5. cvičení z STP

18. března 2021

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

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

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

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

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

(c)

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

$$= 0.324135 + 0.302526 + 0.117649 = 0.74431 \ .$$

Příklad 5.2 *Semena mají klíčivost $p \in (0, 1)$. Jaký je optimální počet n semen v jamce, aby byla co nejvyšší pravděpodobnost, že vyklíčí právě jedno? Řešte obecně a pro $p = 1/3$.*

Řešení:
Vyklíčení jednotlivých semen pokládáme za nezávislé jevy, takže veličina

$$X = \text{“}\textit{počet vyklíčených semen z } n \textit{ semen v jamce}\text{”}$$

má binomické rozdělení $\mathrm{Bi}(n, p)$. Hledáme teď $n \in \mathbb{N}$ ($n \geq 1$), které maximalizuje funkci

$$g(n) = P(X = 1) = \tbinom{n}{1}\, p^1\, (1-p)^{n-1} = n\, p\, (1-p)^{n-1} \left(= np \cdot e^{(n-1)\ln(1-p)}\right) \ .$$

Pro vyšetření této funkce můžeme uvažovat n jako reálnou proměnnou v intervalu $(0, +\infty)$, abychom pak mohli využít derivaci (podle n) a to především pro zjištění, kde je funkce g rostoucí a kde klesající:

$$g'(n) = p\,(1-p)^{n-1} + n\, p\,(1-p)^{n-1}\, \ln(1-p) = p\,(1-p)^{n-1}\big(1 + n\, \ln(1-p)\big) \ .$$

Dostáváme tak, že g je rostoucí až do bodu $n = \frac{-1}{\ln(1-p)}$ a pak je klesající. V uvedeném bodě tak nastává maximum v rámci reálné proměnné. Maximum v oboru přirozených čísel $\mathbb{N}$ nastává pro jedno (nebo obě) ze dvou celých čísel, která jsou nejblíže hodnotě $\frac{-1}{\ln(1-p)}$. Zjistit, které z těchto dvou čísel to vlastně obecně je, ale už dá více práce.

Pro $p = \frac{1}{3}$ máme $\frac{-1}{\ln(1-p)} = \frac{-1}{\ln(2/3)} \doteq 2.466$. Nejbližší čísla jsou tedy 2 a 3 a v nich stačí porovnat hodnoty funkce g:

$$g(2) = 2 \cdot \tfrac{1}{3} \left(\tfrac{2}{3}\right)^1 = \tfrac{4}{9}$$

$$g(3) = 3 \cdot \tfrac{1}{3} \left(\tfrac{2}{3}\right)^2 = \tfrac{4}{9}$$

Obě možnosti 2 a 3 tedy představují optimální počty semen pro $p = \frac{1}{3}$.

Poznámka: *Pro $p \to 0$ je $\ln(1-p) \approx -p$ a $n \approx 1/p$, což je v souladu s očekáváním. Pro $p \to 1$ vychází $n \to 0$, což vypadá překvapivě. Znamená to ale jen, že funkce g je na množině přirozených čísel klesající a maxima nabývá v 1.*

- -

Zde je ještě (pro zájemce) podrobnější a úplné řešení:
Příklad se dá řešit ještě jinak a (překvapivě) i jednodušeji (ale zápis je delší). Zřejmě pro $n \in \mathbb{N}$ máme

$$g(n) \leq g(n+1) \quad \Leftrightarrow \quad 1 \leq \frac{g(n+1)}{g(n)} = \frac{(n+1)\,p\,(1-p)^n}{n\,p\,(1-p)^{n-1}} = \frac{n+1}{n}(1-p)$$

což po úpravě dává

$$\frac{1}{1-p} \leq \frac{n+1}{n} = 1 + \frac{1}{n}$$
$$\underbrace{\frac{1}{1-p} - 1}_{\frac{p}{1-p}} \leq \frac{1}{n}$$

a nakonec

$$n \leq \frac{1-p}{p} = \frac{1}{p} - 1 \,.$$

Speciálně vidíme, že to samé platí i pro ostré nerovnosti, tj.

