Pravděpodobnost, náhodné veličiny a jejich charakteristiky

—

Supplementum ke cvičení 4ST201 Statistika

Lubomír Štěpánek[1, 2]

[1]Oddělení biomedicínské statistiky
Ústav biofyziky a informatiky
1. lékařská fakulta
Univerzita Karlova, Praha

[2]Katedra biomedicínské informatiky
Fakulta biomedicínského inženýrství
České vysoké učení technické v Praze

11. října 2019

Obsah

Příklad

- Jev A značí, že náhodně vybrané přirozené číslo je dělitelné pěti. Jev B značí, že dané číslo končí nulou. Co znamenají následující jevy?
  - (i) $A \cap B$
  - (ii) $A \cup B$
  - (iii) $\overline{A} \cap B$
  - (iv) $A \cup \overline{B}$

Příklad

- Házíme běžnou šestistěnnou hrací kostkou. Jaká je pravděpodobnost, že
  - (i) při jednom hodu padne šestka?
  - (ii) při jednom hodu padne sudé číslo?
  - (iii) při dvou hodech padne dvakrát šestka?
  - (iv) při dvou hodech nepadne ani jednou šestka?
  - (v) při dvou hodech padne alespoň jednou šestka?
  - (vi) při dvou hodech padne právě jednou šestka?
  - (vii) při dvou hodech padne alespoň jednou sudé číslo?

  Kolikrát musíme minimálně hodit kostkou, aby byla pravděpodobnost, že během těchto hodů padne alespoň jedna šestka, větší než $0{,}95$?

Příklad

- V krabici je $b \in \mathbb{N}$ bílých koulí a $c \in \mathbb{N}$ černých koulí. Z krabice vytáhneme najednou tři koule.
  - S jakou pravděpodobností jsou všechny tři bílé?
  - S jakou pravděpodobností jsou všechny dvě bílé a jedna černá?

Příklad

- Z intervalu $\langle 0, 1\rangle$ náhodně vybereme dvě čísla x a y. Jaká je pravděpodobnost jevu, že $\frac{1}{2} \leq x + y \leq \frac{3}{2}$?

Příklad

- Směs sazenic je tvořena ze $80~\%$ sazenicemi první jakosti a z $20~\%$ sazenicemi druhé jakosti. Sazenice první jakosti vyklíčí s pravděpodobností $0{,}9$, sazenice druhé jakosti vyklíčí s pravděpodobností $0{,}6$. S jakou pravděpodobností vyklíčí náhodně vybraná sazenice ze směsi?

Příklad

- V nákladu 150 pytlů ořechů z Turecka je 5 pytlů se zkaženými ořechy. Podobně v nákladu 250 pytlů z Afghánistánu je 5 pytlů se zkaženými ořechy.
    - S jakou pravděpodobností obsahuje náhodně vybraný pytel zkažené ořechy, pokud oba náklady zkompletujeme do jednoho?
    - S jakou pravděpodobností obsahuje náhodně vybraný pytel zkažené ořechy, pokud nejdříve vybereme náhodně náklad a pak náhodně jeden pytel?

Příklad

- V urně je sedm bílých a tři černé koule. Náhodně vybereme tři koule.
  - S jakou pravděpodobností jsou všechny bílé, pokud jsme je vybírali tak, že byly vždy vraceny zpět do urny po každém tahu?
  - S jakou pravděpodobností jsou všechny bílé, pokud jsme je vybírali tak, že nebyly po jednotlivých tazích vraceny zpět do urny?

Příklad

- V zásilce třiceti žárovek jsou tři vadné.
  - Vybereme-li ze zásilky náhodně pět žárovek, s jakou pravděpodobností mezi nimi nebude žádná vadná?
  - Vybereme-li ze zásilky náhodně pět žárovek, s jakou pravděpodobností mezi nimi bude právě jedna vadná?
  - Kolik žárovek musíme náhodně ze zásilky vyzkoušet, aby pravděpodobnost, že objevíme alespoň jednu vadnou, byla větší než $0{,}7$?

