Pravděpodobnostní metody ve strojírenství

3. Náhodná veličina

FAKULTA STROJNÍ ČVUT V PRAZE

prof. RNDr. Gejza Dohnal, CSc.

ak. rok 2023/2024

3. Náhodná veličina

Klíčové pojmy:

- Náhodná veličina, diskrétní, spojitá;
- Rozdělení pravděpodobnosti, distribuční funkce;
- pravděpodobnostní funkce, hustota pravd.;
- Střední hodnota n.v., centrální momenty (rozptyl, směrodatná odchylka, šikmost, špičatost);
- p-kvantil rozdělení n.v.;
- medián, modus;

Klíčové vztahy:

- Alternativní model
- Binomický model
- Hypergeometrický model
- Geometrický model
- Model rovnoměrného rozdělení

Co máme?

1. elementární náhodné jevy Ω

2. jevové pole $\mathcal{F}$:
   (1) $\emptyset \in \mathcal{F}$
   (2) $A \in \mathcal{F} \Rightarrow \bar{A} \in \mathcal{F}$
   (3) $A_1, A_2, \dots \in \mathcal{F} \Rightarrow \bigcup_{i=1}^{\infty} A_i \in \mathcal{F}$

3. Pravděpodobnost
   (a) $\forall A \in \mathcal{F}: \quad P(A) \in \langle 0, 1 \rangle$
   (b) $P(\emptyset) = 0, \quad P(\Omega) = 1$
   (c) jsou-li $A_1, A_2, \dots \in \mathcal{F}$, po dvou disjunktní, potom
   $P\left(\bigcup_{i=1}^{\infty} A_i\right) = \sum_{i=1}^{\infty} P\left(A_i\right)$

$\Rightarrow \quad (\Omega, \mathcal{F}, P)$ …. pravděpodobnostní prostor

Co nám chybí?

4. Náhodná veličina: $X : \mathbf{\Omega} \to \mathbf{R}$

Přesněji: měřitelná funkce $X : (\Omega, \mathcal{F}) \to (\mathbf{R}, \mathcal{B})$

$$\forall x \in \mathbf{R}: \quad \{\omega \in \Omega : X(\omega) \leq x\} \in \mathcal{F}$$

$$P\left(\{\omega \in \Omega : X(\omega) \leq x\}\right) = F(x), \quad x \in \mathbf{R}$$

Náhodná veličina

$$X : (\Omega, \mathcal{F}) \to (\mathbf{R}, \mathcal{B})$$

$$\{\omega \in \Omega : X(\omega) \in (-\infty, x\rangle\} \in \mathcal{F}$$

$$\{X \leq x\} \in \mathcal{F}$$

Distribuční funkce náhodné veličiny: $F(x) = P\,(X \leq x)\,, \quad x \in \mathbf{R}$

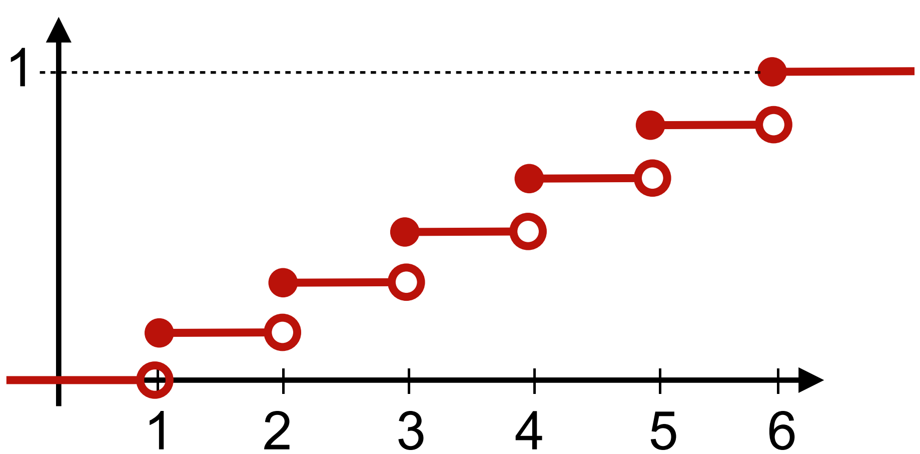
Distribuční funkce odpovídá na otázku: S jakou pravděpodobností náhodná veličina nepřekročí hodnotu x?

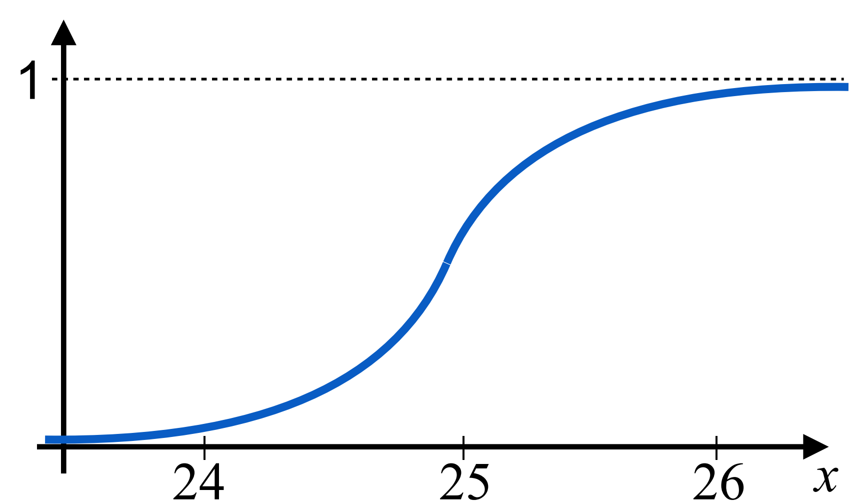
Pro distribuční funkci platí:

1) je neklesající $\quad x \leq y \ \Rightarrow \{X \leq x\} \subset \{X \leq y\} \Rightarrow P(X \leq x) \leq P(X \leq y) \Rightarrow F(x) \leq F(y).$

2) $\lim\limits_{x \to -\infty} F(x) = \lim\limits_{x \to -\infty} P(X \leq x) = 0$; $\lim\limits_{x \to +\infty} F(x) = \lim\limits_{x \to +\infty} P(X \leq x) = 1$

3) $F(x)$ je zprava spojitá $\quad \lim\limits_{x \to x_0+} F(x) = F(x_0)$





Náhodná veličina

$$X : (\Omega, \mathcal{F}) \to (\mathbf{R}, \mathcal{B})$$

Rozlišujeme dva případy:

Náhodná veličina je funkce z Ω do

diskrétní náhodná veličina

a) spočetné podmnožiny $S \subset \mathbf{R}$

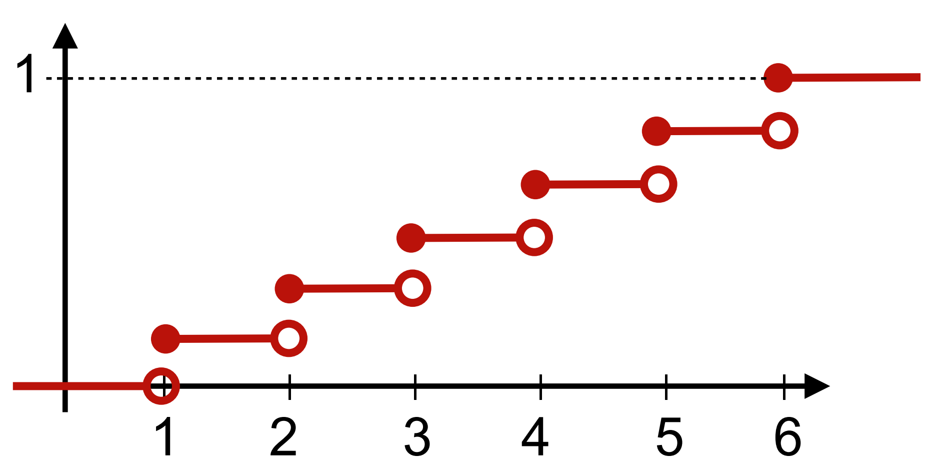
b) souvislé podmnožiny $Q \subset \mathbf{R}$

spojitá náhodná veličina

a) Diskrétní náhodná veličina:

- obor hodnot = $\{x_1, x_2, \ldots, x_n, \ldots\}$
- pravděpodobnostní funkce: $P(X = x_i) = p_i,\ i=1, 2, \ldots$

  vždycky musí platit, že $p_i \in \langle 0, 1 \rangle, \quad \sum_{i=1}^{\infty} p_i = 1$
- distribuční funkce: $F(x) = P(X \leq x) = \sum_{i=:x_i \leq x} p_i, \quad x \in \mathbf{R}$



Náhodná veličina

$X : (\Omega, \mathcal{F}) \to (\mathbf{R}, \mathcal{B})$

Rozlišujeme dva případy:

Náhodná veličina je funkce z Ω do

a) spočetné podmnožiny $S \subset \mathbf{R}$

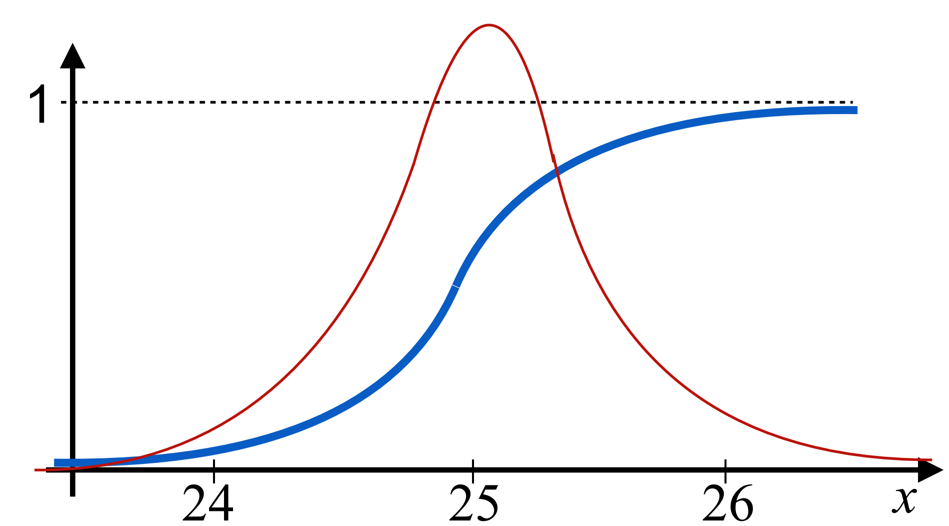
b) souvislé podmnožiny $Q \subset \mathbf{R}$

b) Spojitá náhodná veličina:

- obor hodnot = (a,b) či $\langle a, \infty)$ nebo celé $\mathbf{R}$
- nutně platí: $P(x) = 0, \ \forall x \in (a,b)$
- hustota pravděpodobnosti: nezáporná, spojitá funkce $f(x)$ pro kterou platí:

$$F(x) = \int_{-\infty}^{x} f(t)\mathrm{d}t, \quad \forall x \in \mathbf{R}$$

vždycky musí být $\int_{-\infty}^{+\infty} f(t)\mathrm{d}t = 1$



Charakteristiky náhodné veličiny - Momenty

Střední hodnota: $E(X) = \int_{-\infty}^{+\infty} x \mathrm{d}F(x)$

a) v případě diskrétní náhodné veličiny $E(X) = \sum_{i=1}^{\infty} x_i p_i$

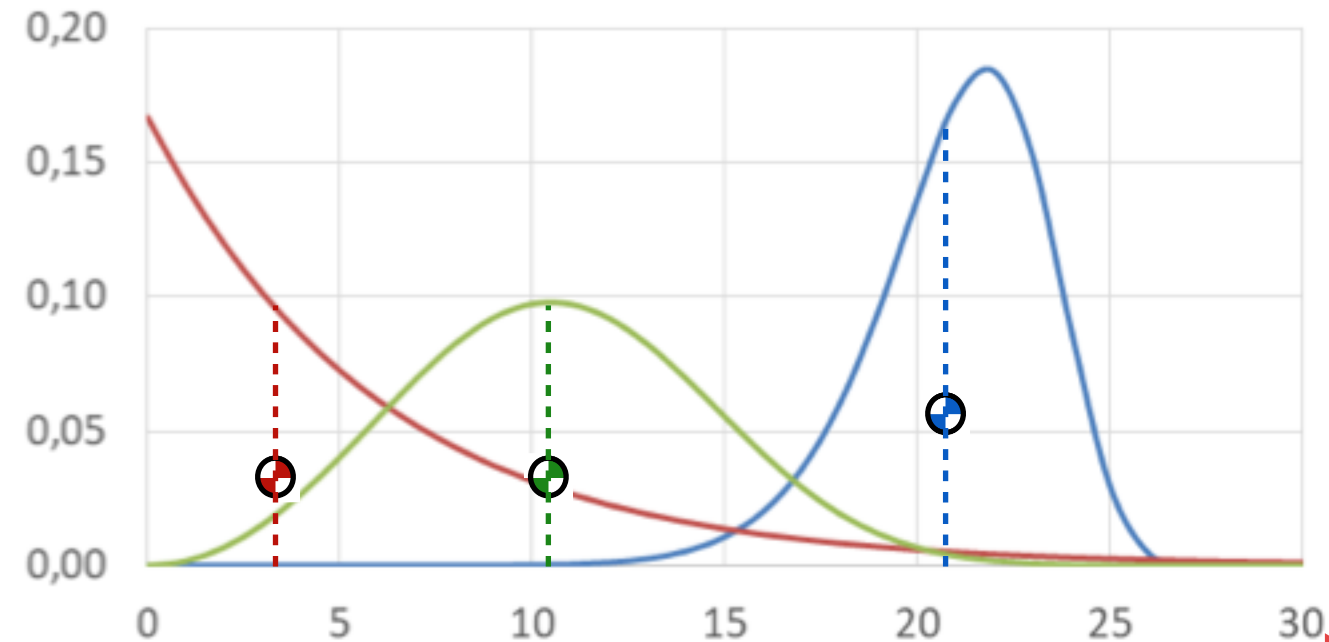
b) v případě spojité náhodné veličiny $E(X) = \int_{-\infty}^{\infty} x f(x) \mathrm{d}x$

