5. cvičení z PST

14. - 18. března 2022

Poznámky k binomickému rozdělení: Mějme n nezávislých opakování daného pokusu, jehož úspěšnost je $0 < p < 1$. Veličina

$$X = \text{“}\textit{počet úspěchů během } n \textit{ pokusů}\text{”}$$

s hodnotami $X \in \{0, 1, \ldots, n\}$ má pak tzv. *binomické* rozdělení $\mathrm{Bi}(n, p)$. Pro $k = 0, 1, \ldots, n$ pak je

$$P(X = k) = \binom{n}{k} p^k \cdot (1-p)^{n-k}$$

což je jeden z členů v binomické větě (odtud také ten název): $(a+b)^n = \sum_{k=0}^{n} \binom{n}{k} a^k \cdot b^{n-k}$

Důvod, proč tu máme kombinační číslo, je v tom, že daných k úspěchů je rozmístěno mezi n pokusů právě $\binom{n}{k}$ způsoby.

Příklad 5.1 *Pravděpodobnost, že atlet v oddíle skočí do dálky přes* $5\,m$*, je* 0.7*. V oddíle je* 6 *atletů.*

(a) Určete rozdělení náhodné veličiny

$$X = \begin{cases} 1, & \textit{atlet skočil přes } 5\,m \\ 0, & \textit{atlet neskočil přes } 5\,m, \end{cases}$$

její střední hodnotu $E(X)$ *a rozptyl* $\mathrm{var}(X)$.

(b) Určete rozdělení náhodné veličiny

$$Y = \text{“ }\textit{počet atletů v oddíle, kteří skočili přes } 5\,m\text{ ”}$$

její střední hodnotu $E(Y)$ *a rozptyl* $\mathrm{var}(Y)$.

(c) Jaká je pravděpodobnost, že přes $5\,m$ *skočí v oddíle alespoň* 4 *atleti?*

Řešení:

(a) Veličina X má *alternativní* rozdělení, tj. $X \sim \mathrm{Alt}(p)$, kde $p = P(X = 1) = 0.7$ je pravděpodobnost úspěšného pokusu a $P(X = 0) = 1 - p = 0.3$.

A dále máme

$$E(X) = \sum_{i \in \mathbb{R}} i \cdot P(X = i) = 0 \cdot (1-p) + 1 \cdot p = p = 0.7$$

$$E(X^2) = \sum_{i \in \mathbb{R}} i^2 \cdot P(X = i) = 0^2 \cdot (1-p) + 1^2 \cdot p = p = 0.7$$

a

$$\mathrm{var}(X) = E(X^2) - \big(E(X)\big)^2 = p - p^2 = p(1-p) = 0.7 \cdot 0.3 = 0.21\ .$$

(b) Veličina Y představuje počet úspěchů při $n = 6$ nezávislých pokusech, s pravděpodobností úspěchu $p = 0.7$, takže Y má *binomické rozdělení* $\text{Binom}(n, p)$. Hodnoty veličiny Y jsou $k = 0, 1, \ldots, n$ a jejich pravděpodobnosti jsou

$$P(Y = k) = \binom{n}{k} p^k (1-p)^{n-k} = \binom{6}{k} 0.7^k \cdot 0.3^{6-k} \ .$$

Pro další výpočty se hodí všimnout si, že $Y = \sum_{i=1}^{n} X_i$ kde

$$X_i = \begin{cases} 1, & i\text{-tý atlet skočil přes } 5\,\text{m} \\ 0, & i\text{-tý atlet neskočil přes } 5\,\text{m}, \end{cases}$$

jsou navzájem nezávislé náhodné veličiny a $X_i \sim \text{Alt}(p)$.