Náhodná veličina

- je popsána číselným (nebo symbolickým) výsledkem náhodného pokusu
  - např. příjem náhodně vybrané domácnosti, doba do poruchy přístroje, počet dětí v rodině, počet porodů za noc v dané porodnici apod.
- náhodné veličiny se obvykle značí velkými písmeny z konce abecedy, tedy X, Y, Z, $\dots$, jejich konkrétní číselné realizace pak malými písmeny x, y, z, $\dots$
  - např. je-li X počet porodů za noc v dané porodnici, pak $x = 6$ značí, že zde došlo k $X = x = 6$ porodům za noc
  - X, Y, Z, $\dots$, mají tedy charakter proměnných, zatímco x, y, z, $\dots$, mají charakter konkrétních čísel (konstant)
- náhodné veličiny mohou být diskrétní, či spojité

Rozdělení náhodné veličiny

- dáno *zákonem rozdělení náhodné veličiny*
  - tj. že náhodná veličina nabude konkrétní hodnoty nebo hodnoty z nějakého intervalu s určitou pravděpodobností
- lze postupně popsat
  - řadou rozdělení
  - distribuční funkcí
  - pravděpodobnostní funkcí, resp. hustotou
  - momentovou funkcí, resp. charakteristickou funkcí
- vhodnými ukazateli jsou pak střední hodnota a variabilita náhodné veličiny

Diskrétní náhodná veličina

- je popsána diskrétním (nespojitým) číselným (nebo symbolickým) výsledkem náhodného pokusu
  - např. počet ok v rámci hodu klasickou šestistěnnou kostkou
- diskrétní náhodná veličina X tedy může nabýt každé z k hodnot $x_1, x_2, \ldots, x_k$ s pravděpodobnostmi $P(x_1), P(x_2), \ldots, P(x_k)$, respektive, pro vhodné $k \in \mathbb{N}$
  - např. pro diskrétní náhodnou veličinu X danou počtem ok v rámci hodu klasickou šestistěnnou kostkou je $k = 6$ a dále je $x_i = i$ a $P(X = x_i) = P(x_i) = \frac{1}{6}$ vždy $\forall i \in \{1, 2, \ldots, 6\}$
- počet možných hodnot k, kterých náhodná veličina může nabývat, je nanejvýš spočetný
- platí tedy

$$\sum_{i=1}^{k} P(X = x_i) = \sum_{i=1}^{k} P(x_i) = 1$$

Řada rozdělení diskrétní náhodné veličiny

- obvykle dána tabulkou typu

x_1	x_2	$\dots$	x_k
$P(x_1)$	$P(x_2)$	$\dots$	$P(x_k)$

- pro diskrétní náhodnou veličinu X danou počtem ok v rámci hodu klasickou šestistěnnou kostkou

1	2	3	4	5	6
$\frac{1}{6}$	$\frac{1}{6}$	$\frac{1}{6}$	$\frac{1}{6}$	$\frac{1}{6}$	$\frac{1}{6}$

Pravděpodobnostní funkce diskrétní náhodné veličiny

- je funkce typu $P_X : \mathbb{Z} \to \langle 0, 1 \rangle$, konkrétněji $P_X : x_i \to P(X = x_i)$
- vrací tedy pro hodnotu x_i diskrétní náhodné veličiny X pravděpodobnost $P(X = x_i) = P(x_i)$, že veličina této hodnoty nabude, kde $i \in \{1, 2, \dots, k\}$ a $k \in \mathbb{N}$
- stále platí, že

$$\sum_{i=1}^{k} P(X = x_i) = \sum_{i=1}^{k} P(x_i) = 1$$

- a také

$$P(X \in \{x_i, x_j, \dots\}) = P(\{x_i, x_j, \dots\}) = \sum_{\chi \in \{x_i, x_j, \dots\}} P(\chi)$$