Střední hodnota je lineární funkcionál:

je aditivní: $E(X+Y) = E(X) + E(Y)$

a je homogenní: $E(kX) = kE(X), \quad k \in \mathbf{R}$

Střední hodnota tvoří jakési
"těžiště rozdělení" náhodné veličiny:



Charakteristiky náhodné veličiny - Momenty

Obecné momenty: $\mu_k(X) = E(X^k) = \int_{-\infty}^{+\infty} x^k \mathrm{d}F(x)$

$\mu_k(X) = \sum_{i=1}^{\infty} x_i p_i$ v diskrétním případě (pro d.n.v.)

$\mu_k(X) = \int_{-\infty}^{+\infty} x^k f(x) \mathrm{d}x$ ve spojitém případě (pro s.n.v.)

Centrální momenty: $X_C = X - \mathrm{E}(X) = X - \mu$ ….. centrovaná náhodná veličina

$$\nu_k(X) = E(X_C^k) = E(X - \mu)^k = \int_{-\infty}^{+\infty} (x - \mu)^k \mathrm{d}F(x)$$

1) $\nu_1(X) = E(X_C) = E(X - \mu) = E(X) - \mu = E(X) - E(X) = 0$

Charakteristiky náhodné veličiny - Momenty

2) $$\nu_2(X) = E(X-\mu)^2 = E(X^2 - 2\mu X + \mu^2) \; = E(X^2) - 2\mu E(X) + \mu^2$$
$$= E(X^2) - 2\Big(E(X)\Big)^2 + \Big(E(X)\Big)^2 = E(X^2) - \Big(E(X)\Big)^2$$
$$= \mu_2(X) - \Big(\mu_1(X)\Big)^2$$

druhý centrální moment = **rozptyl** : $Var(X) = E(X - EX)^2 \; = \mu_2 - \mu_1^2$
(míra variability)

odvozenou charakteristikou je
směrodatná odchylka: $\sigma(X) = \sqrt{Var(X)} = \sqrt{E(X-\mu)^2} \; = \sqrt{\nu_2}$

další odvozenou charakteristikou je
variační koeficient: $$V = \frac{\sigma(X)}{E(X)} = \frac{\sqrt{(\nu_2)}}{\mu_1}$$

3) Třetí centrální moment je mírou symetrie; **koeficient šikmosti:** $$Skew(X) = \frac{E(X-EX)^3}{(Var(X))^{3/2}} = \frac{\nu_3}{\mu_2^{3/2}}$$

4) Čtvrtý centrální moment je mírou koncentrace; **koeficient špičatosti:** $$Kurt(X) = \frac{E(X-EX)^4}{(Var(X))^2} = \frac{\nu_4}{\mu_2^2}$$

Charakteristiky náhodné veličiny - Kvantily

Kvantilová funkce: $\tilde{x}_\alpha = F^{-}(\alpha)$

Často se ptáme: Jakou hodnotu sledovaná náhodná veličina nepřekročí s danou pravděpodobností?

… tedy pro zadané $\alpha \in (0,1)$ hledáme $\tilde{x}_\alpha \in \mathbf{R}$ tak, že $F(\tilde{x}_\alpha) \leq \alpha$

- je-li $F(x)$ spojitá a prostá, potom je $\tilde{x}_\alpha = F^{-1}(\alpha)$
- v ostatních případech je to taková maximální hodnota $\tilde{x}_\alpha$, pro kterou platí: $P(X \leq \tilde{x}_\alpha) \geq \alpha$ a zároveň $P(X < \tilde{x}_\alpha) \leq \alpha$ (tzv. pseudoinverze).
- někdy hovoříme o $\tilde{x}_\alpha$ jako o α-kvantilu nebo o 100α%-kvantilu náhodné veličiny X.

Patří sem:

- minimum ($\tilde{x}_0$), maximum ($\tilde{x}_{1,0}$),
- medián ($\tilde{x}_{0,5}$),
- dolní decil ($\tilde{x}_{0,1}$) a horní decil ($\tilde{x}_{0,9}$),
- dolní kvartil ($\tilde{x}_{0,25}$), horní kvartil ($\tilde{x}_{0,75}$)
- a další ….