Pro střední hodnotu veličiny Y pak máme

$$E(Y) = E\left(\sum_{i=1}^{n} X_i\right) = \sum_{i=1}^{n} \underbrace{E(X_i)}_{p} = n \cdot p = 6 \cdot 0.7 = 4.2$$

a z nezávislosti X_i pak pro rozptyl máme

$$\text{var}(Y) = \sum_{i=1}^{n} \underbrace{\text{var}(X_i)}_{p(1-p)} = np(1-p) = 6 \cdot 0.7 \cdot 0.3 = 1.26 \ .$$

(c)

$$P(Y \geq 4) = \sum_{k=4}^{6} \binom{6}{k} 0.7^k \cdot 0.3^{6-k} = 15 \cdot 0.7^4 \cdot 0.3^2 + 6 \cdot 0.7^5 \cdot 0.3^1 + 1 \cdot 0.7^6 \cdot 1 =$$
$$= 0.324135 + 0.302526 + 0.117649 = 0.74431 \ .$$

Připomenutí: Veličina

$$X = \textit{“počet neúspěchů než nastane první úspěch”}$$

má geometrické rozdělení $\text{Geom}(p)$, pokud můžeme opakovat libovolné množství nezávislých pokusů, které mají všechny stejnou pravděpodobnost úspěchu p. Příkladem je třeba situace, že se chceme trefit míčem do koše apod.

Hodnoty veličiny X jsou $\{0, 1, 2, \ldots\}$. Pro odvození rozdělení X si pro $i = 1, 2, 3\ldots$ označme jevy

$$A_i = \textit{“i-tý pokus je úspěšný”}$$

které budou nezávislé s budou mít pravděpodobnosti $P(A_i) = p$. Pak máme pravděpodobnosti

$$P(X = k) = P(A_1^c \cap \cdots \cap A_k^c \cap A_{k+1}) = P(A_1^c) \cdots P(A_k^c) \cdot P(A_{k+1}) = (1-p)^k p$$

pro $k = 0, 1, 2, \ldots$ Pro střední hodnotu a rozptyl pak máme:

$$E(X) = \sum_{k=0}^{\infty} k \cdot P(X = k) = \sum_{k=0}^{\infty} kp(1-p)^k = \cdots = \tfrac{1-p}{p}$$

$$\text{var}(X) = \frac{1-p}{p^2}$$

což si můžeme lépe zapamatovat jako

$$\text{var}(X) = \frac{1}{p} \cdot E(X) \ .$$

Příklad 5.2 *Pravděpodobnost narození chlapce je* 0.51. *Jaká je pravděpodobnost, že v dané porodnici dnes bylo nejpozději (v časovém pořadí) čtvrté narozené dítě holka?*

Řešení:
Lze použít dva přístupy:

(a) Vezmeme si náhodnou veličinu

$$X = \text{“ } \textit{počet narozených chlapců před první narozenou holkou.”}$$

Ta má *geometrické* rozdělení Geom(p) s pravděpodobností $p = 1 - 0.51 = 0.49$ úspěšného pokusu (tj. narození holky). Hodnoty veličiny X jsou $k = 0, 1, 2, \dots$ s pravděpodobnostmi

$$P(X = k) = (1-p)^k \cdot p = 0.51^k \cdot 0.49$$

Pro jednodušší výpočet si ještě označme $q = 1 - p = 0.51$. Hledaná pravděpodobnost je tedy

$$P(X < 4) = \sum_{k=0}^{3} q^k \cdot (1-q) = (1-q) \cdot \underbrace{(1 + q + \cdots + q^3)}_{\frac{1-q^4}{1-q}} = 1 - q^4 = 1 - 0.51^4 \doteq 0.9323.$$

(b) Vezmeme si náhodnou veličinu

$$Y = \text{“ } \textit{počet narozených holek mezi prvními } 4 \textit{ narozenými dětmi.”}$$

Ta má binomické rozdělení Bi$(4, p)$, kde je opět $p = 0.49$. Hledaná pravděpodobnost je tedy

$$P(Y \geq 1) = 1 - P(Y = 0) = 1 - \binom{4}{0} p^0 \cdot (1-p)^4 = 1 - q^4 = 1 - 0.51^4.$$

Příklad 5.3 *Při hodu na koš se trefíme s pravděpodobností* $p = 0.2$. *Náhodná veličina* X *je počet hodů, než se trefíme.*

(a) Určete rozdělení veličiny X.