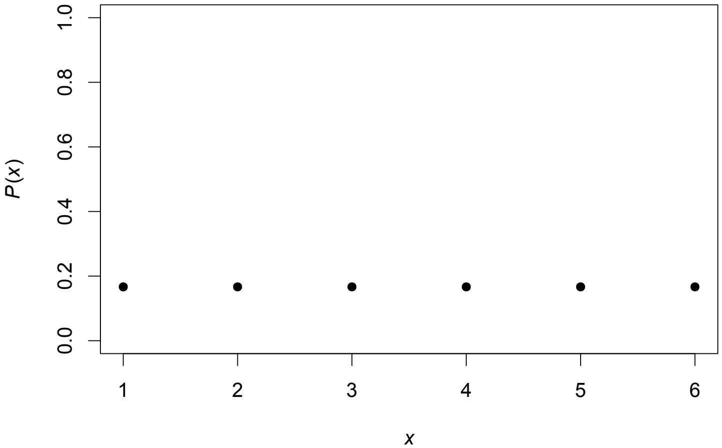
Pravděpodobnostní funkce diskrétní náhodné veličiny

- pro diskrétní náhodnou veličinu X danou počtem ok v rámci hodu klasickou šestistěnnou kostkou platí

$$P_X(i) = P(X = i) = \begin{cases} \frac{1}{6}, & i \in \{1, 2, 3, \ldots, 6\} \\ 0, & \text{jinak} \end{cases}$$

Pravděpodobnostní funkce diskrétní náhodné veličiny

- pro diskrétní náhodnou veličinu X danou počtem ok v rámci hodu klasickou šestistěnnou kostkou



Distribuční funkce (diskrétní) náhodné veličiny

- je funkce typu $F_X : \mathbb{R} \rightarrow \langle 0, 1 \rangle$, konkrétněji $F_X : x \rightarrow P(X \leq x)$
- vrací tedy pro libovolnou hodnotu $x \in \mathbb{R}$ pravděpodobnost $P(X \leq x)$, s jakou (diskrétní i spojitá) náhodná veličina X nepřevýší reálné číslo x
- píšeme tedy $F_X(x) = P(X \leq x)$

Vlastnosti distribuční funkce (diskrétní) náhodné veličiny

$\forall x \in \mathbb{R}$ je $0 \leq F_X(x) \leq 1$

$\forall x \in \mathbb{R}$ je $F_X(x) = P(X \leq x)$ a nabývá tedy pouze hodnot z intervalu $\langle 0, 1 \rangle$ $\square$

$F_X(x)$ je neklesající funkce

$\forall a, b \in \mathbb{R} : a < b$ je $F_X(a) = P(X \leq a)$ a $F_X(b) = P(X \leq b) = P(X \leq a \cup X \in \langle a, b \rangle) = P(X \leq a) + P(X \in \langle a, b \rangle) > P(X \leq a)$, tedy $F_X(x)$ je neklesající $\square$

Platí, že $P(a \leq X \leq b) = F_X(b) - F_X(a)$

$\forall a, b \in \mathbb{R} : a < b$ je $F_X(a) = P(X \leq a)$ a $F_X(b) = P(X \leq b) = P(X \leq a \cup X \in \langle a, b \rangle) = P(X \leq a) + P(X \in \langle a, b \rangle)$, tedy $F_X(b) - F_X(a) = P(X \leq a) + P(X \in \langle a, b \rangle) - P(X \leq a) = P(X \in \langle a, b \rangle)$, tudíž je $P(a \leq X \leq b) = F_X(b) - F_X(a)$ $\square$

