16.12.2015, 2. skupina (18:00 - 19:30) Jméno:

Zápočtový test z PSI

Nezapomeňte podepsat VŠECHNY papíry, které odevzdáváte. Škrtejte zřetelně a stejně zřetelně pište i věci, které platí. Co je škrtnuto, nebude bráno v úvahu a naopak. Jestliže něčemu nerozumíte, zeptejte se. Postup je třeba odůvodnit (okomentovat) nebo uvést výpočet. Výsledek bez uvedení jakéhokoliv postup či výpočtu není akceptován. Abyste uspěli v testu, potřebujete získat alespoň 15 bodů.

(1) (10 bodů) Máme 4 krabice stejného vzhledu. V první jsou 3 bílé a 2 černé koule, ve druhé jsou 2 bílé a 2 černé koule, ve třetí je 1 bílá a 4 černé koule, ve čtvrté 5 bílých a 1 černá koule. Pokud nám někdo řekl, že náhodně vybral jednu z krabic a vytáhl 1 kouli, která byla bílá, s jakou pravděpodobností můžeme usuzovat, že v téže krabici se nachází alespoň 3 černé koule?

Řešení:
Označme jevy:

$A_i =$ "byla vybrána i-tá krabice",
$B =$ "koule vytažená z vybrané krabice je bílá".

pro $i = 1, 2, 3, 4$. Víme, že A_1, A_2, A_3 a A_4 je úplný disjunktní systém jevů, o kterých předpokládáme, že jsou stejně pravděpodobné. Tedy

$$P(A_1) = P(A_2) = P(A_3) = P(A_4) = \frac{1}{4}$$

$$P(B|A_1) = \frac{3}{5} \quad P(B|A_2) = \frac{2}{4} \quad P(B|A_3) = \frac{1}{5} \quad P(B|A_4) = \frac{5}{6} .$$

Protože jediná krabice obsahující alespoň 3 černé koule je třetí krabice, zajímá nás $P(A_3|B)$. Z Bayesovy věty máme:

$$P(A_3|B) = \frac{P(B|A_3) \cdot P(A_3)}{P(B)}$$

$$P(B) = \sum_{i=1}^{4} P(B|A_i) \cdot P(A_i) = \frac{1}{4} \cdot \left(\frac{3}{5} + \frac{2}{4} + \frac{1}{5} + \frac{5}{6}\right) = \frac{8}{15} \doteq 0.5333$$

a tedy

$$P(A_3|B) = \frac{1/5 \cdot 1/4}{8/15} = \frac{3}{32} = 0.09375 .$$

Ještě je dobré uvědomit si, jak vypadá Kolmogorův model, speciálně jaké jsou pravděpodobnosti elementárních jevů:

- elementární jev ω bude dvojice "(výběr krabice, vytažení koule z této krabice)" a Ω bude tedy množina všech takových ω,
- $A_i = \{\omega \in \Omega \mid \omega = (\text{výběr } i\text{-té krabice, vytažení koule z této krabice})\}$,
- pro počet krabic $k = 4$ a počet koulí k_i v i-té krabici pak pro $\omega \in A_i$ máme $P(\{\omega\}) = \frac{1}{k \cdot k_i}$,
- pravděpodobnost jevu A_i, tedy výběr i-té krabice, pak je $P(A_i) = \sum_{\omega \in A_i} \frac{1}{k \cdot k_i} = \frac{k_i}{k \cdot k_i} = \frac{1}{k}$.

Můžeme si tedy všimnout, že pravděpodobnost vytažení koule z dané vybrané krabice nejsou všechny stejné, zatímco pravděpodobnosti výběru dané krabice ano.

Jak by situace vypadala, kdybychom v zadání předpokládali, že koule nevybíráme z krabic, ale sesypeme je do jedné nádoby, přičemž si na každou z nich napíšeme z jaké krabice pochází? Pravděpodobností se pak změní:

- elementární jev ω' bude dvojice "(číslo krabice, koule pocházející z krabice s tímto číslem)" a Ω' bude tedy množina všech takových ω',

- jev $A_i' = \{\omega' \in \Omega' \mid \omega' = (i, \text{koule pocházející z } i\text{-té krabice})\}$ představuje všechny koule, které pocházejí z i-té krabice,
- protože nyní taháme koule z nádoby, budeme považovat pravděpodobnost jejich vytažení za stejnou pro všechny koule, tj. $P(\{\omega\}) = \frac{1}{K}$, kde $K = \sum_j k_j$,
- pravděpodobnost jevu A_i, tedy podíl počtu koulí z i-té krabice, pak bude $P(A_i) = \sum_{\omega \in A_i'} \frac{1}{K} = \frac{k_i}{K}$.