(b) V kterém kole se průměrně poprvé trefíme?

(c) Kolikrát musíme nejméně hodit, abychom se s pravděpodobností alespoň 90% *alespoň jednou (v rámci těchto hodů) trefili?*

Řešení:
(a) Veličina X počítá počet neúspěšných kol. Má tedy geometrické rozdělení Geom(0.2) s oborem hodnot $\{0, 1, 2, \dots\}$ a jejich pravděpodobnostmi

$$P(X = k) = p(1-p)^k = 0.2 \cdot 0.8^k$$

pro $k = 0, 1, 2, \ldots$.

POZOR: Neplețte si *geometrickou pravděpodobnost*, což je způsob počítání pravděpodobnosti jevu (název je odvozen od geometrických obrazců) a *geometrické rozdělení*, které zase přísluší náhodné veličině (název je odvozen od geometrické posloupnosti, kterou tvoří hodnoty pravděpodobnostní funkce dané veličiny).

(b) Hledáme střední hodnotu veličiny

$$Y = \text{"}\textit{pořadí hodu, při kterém se poprvé trefíme}\text{"} .$$

Zřejmě je $Y = X + 1$ a tedy

$$E(Y) = E(X+1) = E(X) + 1 = \frac{1-p}{p} + 1 = \frac{1}{p} = \frac{1}{0.2} = 5 .$$

(c) Abychom si lépe představili situaci, předpokládejme, že i po té, co se trefíme, pokračujeme v házení (a veličina Y zaznamená pouze ten první úspěšný pokus). Ptáme se tedy, jaký je nejmenší počet hodů $n \in \mathbb{N}$, abychom se v jejich průběhu alespoň jednou trefili. Chceme tudíž znát nejmenší $n \in \mathbb{N}$ takové, že $P\Big(Y \leq n\Big) \geq 0.9$, neboli

$$0.9 \leq P\Big(X + 1 \leq n\Big) = P\Big(X \leq n-1\Big) = F_X(n-1) .$$

K tomu potřebujeme tudíž znát distribuční funkci X pro $k \in \mathbb{N}_0$:

$$F_X(k) = \sum_{i \leq k} P(X = i) = \sum_{i=0}^{k} p(1-p)^i = p \cdot \frac{1-(1-p)^{k+1}}{1-(1-p)} = 1 - (1-p)^{k+1} .$$

Po dosazení tedy dostaneme podmínku

$$0.9 \leq F_X(n-1) = 1 - (1-p)^n$$

neboli

$$\begin{aligned} 0.1 &\geq (1-p)^n = (1-0.2)^n = 0.8^n \\ \log 0.1 &\geq n \log 0.8 \\ n &\geq \tfrac{\log 0.1}{\log 0.8} \doteq 10.32 \end{aligned}$$

Pozor: logaritmus hodnoty 0.8 je záporný! Tedy musíme hodit alespoň $n = 11$ -krát.

- -

Úlohu (c) můžeme vyřešit i s pomocí binomického rozdělení. Pro $n \in \mathbb{N}$ uvažujme veličinu

$$Z_n = \text{"}\textit{počet tolika hodů (z n možných), ve kterých se trefíme}\text{"}$$

která zřejmě má binomické rozdělení $\mathrm{Bi}(n, p)$. Hledáme nyní nejmenší $n \in \mathbb{N}$ tak, aby

$$P(Z_n \geq 1) \geq 0.9 .$$

Máme tedy

$$0.9 \leq P(Z_n \geq 1) = 1 - P(Z_n = 0) = 1 - (1-p)^n$$

což je stejná nerovnost jako výše a tím dostaneme i stejné řešení.