∗ $g(n) < g(n+1)$ platí právě když $n < \frac{1}{p} - 1$.

∗ $g(n) > g(n+1)$ platí právě když $n > \frac{1}{p} - 1$.

∗ $g(n) = g(n+1)$ platí právě když $n = \frac{1}{p} - 1$.

Odsud ihned máme, že

- pro $\frac{1}{p} - 1 < 1$, tj. $p \in \left(\frac{1}{2}, 1\right)$ je g ostře klesající (na $\mathbb{N}$), takže maximum nastává pro $n_0 = 1$ a pravděpodobnost pak je $P(X=1) = g(1) = p$.
- pokud $\frac{1}{p} \notin \mathbb{N}$ a $p \in \left(0, \frac{1}{2}\right)$, budou všechny nerovnosti mezi hodnotami funkce g ostré a největší hodnota bude dosažena pro $n_0 = \left[\frac{1}{p}\right]$ (tj. celá část z $\frac{1}{p}$).

  Je to proto, že z $n_0 - 1 < \frac{1}{p} - 1$ plyne $g(n_0 - 1) < g(n_0)$ a současně z $\frac{1}{p} - 1 < n_0$ plyne $g(n_0) > g(n_0 + 1)$.

  Navíc ještě z nerovnosti $n_0 < \frac{1}{p} < n_0 + 1$ dostaneme odhad pravděpodobnosti

$$\underbrace{\left(\frac{1}{p} - 1\right) p(1-p)^{\frac{1}{p}-1}}_{(1-p)^{\frac{1}{p}}} \;<\; \underbrace{n_0 p(1-p)^{n_0-1}}_{P(X=1)} \;<\; \underbrace{\frac{1}{p}\, p(1-p)^{\frac{1}{p}-2}}_{(1-p)^{\frac{1}{p}-2}}$$

- a konečně pro $n_0 = \frac{1}{p} \in \mathbb{N}$ budou všechny nerovnosti mezi hodnotami $g(n)$ také ostré až na případ $g(n_0 - 1) = g(n_0)$, který odpovídá maximu funkce g na přirozených číslech.

  V tomto případě totiž platí, že $n_0 - 1 = \frac{1}{p} - 1$, tedy skutečně $g(n_0 - 1) = g(n_0)$.

  Tedy maximum nastává pro dva počty semen $n_0 - 1$ a n_0 a pravděpodobnost vyklíčení právě jednoho semene je

$$P(X=1) = \frac{1}{p} p(1-p)^{\frac{1}{p}-1} = (1-p)^{\frac{1}{p}-1}$$

Mimo jiné z tohoto všeho vidíme, že pro $p \to 0$ se pravděpodobnost vyklíčení jednoho semene (při optimálním počtu semen) blíží k $e^{-1} \doteq 0.3679$ (protože $\lim_{p \to 0+} (1-p)^{\frac{1}{p}} = \lim_{p \to 0+} e^{\frac{\ln(1-p)}{p}} = e^{-1}$).

Pro konkretní volbu $p = \frac{1}{3}$ máme $\frac{1}{p} = 3 \in \mathbb{N}$ (tedy třetí případ) a tak opět dostáváme, že: optimální počet semen je $n \in \left\{\frac{1}{p} - 1, \frac{1}{p}\right\} = \{2, 3\}$ a pravděpodobnost bude $P(X = 1) = (1-p)^{\frac{1}{p}-1} = \frac{4}{9}$.

Připomenutí: Veličina

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

má geometrické rozdělení 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 $\{0, 1, 2, \dots\}$. Pro odvození rozdělení X si pro $i = 1, 2, 3 \dots$ označme jevy

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

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

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

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

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

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

což si můžeme lépe zapamatovat jako

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

Příklad 5.3 *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

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

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

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

(b) Vezmeme si náhodnou veličinu

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

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

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