Charakteristiky náhodné veličiny - další …

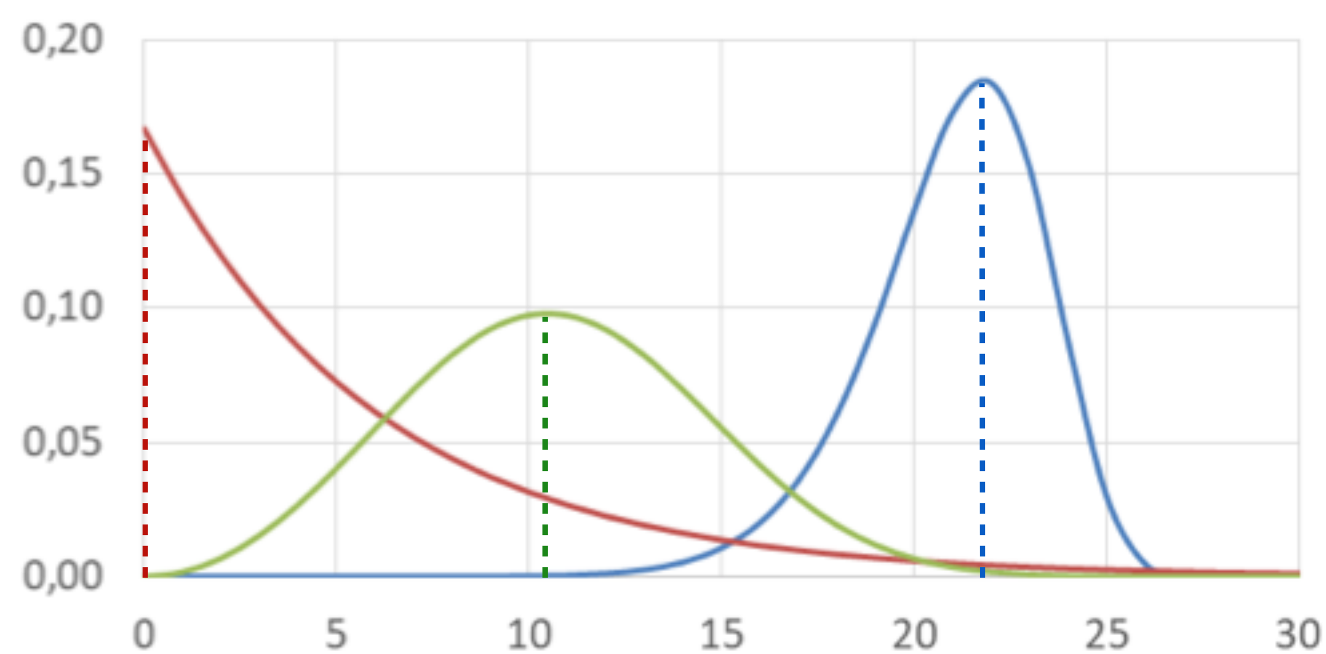
Patří sem například :

- modus (nejpravděpodobnější hodnota)
- rozpětí ($\tilde{x}_1 - \tilde{x}_0$),
- mezikvartilová odchylka ($\tilde{x}_{0,75} - \tilde{x}_{0,25}$)
- a další ….

v diskrétním případě: $x_{\mathrm{mod}} = \max\{p_i; i = 1, 2, \ldots\}$

ve spojitém případě: $x_{\mathrm{mod}} = \mathrm{argmax}\{f(x); x \in \mathbf{R}\}$

$$P(X = x_0) = 0$$
$$P(X \in \langle x_0; \mathrm{d}x\,\rangle) \doteq f(x_0)\mathrm{d}x$$



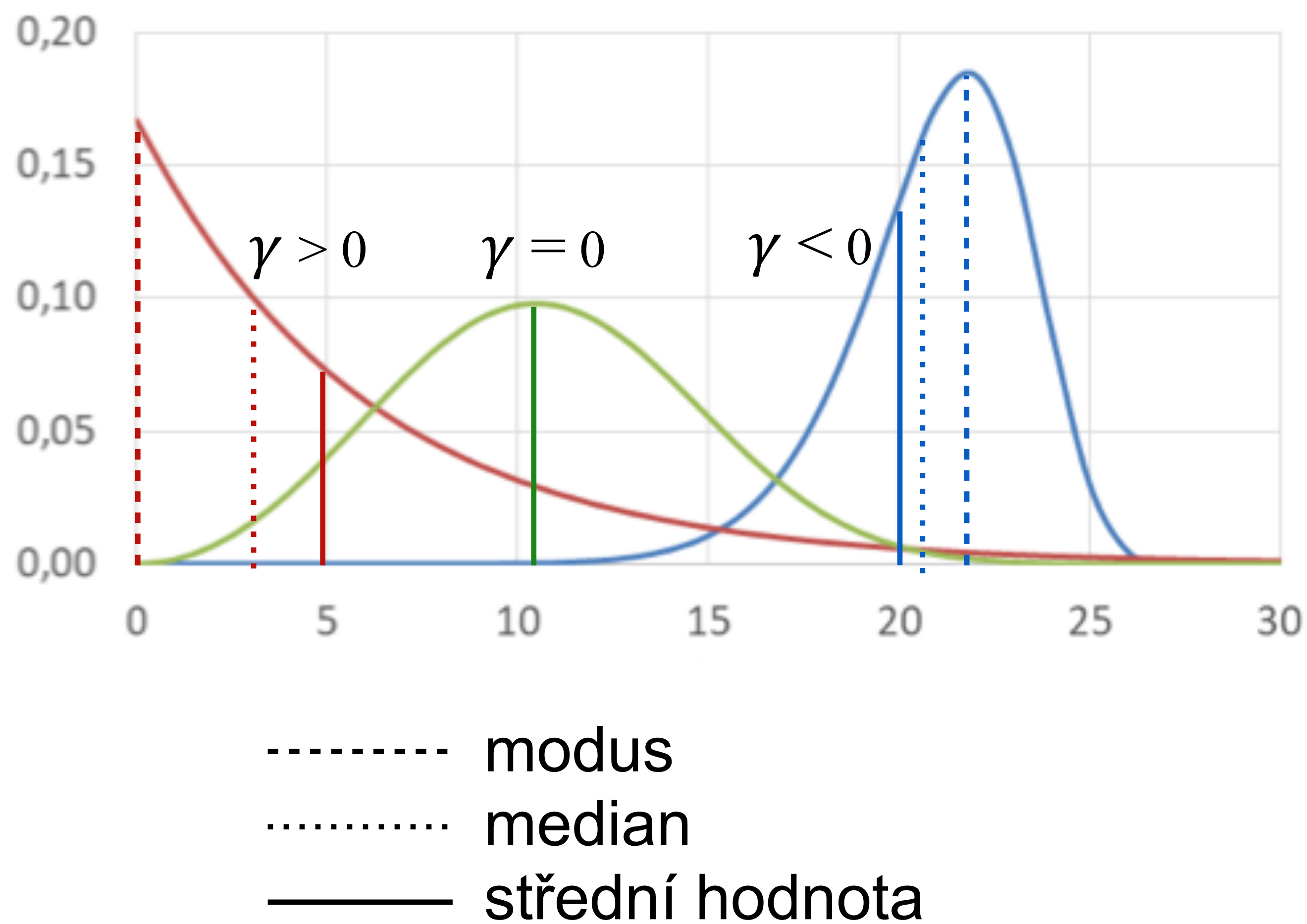
Charakteristiky náhodné veličiny - ...