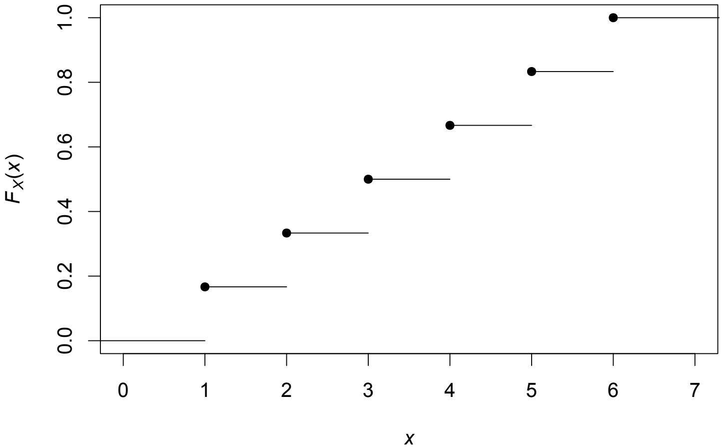
Distribuční funkce diskrétní náhodné veličiny

- pro diskrétní náhodnou veličinu X je také

$$F_X(x) = P(X \leq x) = \sum_{\forall \chi \in \{x_1, x_2, \ldots, x_k\} : \chi \leq x} P(X = \chi)$$

Distribuční funkce diskrétní náhodné veličiny

- pro diskrétní náhodnou veličinu X danou počtem ok v rámci hodu klasickou šestistěnnou kostkou



Střední hodnota diskrétní náhodné veličiny

- ať diskrétní náhodná veličina X může nabýt každé z k hodnot $x_1, x_2, \ldots, x_k$ s pravděpodobnostmi $P(x_1), P(x_2), \ldots, P(x_k)$, respektive, pro vhodné $k \in \mathbb{N}$
- pak střední hodnota $\mathbb{E}(X)$ diskrétní náhodné veličiny X je

$$\mathbb{E}(X) = \sum_{i=1}^{k} x_i P(X = x_i) = \sum_{i=1}^{k} x_i P(x_i)$$

- zřejmě dále platí
  - pro $a, b \in \mathbb{R}$ je $\mathbb{E}(aX + b) = a\mathbb{E}(X) + b$
  - pro libovolné náhodné veličiny $X_1, X_2, \ldots, X_p$ je $\mathbb{E}(X_1 + X_2 + \ldots + X_p) = \mathbb{E}(X_1) + \mathbb{E}(X_2) + \ldots + \mathbb{E}(X_p)$
  - pro nezávislé náhodné veličiny $X_1, X_2, \ldots, X_p$ je $\mathbb{E}(X_1 \cdot X_2 \cdot \ldots \cdot X_p) = \mathbb{E}(X_1) \cdot \mathbb{E}(X_2) \cdot \ldots \cdot \mathbb{E}(X_p)$

Střední hodnota diskrétní náhodné veličiny

- pro diskrétní náhodnou veličinu X danou počtem ok v rámci hodu klasickou šestistěnnou kostkou

$$
\begin{aligned}
\mathbb{E}(X) &= \sum_{i=1}^{k} x_i P(X = x_i) = \sum_{i=1}^{k} x_i P(x_i) \\
&= \sum_{i=1}^{6} i \cdot \frac{1}{6} = \frac{1}{6}(1+2+3+4+5+6) \\
&= \frac{7}{2}
\end{aligned}
$$

Rozptyl diskrétní náhodné veličiny

- ať diskrétní náhodná veličina X může nabýt každé z k hodnot $x_1, x_2, \ldots, x_k$ s pravděpodobnostmi $P(x_1), P(x_2), \ldots, P(x_k)$, respektive, pro vhodné $k \in \mathbb{N}$
- pak rozptyl $\mathrm{var}(X)$ diskrétní náhodné veličiny X je

$$\begin{aligned}\mathrm{var}(X) &= \sum_{i=1}^{k}(x_i - \mathbb{E}(X))^2 P(X = x_i) = \sum_{i=1}^{k}(x_i - \mathbb{E}(X))^2 P(x_i) \\ &= \sum_{i=1}^{k} x_i^2 P(x_i) - \left(\sum_{i=1}^{k} x_i^2 P(x_i)\right)^2\end{aligned}$$

