/100} = 90$, $\text{var}(Z) = c^2 \cdot \text{var}(Y) = \frac{1}{100} \cdot \frac{99/100}{1/10000} = 99$. Směrodatné odchylky jsou pak $\sigma_X = \sqrt{90}, \sigma_Y = \sqrt{99}$ a variační koeficienty jsou totéž, jen vydělené deseti.