Charakteristiky podle funkce:

- míry polohy - posouvají se s náhodnou veličinou:
  je-li θ mírou polohy veličiny X, potom veličina $X+d$ má míru polohy $\theta+d$
  (patří sem střední hodnota, kvantily, modus, …)

- míry rozptýlenosti (variability) - jsou invariantní vůči posunutí:
  je-li ϑ mírou variability veličiny X, potom je mírou variability i veličiny $X+d$
  (patří sem rozptyl, směrodatná odchylka, různá rozpětí, …)

- míry symetrie (šikmost)



označíme-li šikmost = γ, potom

$\gamma > 0$ – zešikmené zprava (doleva)
$\gamma = 0$ – symetrické
$\gamma < 0$ – zešikmené zleva (doprava)

Rozdělení pravděpodobnosti náhodné veličiny - různé modely

Pravděpodobnostní model = popis experimentu, při němž se realizuje náhodná veličina (jedna či více) s nějakým rozdělením pravděpodobnosti

$\Rightarrow \quad (\Omega, \mathcal{F}, P)$ …. pravděpodobnostní prostor

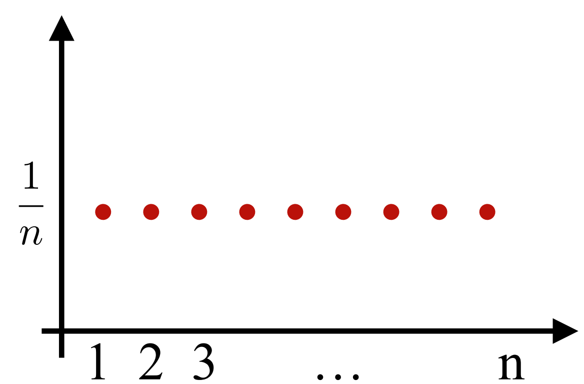
1) Model diskrétního rovnoměrného rozdělení

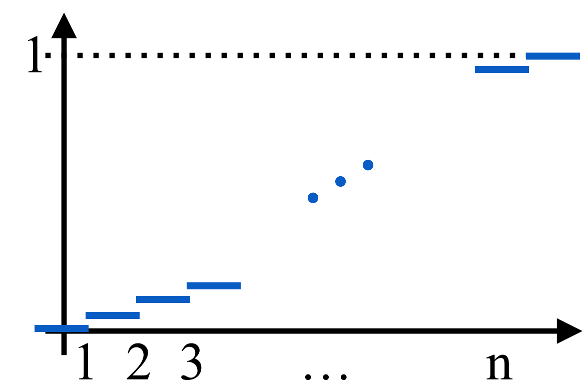
Experiment, při němž může nastat jeden z n možných výsledků. Předpokládáme, že každý z nich je stejně možný, tedy všechny mají stejnou pravděpodobnost.

$\Omega = \{a_1, a_2, \dots, a_n\}, \quad X \in \{x_1, x_2, \dots, x_n\}, \quad P(X = x_i) = 1/n, \quad i = 1, 2, \dots, n.$

$$E(X) = \sum_{i=1}^{n} x_i \frac{1}{n} = \frac{1}{n} \sum_{i=1}^{n} x_i = \bar{x}$$

$$Var(X) = \frac{1}{n} \sum_{i=1}^{n} x_i^2 - \bar{x}^2$$





$X \in \{1, 2, \dots, n\} \Rightarrow E(X) = \frac{n+1}{2} \qquad E(X^2) = \frac{(n+1)(2n+1)}{6}$

$$Var(X) = E(X^2) - (E(X))^2 = \frac{n^2 - 1}{12}$$

Rozdělení pravděpodobnosti náhodné veličiny - různé modely

2) Alternativní model (model alternativního rozdělení)

Experiment, při němž může nastat pouze jeden ze dvou možných výsledků.

$\Omega = \{a_1, a_2\}$, $X \in \{0, 1\}$, $\mathrm{P}(X = 1) = p$ => $\mathrm{P}(X = 0) = 1 - p$

$E(X) = 0 \cdot (1-p) + 1 \cdot p = p$ $\qquad$ $E(X^2) = 0^2 \cdot (1-p) + 1^2 \cdot p = p$

$Var(X) = E(X^2) - (E(x))^2 = p - p^2 = p(1-p)$

3) Bernoulliho model (model binomického rozdělení)

Experiment, při němž opakujeme n-krát nezávisle na sobě alternativní pokus za stejných podmínek. Zajímá nás počet „úspěchů", tedy kolikrát nastal výsledek „1".

absolutní četnost kladných výsledků = součet pozorování $Y = \sum_{i=1}^{n} X_i$

$$\boxed{P(Y = k) = \binom{n}{k} p^k (1-p)^{n-k}, \quad k = 0, 1, \ldots, n.} \qquad Y \sim Bin(n, p)$$

Cvičení: 1) Ukažte, že součet těchto pravděpodobností je roven jedné.
2) Spočtěte střední hodnotu a rozptyl veličiny Y podle definice.

Rozdělení pravděpodobnosti náhodné veličiny - různé modely

3) Bernoulliho model (model binomického rozdělení)

absolutní četnost kladných výsledků = součet pozorování $Y = \sum_{i=1}^{n} X_i$

$$E(Y) = E\left(\sum_{i=1}^{n} X_i\right) = \sum_{i=1}^{n} E(X_i) = np \qquad Var(Y) = E(Y - np)^2 = np(1-p)$$

relativní četnost kladných výsledků = aritmetický průměr pozorování $\bar{X} = \frac{1}{n}\sum_{i=1}^{n} X_i$

$$E(\bar{X}) = E\left(\frac{1}{n}\sum_{i=1}^{n} X_i\right) = \frac{1}{n}E\left(\sum_{i=1}^{n} X_i\right) = \frac{1}{n}\sum_{i=1}^{n} E(X_i) = \frac{1}{n}np = p$$

$$Var(\bar{X}) = E(\bar{X} - p)^2 = \frac{p(1-p)}{n}$$

Rozdělení pravděpodobnosti náhodné veličiny - různé modely

3) Bernoulliho model (model binomického rozdělení)

Experiment, při němž z množiny N prvků, mezi nimiž je M prvků s určitou vlastností, vybíráme náhodně postupně n prvků tak, že po každém výběru vybraný prvek vrátíme zpět mezi ostatní. Zajímá nás kolikrát vybereme prvek s danou vlastností.

$$\mathrm{P}(\text{vybereme prvek s danou vlastností}) = \frac{M}{N}, \quad Y = \text{počet vybraných prvků s danou vlastností}$$

$$P(Y=k) = \binom{N}{n}\left(\frac{M}{N}\right)^k\left(1-\frac{M}{N}\right)^{n-k} \qquad Y \sim Bin\left(n, \frac{M}{N}\right)$$

počet k-tic
v n prvcích

k-krát vybereme prvek
s pravděpodobností $\frac{M}{N}$

(n-k)-krát vybereme prvek
s pravděpodobností (1- $\frac{M}{N}$)