Sesypáním koulí jsme tak způsobili to, že šance na vytažení kterékoliv koule se teď vyrovnaly a naopak pravděpodobnosti "příslušející krabicím"jsou různé. Pokud bychom tedy řešili takto pozměněné zadání, tak pro jev

B' = "koule vytažená z nádoby je bílá"

a úplný disjunktní systém jevů A_1', A_2', A_3' a A_4' dostáváme (s celkovým počtem koulí $5 + 4 + 5 + 6 = 20$), že

$$P(A_1') = \frac{5}{20} \qquad P(A_2') = \frac{4}{20} \qquad P(A_3') = \frac{5}{20} \qquad P(A_4') = \frac{6}{20}$$

$$P(B'|A_1') = \frac{3}{5} \qquad P(B'|A_2') = \frac{2}{4} \qquad P(B'|A_3') = \frac{1}{5} \qquad P(B'|A_4') = \frac{5}{6} \ .$$

Opět jediná krabice obsahující alespoň 3 černé koule je třetí krabice, takže nás bude zajímat $P(A_3'|B')$. Z Bayesovy věty máme:

$$P(A_3'|B') = \frac{P(B'|A_3') \cdot P(A_3')}{P(B')}$$

$$P(B') = \sum_{i=1}^{4} P(B'|A_i') \cdot P(A_i') = \frac{3}{5} \cdot \frac{5}{20} + \frac{2}{4} \cdot \frac{4}{20} + \frac{1}{5} \cdot \frac{5}{20} + \frac{5}{6} \cdot \frac{6}{20} = \frac{11}{20} = 0.55$$

(což se snadněji dalo spočítat prostě podílem bílých koulí v nádobě jako $P(B') = \frac{3+2+1+5}{20} = \frac{11}{20}$) a tedy

$$P(A_3'|B') = \frac{1/5 \cdot 5/20}{11/20} = \frac{1}{11} \doteq 0.0909$$

(což se opět snadněji dalo spočítat prostě podílem bílých koulí pocházejících ze 3. krabice v rámci všech bílých koulí, tj. $P(A_3'|B') = \frac{1}{3+2+1+5} = \frac{1}{11}$).

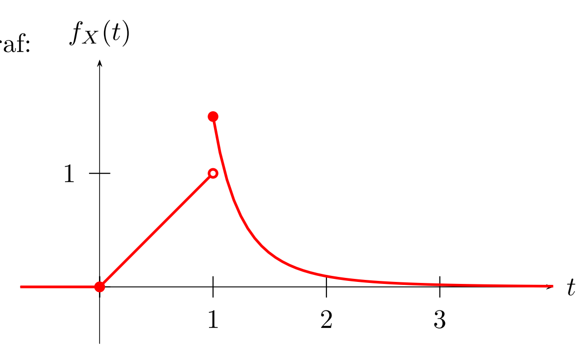
(2) (10 bodů) Rozdělení náhodné veličiny X je dáno hustotou

$$f_X(t) = \begin{cases} t & , \ t \in \langle 0, 1) \\ \frac{3}{2t^4} & , \ t \in \langle 1, +\infty) \\ 0 & , \ \text{jinak.} \end{cases}$$

Určete střední hodnotu a rozptyl veličiny X a hodnotu $P(-1 \leq X < 2)$.

Řešení:

Hustota má graf:



Pomocí ní spočítáme střední hodnotu $E(X)$ a druhý moment $E(X^2)$:

$$E(X) = \int_{-\infty}^{\infty} t \cdot f_X(t)\ dt = \int_0^1 t \cdot t\ dt + \int_1^{\infty} t \cdot \frac{3}{2t^4}\ dt = \left[\frac{t^3}{3}\right]_0^1 + \left[-\frac{3}{4t^2}\right]_1^{\infty} = \frac{1}{3} + \frac{3}{4} = \frac{13}{12} \doteq 1.083$$

$$E(X^2) = \int_{-\infty}^{\infty} t^2 \cdot f_X(t)\ dt = \int_0^1 t^2 \cdot t\ dt + \int_1^{\infty} t^2 \cdot \frac{3}{2t^4}\ dt = \left[\frac{t^4}{4}\right]_0^1 + \left[-\frac{3}{2t}\right]_1^{\infty} = \frac{1}{4} + \frac{3}{2} = \frac{7}{4} = 1.75$$

Tedy

$$D(X) = E(X^2) - \big(E(X)\big)^2 = \frac{7}{4} - \left(\frac{13}{12}\right)^2 = \frac{83}{144} \doteq 0.5764\ .$$

A nakonec

$$P(-1 \le X < 2) = \int_{-1}^{2} f_X(t)\ dt = \int_0^1 t\ dt + \int_1^2 \frac{3}{2t^4}\ dt =$$

$$= \left[\frac{t^2}{2}\right]_0^1 + \left[-\frac{1}{2t^3}\right]_1^2 = \frac{1}{2} - \frac{1}{16} + \frac{1}{2} = \frac{15}{16} = 0.9375$$

(3) (10 bodů) Před volbami je v populaci státu 52% příznivců koaliční strany. Jaká je pravděpodobnost, že průzkum veřejného mínění o rozsahu $n = 1500$ ukáže nesprávně převahu opozice?

