

## 6. cvičení z PST

21. - 24. března 2022

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

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

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

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

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) spojitě*  $\Rightarrow f_X$  je její hustota pravděpodobnosti a  $p_X(t) = 0$  pro všechna  $t \in \mathbb{R}$ .

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

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

kde  $D$  má diskrétní rozdělení a  $S$  má (absolutně) spojitě 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

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

a také

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

a konečně také

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

Poznamenejme, že pokud existuje dělení  $-\infty = t_0 < t_1 < \dots < t_n = \infty$  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  $S$ .

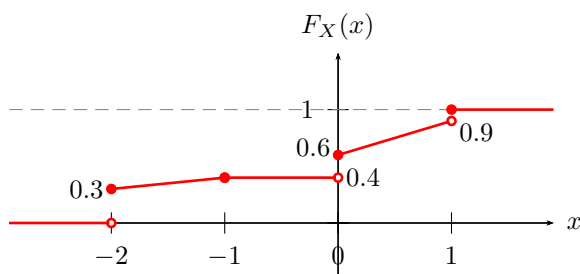
### Příklad 6.1 (rozklad na směs)

Náhodná veličina  $X$  má distribuční funkci

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

(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  $E(X)$ .

**Řešení:**(a) Znázorníme si graf  $F_X$ :

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

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

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

Hodnotu skoku dostaneme pomocí vztahu

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

a podobně

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

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

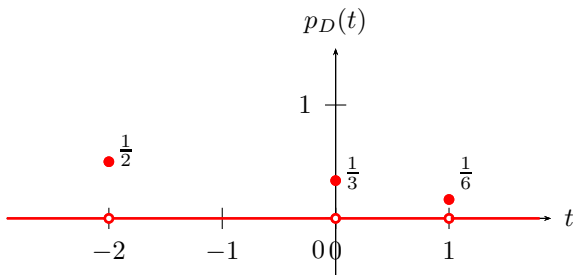
$$F_X(t) = cF_D(t) + (1 - c)F_S(t)$$

- **Popis diskrétní části  $D$ :**

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

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

s grafem:



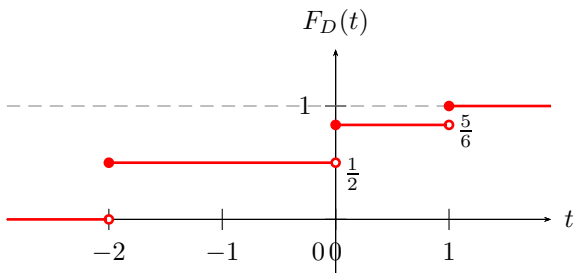
Pro distribuční funkci  $F_D$  diskrétní veličiny  $D$  platí

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

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

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

a graf  $F_D$  je:



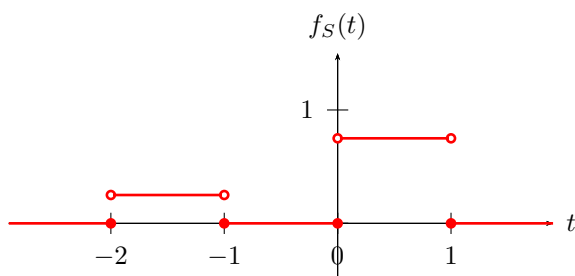
• **Popis spojité části  $S$ :**

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:

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

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

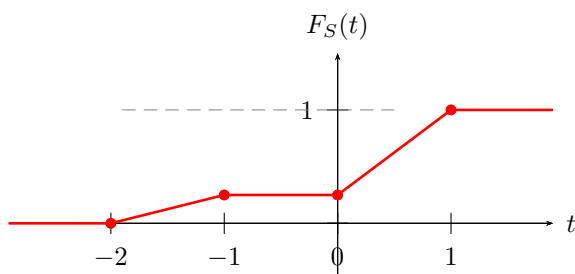
Graf  $f_S$  pak bude:



Pro distribuční funkci spojité veličiny  $S$  pak ze vztahu  $F_X = cF_D + (1 - c)F_S$  dostáváme

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

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

$$E(S) = \int_{-\infty}^{\infty} t \cdot f_S(t) dt = \int_{-2}^{-1} t \cdot \frac{1}{4} dt + \int_0^1 t \cdot \frac{3}{4} dt = \left[ \frac{t^2}{8} \right]_{-2}^{-1} + \left[ \frac{3t^2}{8} \right]_0^1 = \frac{1}{8} - \frac{1}{4} - \frac{3}{8} = 0$$

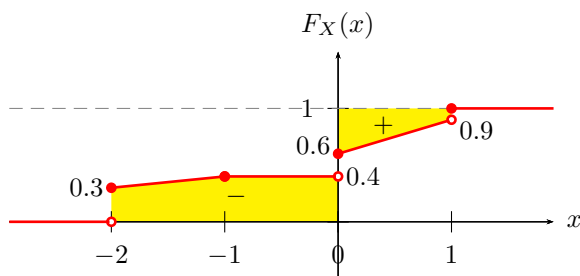
takže

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

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

$$\begin{aligned} E(X) &= - \int_{-\infty}^0 F_X(x) dx + \int_0^{\infty} (1 - F_X(x)) dx = \\ &= - \int_{-2}^{-1} (0.1x + 0.5) dx - \int_{-1}^0 0.4 dx + \int_0^1 (-0.3x + 0.4) dx = \\ &= - \underbrace{\left[ \frac{0.1x^2}{2} \right]_{-2}^{-1}}_{0.15} - 0.5 - 0.4 + \underbrace{\left[ -\frac{0.3x^2}{2} \right]_0^1}_{-0.15} + 0.4 = -0.5 . \end{aligned}$$

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).



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

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

s hodnotami  $X \in \{0, 1, \dots, n\}$  má pak tzv. *binomické* rozdělení  $\text{Bi}(n, p)$ . 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  $n$  pokusy 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

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

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

(b) Určete rozdělení náhodné veličiny

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

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

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

**Řešení:**

- (a) Veličina  $X$  má *alternativní* rozdělení, tj.  $X \sim \text{Alt}(p)$ , 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

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

- (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  $k = 0, 1, \dots, n$  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

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

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)

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

**Připomenutí:** Veličina

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

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

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

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

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

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 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

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

Ta má *geometrické* rozdělení  $\text{Geom}(p)$  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

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

Ta má binomické rozdělení  $\text{Bi}(4, p)$ , 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 6.4** 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  $\text{Geom}(0.2)$  s oborem hodnot  $\{0, 1, 2, \dots\}$  a jejich pravděpodobnostmi

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

pro  $k = 0, 1, 2, \dots$

**POZOR:** Nepleťte 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

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

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  $P(Y \leq n) \geq 0.9$ , neboli

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

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

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

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

$Z_n =$  “počet tolika hodů (z  $n$  možných), ve kterých se trefíme”

která zřejmě má binomické rozdělení  $\text{Bi}(n, p)$ . 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.