Rozdělení pravděpodobnosti náhodné veličiny - různé modely

3) Bernoulliho model (model binomického rozdělení)

Experiment, při němž z množiny N prvků, mezi nimiž je M prvků s určitou vlastností, vybíráme náhodně postupně n prvků tak, že po každém výběru vybraný prvek vrátíme zpět mezi ostatní. Zajímá nás kolikrát vybereme prvek s danou vlastností.

$$\mathrm{P}(\text{vybereme prvek s danou vlastností}) = \frac{M}{N}, \quad Y = \text{počet vybraných prvků s danou vlastností}$$

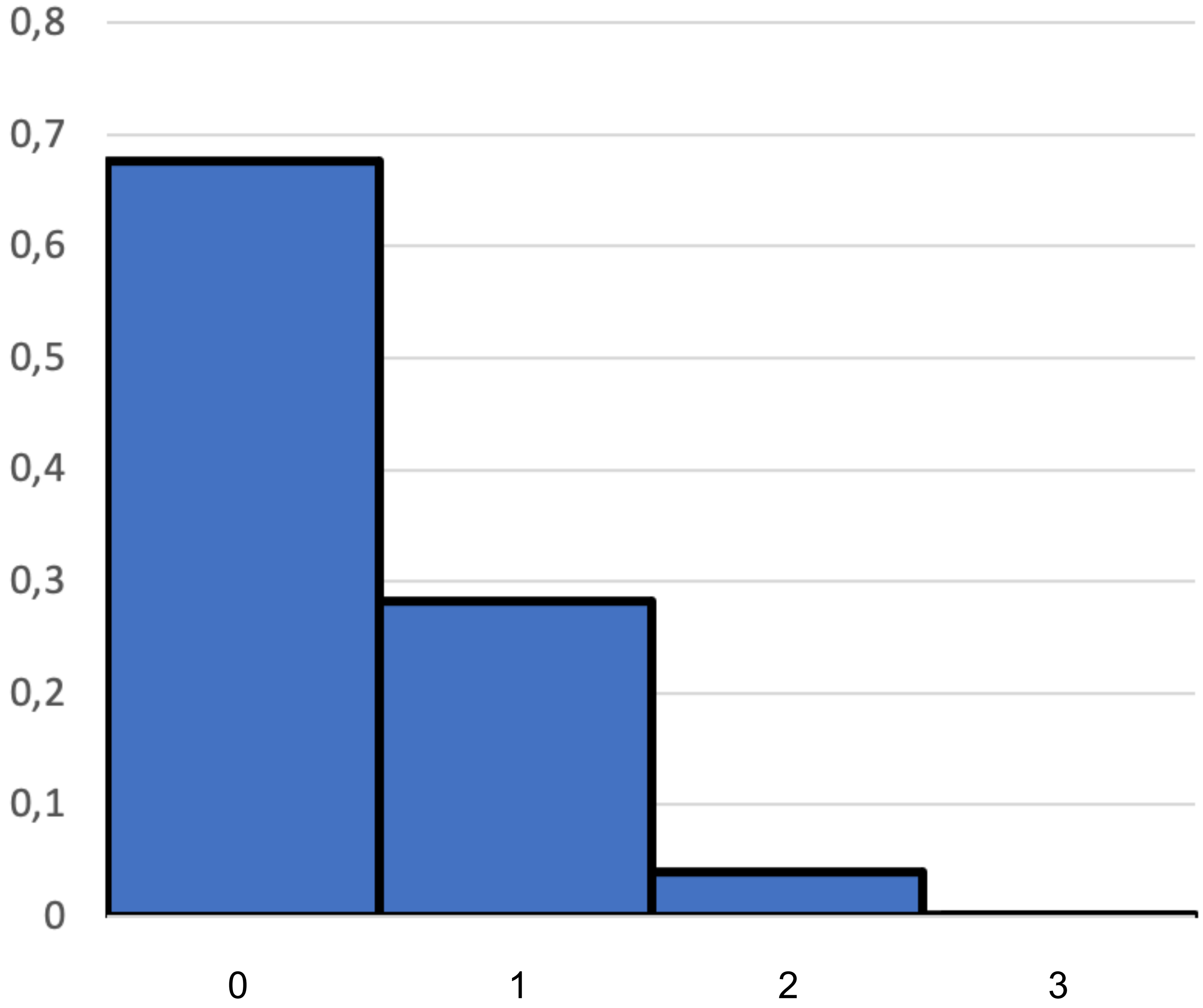
$$P(Y = k) = \binom{N}{n} \left(\frac{M}{N}\right)^k \left(1 - \frac{M}{N}\right)^{n-k} \qquad Y \sim Bin\left(n, \frac{M}{N}\right)$$

Příklad 1: Házení kostkou: házíme třemi hracími kostkami současně (nebo jednou třikrát po sobě). Jaká je pravděpodobnost, že padnou alespoň dvě šestky?

Příklad 2: Kontrola jakosti: Z výrobní linky odebíráme nezávisle na sobě 10 výrobků. Víme, že v produkci jsou 3% vadných. Jaká je pravděpodobnost, že při výběru vybereme alespoň 1 zmetek?

Příklad 3: Losování zaměstnance: každý den v týdnu losujeme jednoho z 15ti zaměstnanců, který provede odpolední úklid. Mezi zaměstnanci 10 žen. Jaká je pravděpodobnost, že v týdnu vybereme samé ženy?

