6. cvičení z PST

24. - 28. října 2022

Poznámka: Pro *libovolnou* veličinu X a její distribuční funkci F_X si můžeme obecně definovat funkci

$$p_X(t) := F_X(t) - F_X(t_-) \;\Big(= P(X = t)\Big) \quad \text{pro všechna } t \in \mathbb{R}$$

(kde $F_X(t_-)$ je limita zleva funkce F_X v bodě t) a funkci

$$f_X(t) := \begin{cases} \frac{\mathrm{d}}{\mathrm{d}t} F_X(t) & \text{, pro taková } t \in \mathbb{R}, \text{ kde konečná derivace existuje} \\ 0 & \text{, jinak.} \end{cases}$$

Nyní platí toto: Jestliže rozdělení X je

- *diskrétní* $\Rightarrow$ p_X je tzv. *pravděpodobnostní funkce pro* X, pro níž je $\sum_{t\in\mathbb{R}} p_X(t) = 1$, a $f_X(t) = 0$ pro všechna $t \in \mathbb{R}$,
- *(absolutně) spojité* $\Rightarrow$ f_X je její hustota pravděpodobnosti a $p_X(t) = 0$ pro všechna $t \in \mathbb{R}$.

Mějme veličinu $X : \Omega \to \mathbb{R}$ na Kolmogorově modelu s pravděpodobností P a nechť X je směs veličin

$$X = \mathrm{Mix}_c(D, S)$$

kde D má diskrétní rozdělení a S má (absolutně) spojité rozdělení. Pak platí

$$F_X(t) = c \cdot F_D(t) + (1 - c) \cdot F_S(t) \quad \text{pro všechna } \; t \in \mathbb{R}$$

a odsud máme

$$P(X = t) = p_X(t) = c \cdot p_D(t) + (1 - c) \cdot \underbrace{p_S(t)}_{=0} \quad \Rightarrow \quad p_D(t) = \frac{P(X = t)}{c}$$

a také

$$\sum_{t\in\mathbb{R}} P(X = t) = c \underbrace{\sum_{t\in\mathbb{R}} p_D(t)}_{=1} \quad \Rightarrow \quad c = \sum_{t\in\mathbb{R}} P(X = t)$$

a konečně také

$$f_X(t) = c \cdot \underbrace{f_D(t)}_{=0} + (1 - c) \cdot f_S(t) \quad \Rightarrow \quad f_S(t) = \begin{cases} \frac{F_X'(t)}{1-c} & \text{, pro taková } t \in \mathbb{R}, \text{ kde konečná derivace existuje} \\ 0 & \text{, jinak.} \end{cases}$$

Poznamenejme, že pokud existuje dělení $-\infty = t_0 < t_1 < \cdots < t_n = \infty$ celé přímky $\mathbb{R}$ tak, aby na otevřených intervalech (t_i, t_{i+1}) měla F_X spojitou derivaci, pak veličina S bude mít hustotu a výše odvozená funkce f_S je příslušná hustota S.

Příklad 6.1 *(rozklad na směs)*
Náhodná veličina X má distribuční funkci

$$F_X(x) = \begin{cases} 0 & , x \in (-\infty, -2) \\ 0.1\,x + 0.5 & , x \in \langle -2, -1) \\ 0.4 & , x \in \langle -1, 0) \\ 0.3\,x + 0.6 & , x \in \langle 0, 1) \\ 1 & , x \in \langle 1, \infty) \end{cases}$$

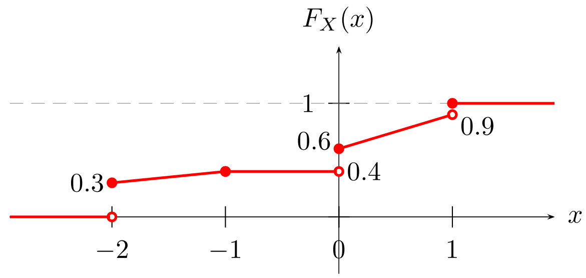
(a) *Vyjádřete X jako směs náhodných veličin D a S, z nichž D je diskrétní a S spojitá; popište a znázorněte jejich rozdělení.*