- zřejmě dále platí
  - pro $a, b \in \mathbb{R}$ je $\mathrm{var}(aX + b) = a^2 \mathrm{var}(X)$
  - pro nezávislé náhodné veličiny $X_1, X_2, \ldots, X_p$ je $\mathrm{var}(X_1 + X_2 + \ldots + X_p) = \mathrm{var}(X_1) + \mathrm{var}(X_2) + \ldots + \mathrm{var}(X_p)$

Rozptyl diskrétní náhodné veličiny

- pro diskrétní náhodnou veličinu X danou počtem ok v rámci hodu klasickou šestistěnnou kostkou

$$
\begin{aligned}
\operatorname{var}(X) &= \sum_{i=1}^{k}(x_i - \mathbb{E}(X))^2 P(X = x_i) = \sum_{i=1}^{k}(x_i - \mathbb{E}(X))^2 P(x_i) \\
&= \sum_{i=1}^{k} x_i^2 P(x_i) - \left(\sum_{i=1}^{k} x_i^2 P(x_i)\right)^2 \\
&= \frac{1}{6}(1^2 + 2^2 + 3^2 + 4^2 + 5^2 + 6^2) - \\
&- \left(\frac{1}{6}(1 + 2 + 3 + 4 + 5 + 6)\right)^2 \\
&= \frac{105}{36}
\end{aligned}
$$

Alternativní rozdělení

- náhodná veličina X sleduje alternativní rozdělení, tedy $X \sim \mathcal{A}(p)$ s parametrem $0 \leq p \leq 1$, pokud nabývá jen dvou hodnot 0 a 1 s pravděpodobnostmi $P(X = 0) = p - 1$ a $P(X = 1) = p$
- pravděpodobnostní funkce je tedy

$$P_X(x_i) = P(X = x_i) = \begin{cases} p^{x_i}(1-p)^{1-x_i}, & x_i \in \{0, 1\} \\ 0, & \text{jinak} \end{cases}$$

- snadno nahlédneme, že

$$\mathbb{E}(X) = p$$
$$\operatorname{var}(X) = p(1-p)$$

Příklad

- Student odpoví na otázku v testu bodovanou jedním bodem správně s pravděpodobností $p = 0{,}7$. Jaká je očekávaná střední hodnota a rozptyl počtu bodů, které student za takovou otázku získá?

Binomické rozdělení

- předpokládejme, že v každém z $n \in \mathbb{N}$ nezávislých náhodných pokusů[1] může nastat úspěch s pravděpodobností p a neúspěch s pravděpodobností $1-p$, kde $0 \leq p \leq 1$
- náhodná veličina X vrací počet úspěchů, které během n pokusů nastanou
- pak náhodná veličina X sleduje binomické rozdělení, tedy $X \sim \text{binom}(n, p)$ s parametry $n \in \mathbb{N}$ a $0 \leq p \leq 1$
- pravděpodobnostní funkce, tedy pravděpodobnost, že během n nezávislých náhodných pokusů nastane úspěch právě x-krát, je

$$P_X(x) = P(X = x) = \begin{cases} \binom{n}{x} p^x (1-p)^{1-x}, & x \in \{0, 1, 2, \dots\} \\ 0, & \text{jinak} \end{cases}$$

[1] všimněme si, že jednotlivé pokusy samy o sobě sledují alternativní rozdělení

Binomické rozdělení

- pokud náhodná veličina X sleduje binomické rozdělení, tedy $X \sim \text{binom}(n, p)$, můžeme odvodit, že

$$\mathbb{E}(X) = np$$
$$\text{var}(X) = np(1-p)$$

Příklad

- Házíme desetkrát klasickou šestistěnnou kostkou. S jakou pravděpodobností padne právě čtyřikrát šestka? S jakou pravděpodobností padne nejvýše dvakrát?