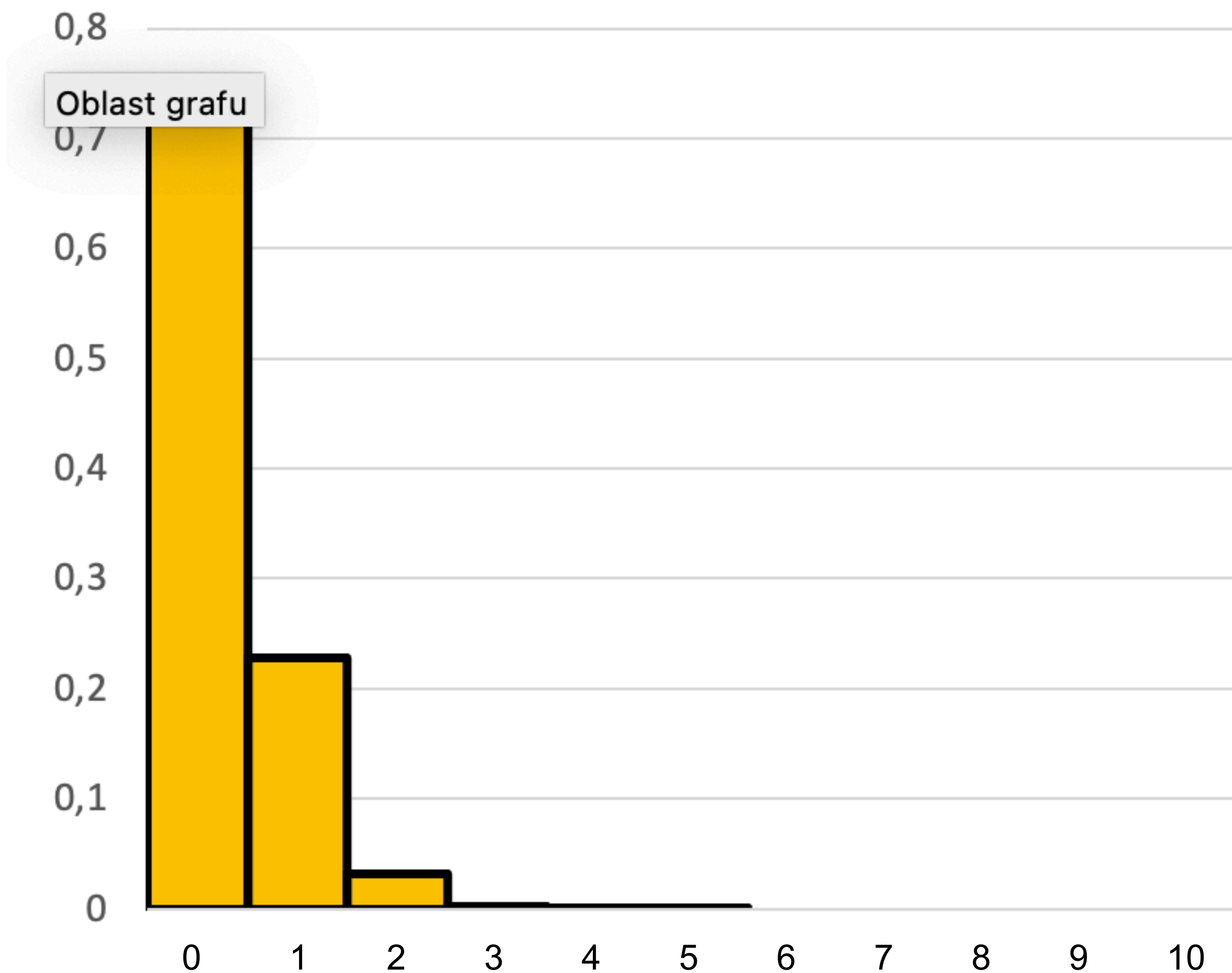
Příklad 1: Házení kostkou: házíme třemi hracími kostkami současně (nebo jednou třikrát po sobě). Jaká je pravděpodobnost, že padnou alespoň dvě šestky?



X = počet šestek

k	P(X=k)
0	0,6757983
1	0,2828923
2	0,0394733
3	0,001836

Příklad 2: Kontrola jakosti: Z výrobní linky odebíráme nezávisle na sobě 10 výrobků. Víme, že v produkci jsou 3% vadných. Jaká je pravděpodobnost, že při výběru vybereme alespoň 1 zmetek?



X = počet vybraných zmetků

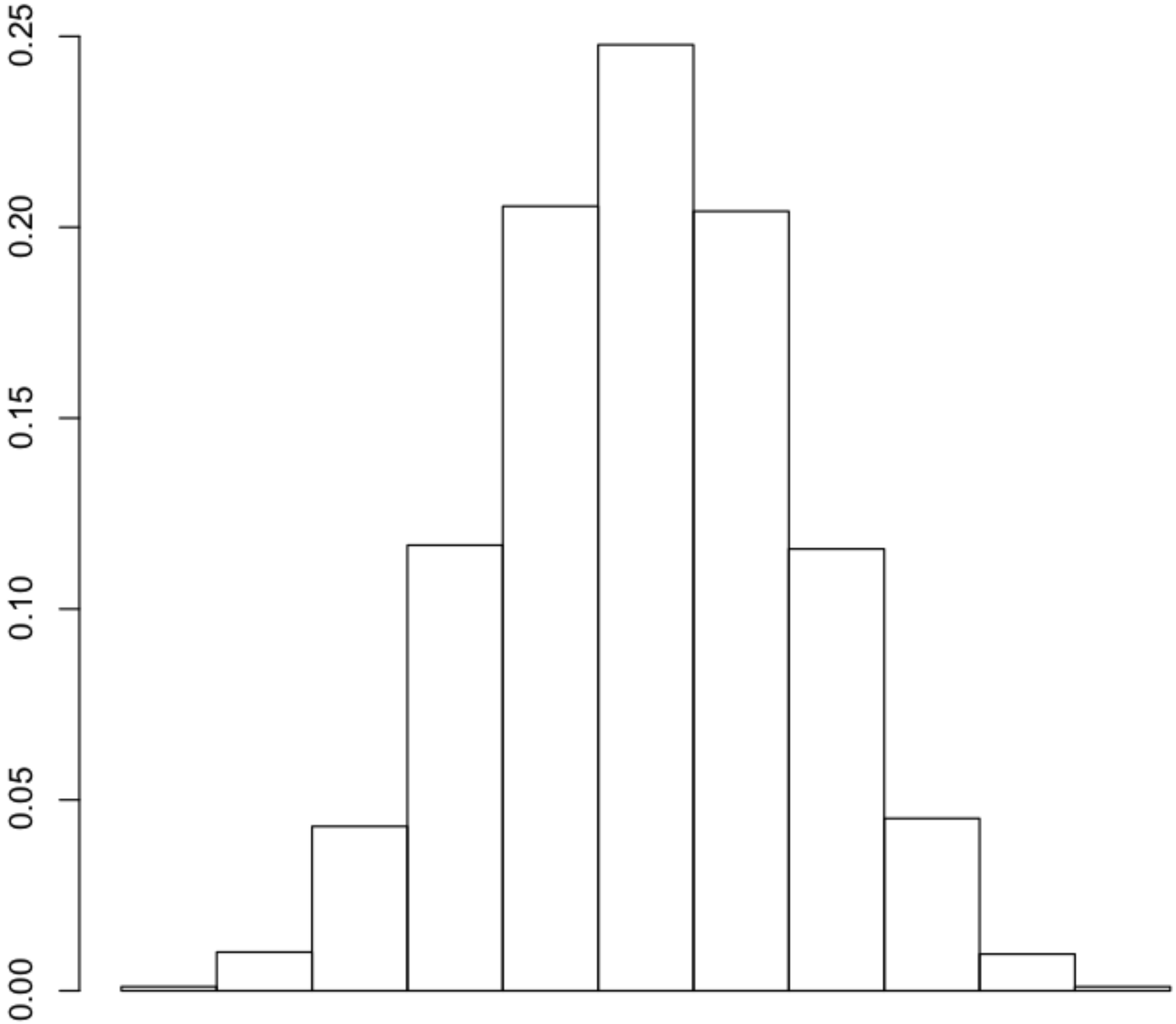
k	P(X=k)
0	0,7374241
1	0,2280693
2	0,0317416
3	0,0026179
4	0,0001417
5	5,259E-06
6	1,355E-07
7	2,395E-09
8	2,778E-11
9	1,909E-13
10	5,905E-16