(b) *Určete $E(X)$.*

(c) *Najděte kvantilovou funkci q_X.*

Řešení:

(a) Znázorníme si graf F_X:



Připomeňme si, jak se hledá diskrétní a spojitá části směsi $X = \mathrm{Mix}_c(D, S)$. Diskrétní část D je zodpovědná za skoky v distribuční funkci F_X. Pro koeficient c ve směsi platí

$$c = P(X \in \underbrace{\text{"\textit{množina všech bodů nespojitosti funkce} } F_X\text{"}}_{=\{-2,0,1\}}) =$$

$$= P\Big(X \in \{-2, 0, 1\}\Big) = P(X = -2) + P(X = 0) + P(X = 1) \ .$$

Hodnotu skoku dostaneme pomocí vztahu

$$P(X = -2) = P(X \leq -2) - P(X < -2) = F_X(-2) - \lim_{t \to -2_-} F_X(t) = 0.3 - 0 = 0.3$$

a podobně

$$P(X = 0) = 0.6 - 0.4 = 0.2 \ \text{ a } \ P(X = 1) = 1 - 0.9 = 0.1$$

Takže

$$c = P(X = -2) + P(X = 0) + P(X = 1) = 0.3 + 0.2 + 0.1 = \mathbf{0.6}.$$

Pro distribuční funkce máme vztah

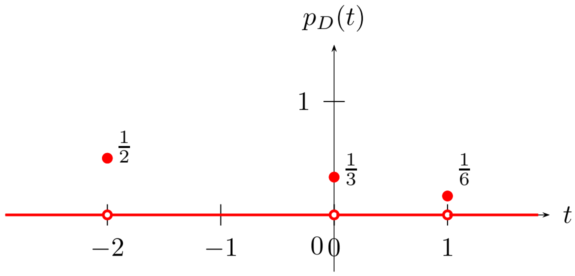
$$F_X(t) = cF_D(t) + (1 - c)F_S(t)$$

• **Popis diskrétní části D:**

Pro popis diskrétní části D stačí znát pravděpodobnosti jejich hodnot, tedy její pravděpodobnostní funkci p_D:

$$p_D(t) = \tfrac{1}{c} \cdot P(X = t) = \tfrac{1}{0.6} \cdot P(X = t) = \begin{cases} \frac{1}{0.6} \cdot 0.3 = \frac{1}{2} & , \ t = -2 \\ \frac{1}{0.6} \cdot 0.2 = \frac{1}{3} & , \ t = 0 \\ \frac{1}{0.6} \cdot 0.1 = \frac{1}{6} & , \ t = 1 \\ 0 & , \ \text{jinak} \end{cases}$$

s grafem:



Dále můžeme ještě pro diskrétní veličinu D zjistit její distribuční funkci:

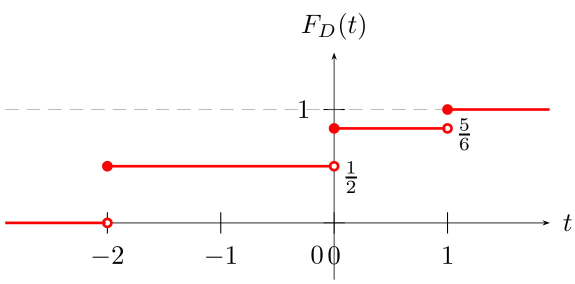
Platí

$$cF_D(t) = \sum_{a \le t} P(X = a) = \begin{cases} 0 & , \ t < -2 \\ 0.3 & , \ t \in \langle -2, 0) \\ 0.3 + 0.2 = 0.5 & , \ t \in \langle 0, 1) \\ 0.3 + 0.2 + 0.1 = 0.6 & , \ t \ge 1 \end{cases}$$

takže pro $\frac{1}{c} = \frac{1}{0.6}$ máme

$$F_D(t) = \begin{cases} 0 & , \ t < -2 \\ \frac{1}{0.6} \cdot 0.3 = \frac{1}{2} & , \ t \in \langle -2, 0) \\ \frac{1}{0.6} \cdot 0.5 = \frac{5}{6} & , \ t \in \langle 0, 1) \\ \frac{1}{0.6} \cdot 0.6 = 1 & , \ t \ge 1 \end{cases}$$

a graf F_D je:



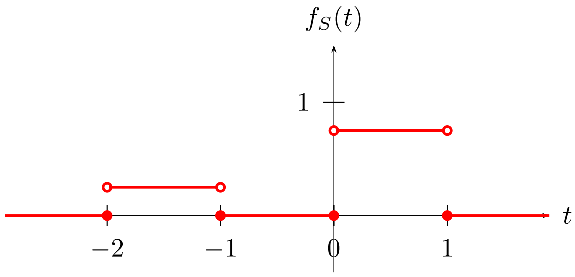
• **Popis spojité části S:**

Pro popis (absolutně) spojité části S stačí znát její hustotu pravděpodobnosti f_S. Tu získáme derivací F_S pro body, kde derivace existuje. V ostatních (v tomto případě konečně mnoha bodech) na (nezáporných) hodnotách nezáleží. Takže máme:

$$f_S(t) = F_S'(t) = \left(\frac{F_X - cF_D}{1 - c} \right)'(t) = \frac{1}{1 - c} \cdot F_X'(t)$$

$$f_S(t) = \begin{cases} \frac{1}{4} & , \ t \in (-2, -1) \\ \frac{3}{4} & , \ t \in (0, 1) \\ 0 & , \ \text{jinak.} \end{cases}$$

Graf f_S pak bude:

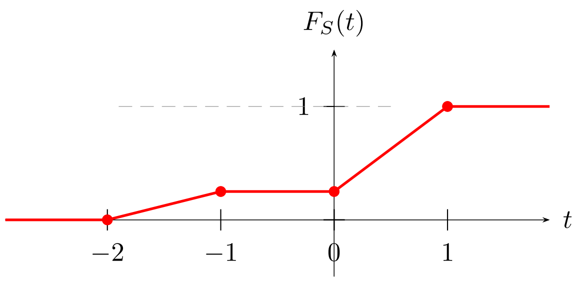


Dále můžeme ještě pro spojitou veličinu S zjistit její distribuční funkci:

Ze vztahu $F_X = cF_D + (1-c)F_S$ dostáváme

$$F_S(t) = \frac{F_X(t) - cF_D(t)}{1-c} = \tfrac{1}{0.4}\Big(F_X(t) - cF_D(t)\Big) = \begin{cases} & 0 \quad , \ t < -1 \\ \tfrac{1}{0.4}\Big(0.1\ t + 0.5 - 0.3\Big) = & \tfrac{t+2}{4} \quad , \ t \in \langle -2, -1) \\ \tfrac{1}{0.4}\Big(0.4 - 0.3\Big) = & \tfrac{1}{4} \quad , \ t \in \langle -1, 0) \\ \tfrac{1}{0.4}(0.3\ t + 0.6 - 0.5) = & \tfrac{3t+1}{4} \quad , \ t \in \langle 0, 1) \\ \tfrac{1}{0.4}(1 - 0.6) = & 1 \quad , \ t \geq 1 \end{cases}$$

Graf funkce F_S tedy dostaneme jednoduše tak, že části grafu F_X, které na sebe nenavazují, posuneme dolů tak, aby výsledek byl spojitý. Celý graf pak ve směru y natáhneme tak, aby v nekonečnu měl limitu rovnou 1:



(b) Střední hodnotu $E(X)$ můžeme spočítat ze vztahu

$$E(X) = cE(D) + (1-c)E(S)\ .$$

Pro diskrétní část je

$$E(D) = \sum_{t \in \mathbb{R}} t \cdot p_D(t) = (-2) \cdot \frac{1}{2} + 0 \cdot \frac{1}{3} + 1 \cdot \frac{1}{6} = -\frac{5}{6}$$

a pro spojitou část je

$$E(S) = \int_{-\infty}^{\infty} t \cdot f_S(t)\ \mathrm{d}t = \int_{-2}^{-1} t \cdot \frac{1}{4}\mathrm{d}t + \int_{0}^{1} t \cdot \frac{3}{4}\mathrm{d}t = \left[\tfrac{t^2}{8}\right]_{-2}^{-1} + \left[\tfrac{3t^2}{8}\right]_{0}^{1} = \frac{1}{8} - \frac{1}{4} - \frac{3}{8} = 0$$

takže

$$E(X) = 0.6 \cdot E(D) + 0.4 \cdot E(S) = 0.6 \cdot \left(-\frac{5}{6}\right) = -0.5\ .$$

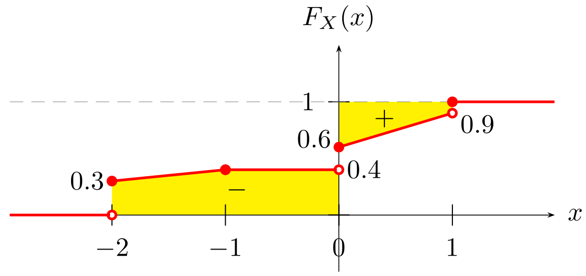
Střední hodnotu $E(X)$ můžeme také spočítat pomocí F_X jako

$$E(X) = -\int_{-\infty}^{0} F_X(x)\ \mathrm{d}x + \int_{0}^{\infty} (1 - F_X(x))\ \mathrm{d}x =$$

$$= -\int_{-2}^{-1} (0.1\,x + 0.5)\ \mathrm{d}x - \int_{-1}^{0} 0.4\ \mathrm{d}x + \int_{0}^{1} (-0.3\,x + 0.4)\ \mathrm{d}x =$$

$$= \underbrace{-\left[\tfrac{0.1\,x^2}{2}\right]_{-2}^{-1}}_{0.15} -0.5 - 0.4 + \underbrace{\left[-\tfrac{0.3\,x^2}{2}\right]_{0}^{1}}_{-0.15} +0.4 = -0.5\ .$$

Geometrický význam střední hodnoty $E(X)$ je rozdíl velikosti plochy nad grafem F_X v intervalu $(0, +\infty)$ a plochy pod grafem funkce F_X v intervalu $(-\infty, 0)$ (viz žluté plochy se znaménky v obrázku).



(c) Kvantilová funkce q_X pro veličinu X je v jistém smyslu "inverzí" k F_X. Je definována jako

$$q_X(\alpha) := \frac{1}{2}\Big(\sup\{t \in \mathbb{R} \mid F_X(t) \le \alpha\} + \inf\{t \in \mathbb{R} \mid F_X(t) \ge \alpha\}\Big) \quad \text{pro} \quad \alpha \in (0, 1).$$

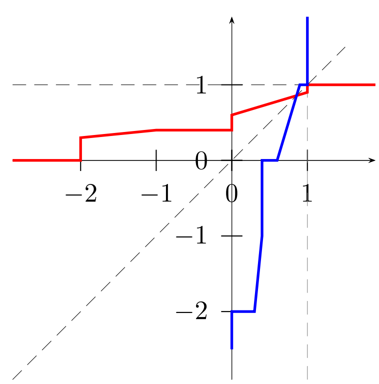
Tato definice vypadá poněkud složitě, ale je to jen kvůli případným úsekům, kde funkce F_X je konstantní. Vždy ale platí toto:

Věta: *Kvantilová funkce q_X je inverzní funkce k distribuční funkci F_X tam, kde F_X je na otevřeném intervalu $I \subseteq \mathbb{R}$ spojitá a ostře rostoucí.*