Řešení:
Jednotlivý volič volí tedy koalici s pravděpodobností $p = 0.52$. Pro $i = 1, \ldots, n$ si zavedeme veličiny

$$X_i = \begin{cases} 1 & ,\ i\text{-tý člověk z průzkumu zvolí koalici,} \\ 0 & ,\ i\text{-tý člověk z průzkumu zvolí opozici.} \end{cases}$$

Veličiny X_i považujeme za nezávislé, s alternativním rozdělení s parametrem p (protože $P(X_i = 1) = p$), střední hodnotou $E(X_i) = p = 0.52$ a rozptylem $D(X_i) = p(1-p) = 0.52 \cdot 0.48 = 0.2496$.

Preference koaliční strany v průzkumu se pak vyjádří jako

$$\overline{X} = \frac{1}{n} \sum_{i=1}^{n} X_i$$

(tj. počet lidí z průzkumu, kteří by volili koalici, ku počtu všech dotázaných). Pravděpodobnost, že se ukáže převaha opozice je $P(\overline{X} < 0.5)$, a její hodnotu odhadneme pomocí centrální limitní věty použité na normovanou veličinu $\mathrm{norm}(\overline{X}) = \frac{\overline{X} - E(\overline{X})}{\sqrt{D(\overline{X})}}$. K tomu potřebujeme znát:

$$E(\overline{X}) = \frac{1}{n} \sum_{i=1}^{n} E(X_i) = p = 0.52$$

$$D(\overline{X}) = \frac{1}{n^2} \sum_{i=1}^{n} D(X_i) = \frac{p(1-p)}{n} = \frac{0.2496}{1500} = 166.4 \cdot 10^{-6}$$

$$\sqrt{D(\overline{X})} = \sqrt{166.4} \cdot 10^{-3} \doteq 1.29 \cdot 10^{-2}$$

Takže

$$\mathrm{norm}(\overline{X}) = \frac{\overline{X} - E(\overline{X})}{\sqrt{D(\overline{X})}} = \frac{\overline{X} - 0.52}{\sqrt{166.4}} \cdot 1000$$

a můžeme psát (pomocí úprav nerovností):

$$P(\overline{X} < 0.5) = P\Big(\frac{\overline{X} - 0.52}{\sqrt{166.4}} \cdot 1000 < \frac{0.5 - 0.52}{\sqrt{166.4}} \cdot 1000\Big) = P\Big(\text{norm}(\overline{X}) < -\frac{20}{\sqrt{166.4}}\Big) \doteq$$

$$\doteq \Phi\Big(-\frac{20}{\sqrt{166.4}}\Big) = 1 - \Phi\Big(\frac{20}{\sqrt{166.4}}\Big) \doteq 1 - \Phi\Big(1.5504\Big) \doteq 1 - 0.93948 = 0.0652$$

Tedy asi 6.52%.

O něco přesnější postup by byl tento:
Vezmeme veličinu

$$X = \sum_{i=1}^{n} X_i$$

vyjadřující počet lidí z průzkumu, kteří by volili koalici. V průzkumu se ukáže převaha opozice, pokud tato bude mít alespoň o 1 hlas více než koalice, tedy ptáme se na $P(X \leq 749)$. To odpovídá situaci kdy $P(\overline{X} \leq \frac{749}{1500})$. Opravený výpočet (s použitím $0.52 = \frac{780}{1500}$) pak dává:

$$P\Big(\overline{X} \leq \frac{749}{1500}\Big) = P\Big(\frac{\overline{X} - 0.52}{\sqrt{166.4}} \cdot 1000 \leq \frac{749/1500 - 780/1500}{\sqrt{166.4}} \cdot 1000\Big) = P\Big(\text{norm}(\overline{X}) \leq -\frac{31}{\sqrt{166.4}} \cdot \frac{2}{3}\Big) \doteq$$

$$\doteq \Phi\Big(-\frac{62}{3\sqrt{166.4}}\Big) = 1 - \Phi\Big(\frac{62}{3\sqrt{166.4}}\Big) \doteq 1 - \Phi\Big(1.6021\Big) \doteq 1 - 0.94543 = 0.05457$$

tedy asi 5, 46%, což je skoro o 1% méně než v předchozím postupu.