Poissonovo rozdělení

- ať náhodná veličina X sleduje původně binomické rozdělení, tedy $X \sim \text{binom}(n, p)$ s parametry $n \in \mathbb{N}$ a $0 \leq p \leq 1$
- pak je-li n dostatečně velké ($n > 30$) a p dostatečně malé ($p < 0{,}1$), lze binomické rozdělení aproximovat jednoparametrickým, Poissionovým rozdělením s $\lambda = np$, tedy $X \sim \text{Po}(\lambda)$
- náhodná veličina X sledující Poissonovo rozdělení vrací počet úspěchů, které nastanou během „velkého" počtu n pokusů, přičemž dílčí pravděpodobnost p nastání jednoho každého pokusu je „malá"
- pravděpodobnostní funkce, tedy pravděpodobnost, že během n nezávislých náhodných pokusů nastane úspěch právě x-krát, je

$$P_X(x) = P(X = x) = \begin{cases} \frac{\lambda^x e^{-\lambda}}{x!}, & x \in \{0, 1, 2, \dots\} \\ 0, & \text{jinak} \end{cases}$$

Poissonovo rozdělení

- pokud náhodná veličina X sleduje Poissonovo rozdělení, tedy $X \sim \text{Po}(\lambda)$, můžeme odvodit, že

$$\mathbb{E}(X) = \lambda$$
$$\text{var}(X) = \lambda$$

Příklad

- Ve vybraném periodiku se objevují průměrně čtyři překlepy na každých jeho deset stránek. S jakou pravděpodobností bude na náhodně vybrané stránce periodika
  (i) žádný překlep?
  (ii) jeden překlep?
  (iii) dva překlepy?
  (iv) více než dva překlepy?

Hypergeometrické rozdělení

- předpokládejme, že uvažujeme sestavu N prvků, kde M prvků z nich má určitou vlastnost a $N-M$ prvků tuto vlastnost nemá
- ze sestavy těchto prvků postupně vybereme n prvků bez vracení (!)
- náhodná veličina X, která vrací počet prvků s určitou vlastností mezi n prvky, sleduje hypergeometrické rozdělení, $X \sim \mathrm{hy}(N, M, n)$
- pravděpodobnostní funkce, tedy pravděpodobnost, že mezi n náhodně bez vracení vybranými prvky bude x-krát prvek s určitou vlastností, je

$$P(X = x) = \begin{cases} \frac{\binom{M}{x}\binom{N-M}{n-x}}{\binom{N}{n}}, & \max\{n - N + M, 0\} \leq x \leq \min\{M, n\} \\ 0, & \text{jinak} \end{cases}$$

Hypergeometrické rozdělení

- pokud náhodná veličina X sleduje hypergeometrické rozdělení, tedy $X \sim \text{hy}(N, M, n)$, můžeme odvodit, že

$$\mathbb{E}(X) = n\frac{M}{N}$$
$$\text{var}(X) = n\frac{M}{N}\left(1 - \frac{M}{N}\right)\frac{N-n}{N-1}$$

Příklad

- V sérii po dvě stě kusech je deset zmetků. Při přejímce náhodně vybereme pět kusů a podrobíme je destrukční zkoušce. Pokud není mezi pěti vybranými ani jeden zmetek, sérii přijmeme. S jakou pravděpodobností to nastane?

Literatura

- Hindls, Richard, Stanislava Hronová, Jan Seger a Jakub Fischer. *Statistika pro ekonomy*. Praha: Professional Publishing, 2007. ISBN: 978-80-86946-43-6.

- Marek, Luboš. *Statistika v příkladech*. Praha: Professional Publishing, 2015. ISBN: 978-80-7431-153-6.

Děkuji za pozornost!

lubomir.stepanek@vse.cz
lubomir.stepanek@lf1.cuni.cz
lubomir.stepanek@fbmi.cvut.cz