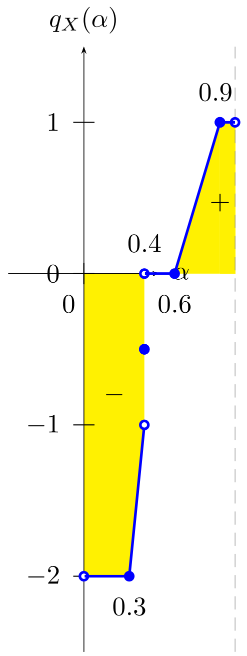
Obecněji pak platí následující: graf kvantilové funkce q_X získáme z grafu distribuční funkce F_X takto

- graf F_X doplníme na "souvislou čáru", tj. skoky funkce F_X nahradíme spojitou svislou úsečkou,
- tento útvar převrátíme podle osy 1. a 3. kvadrantu (tj. podle přímky "$x = y$"),
- tam, kde převrácený útvar není funkcí (tj. obsahuje svislé čáry) tyto úseky odstraníme a nahradíme jedinou hodnotou, a sice průměrem limit zprava a zleva (případné krajní úseky v bodech $\alpha = 0$ a $\alpha = 1$ odstraníme úplně, protože tam se kvantil nedefinuje).

V našem případě tedy graf F_X přejde na



a dostaneme graf q_X:



Kvantilovou funkci určíme také explicitně. Kvantil q_X je inverzní funkcí k tam, kde F_X je ostře rostoucí funkce, tj. pak $q_X(\alpha) = x$ právě když $F_X(x) = \alpha$. Tedy pro $x \in (-2, -1)$ je $F_X(x) = 0.1\,x + 0.5$ a tudíž

$$\alpha = 0.1\,x + 0.5 \quad \Leftrightarrow \quad x = 10\alpha - 5$$

a tudíž

$$q_X(\alpha) = 10\alpha - 5 \quad \text{pro} \quad \alpha \in (0,\ 0.3)$$

Podobně pro $x \in (0, 1)$ je $F_X(x) = 0.3\,x + 0.6$. Tedy

$$\alpha = 0.3\,x + 0.6 \quad \Leftrightarrow \quad x = \frac{10}{3}\alpha - 2$$

a tudíž

$$q_X(\alpha) = \frac{10}{3}\alpha - 2 \quad \text{pro} \quad \alpha \in (0.6,\ 0.9)\ .$$

Celkově:

$$q_X(\alpha) = \begin{cases} -2 & ,\ \alpha \in (0,\ 0.3) \\ 10\alpha - 5 & ,\ \alpha \in \langle 0.3,\ 0.4) \\ -0.5 & ,\ \alpha = 0.4 \\ 0 & ,\ \alpha \in (0.4,\ 0.6) \\ \frac{10}{3}\alpha - 2 & ,\ \alpha \in \langle 0.6,\ 0.9) \\ 1 & ,\ \alpha \in \langle 0.9,\ 1). \end{cases}$$

Jestliže nyní budeme uvažovat Kolmogorův model na intervalu $\Omega = (0, 1)$ s rovnoměrným rozdělením pravděpodobnosti, můžeme kvantilovou funkci $q_X : \Omega \to \mathbb{R}$ chápat jako jeden ze způsobů, jak si představit veličinu X, tj. q_X je nyní náhodná veličina se stejnou distribuční funkcí jako má X (skutečně je $F_{q_X} = F_X$) - tj. q_X je jeden z modelů veličiny X.

Tato představa má tu výhodu, že na $\Omega = (0,1)$ můžeme "běžně" integrovat. Díky tomu, že interval $(0,1)$ má délku jedna, bude střední hodnota z q_X jednoduše integrál z této funkce (viz žluté plochy se znaménky na obrázku). Střední hodnota $E(X)$ se tak dá spočítat také jednoduše jako $E(X) = \int_0^1 q_X(\alpha)\, d\alpha$ (viz žluté plochy se znaménky na obrázku).

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 6.2 *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

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

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

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

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

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

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

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

$$\mathrm{var}(X) = \tfrac{1}{p^2} - \tfrac{1}{p}$$

což si můžeme lépe zapamatovat jako

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

Příklad 6.3 *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

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

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

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

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

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