Příklad 3: Losování zaměstnance: každý den v týdnu losujeme jednoho z 15ti zaměstnanců, který provede odpolední úklid. Mezi zaměstnanci 10 žen. Jaká je pravděpodobnost, že v týdnu vybereme samé ženy?

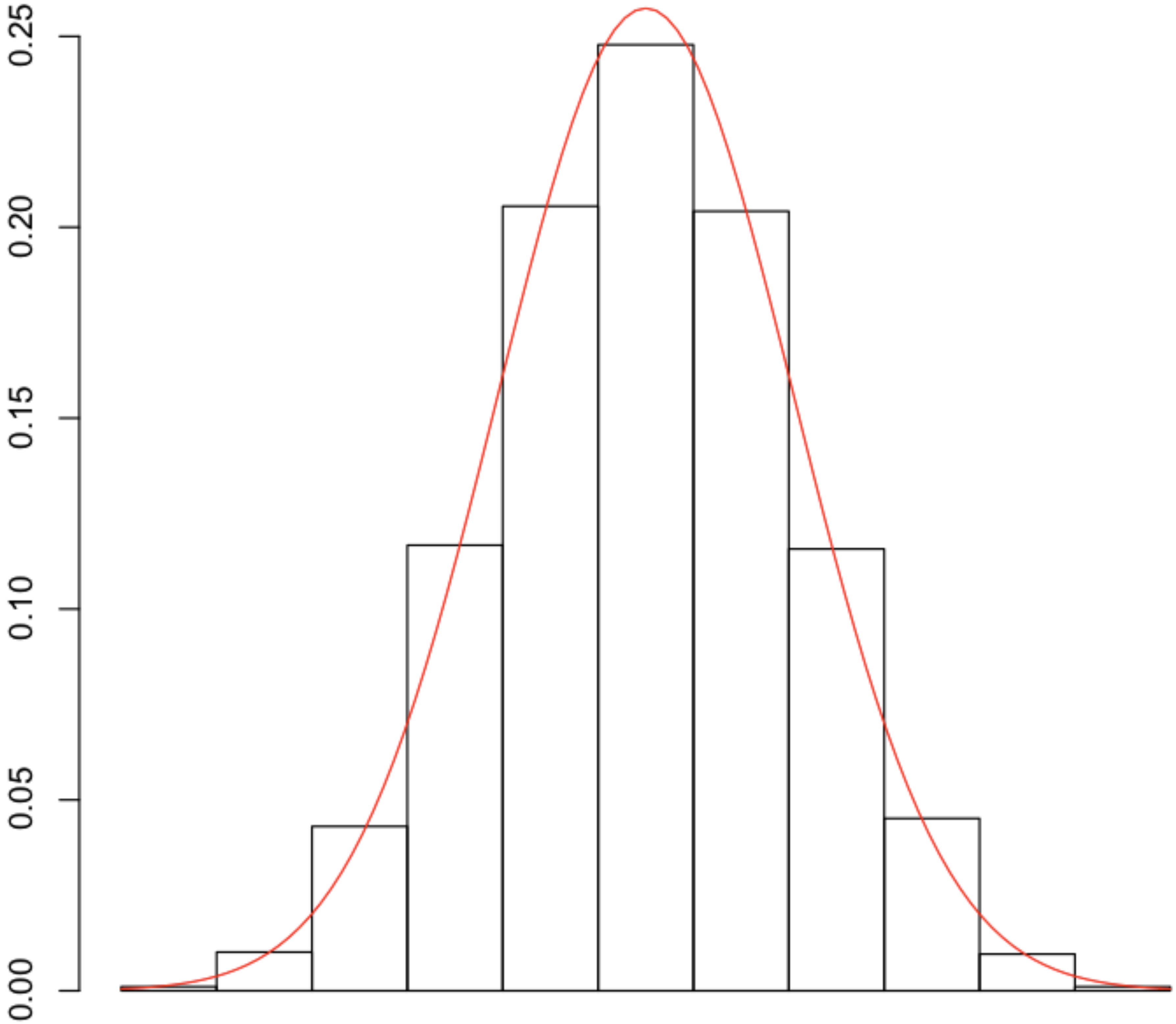
X = počet vybraných žen

k	P(X=k)
0	0,0041152
1	0,0411523
2	0,1646091
3	0,3292181
4	0,3292181
5	0,1316872

Binomické rozdělení lze pro velká n a relativně malá k aproximovat spojitým normálním rozdělením



Binomické rozdělení lze pro velká n a relativně malá k aproximovat spojitým normálním rozdělením



Rozdělení pravděpodobnosti náhodné veličiny - různé modely

4) Hypergeometrický model (model výběru bez vracení)

Experiment, při němž z množiny N prvků, mezi nimiž je M prvků s určitou vlastností, vybíráme náhodně postupně n prvků tak, že po každém výběru vybraný prvek nevracíme zpět mezi ostatní. Zajímá nás kolik prvků s danou vlastností vybereme.

k prvků s danou vlastností z celkového počtu M můžeme vybrat celkem $\binom{M}{k}$ způsoby

zbylých n-k prvků bez této vlastnost můžeme vybrat celkem $\binom{N-M}{n-k}$ způsoby

celkový počet možných n výběrů z N prvků je $\binom{N}{n}$

Tedy: $$P(X=k) = \frac{\binom{M}{k}\binom{N-M}{n-k}}{\binom{N}{n}}$$

$$N = 1, 2, \ldots, \quad M \leq N, \quad n \leq N,$$
$$\max\{\,0,\ n+M-N\,\} \leq k \leq \min\{\,n,\ M\,\}$$

$$E(X) = n\frac{M}{N}$$

$$Var(X) = \frac{nM(N-n)(N-M)}{N^2(N-1)}$$

Hypergeometrické rozdělení s parametry N, M, n.

Rozdělení pravděpodobnosti náhodné veličiny - různé modely

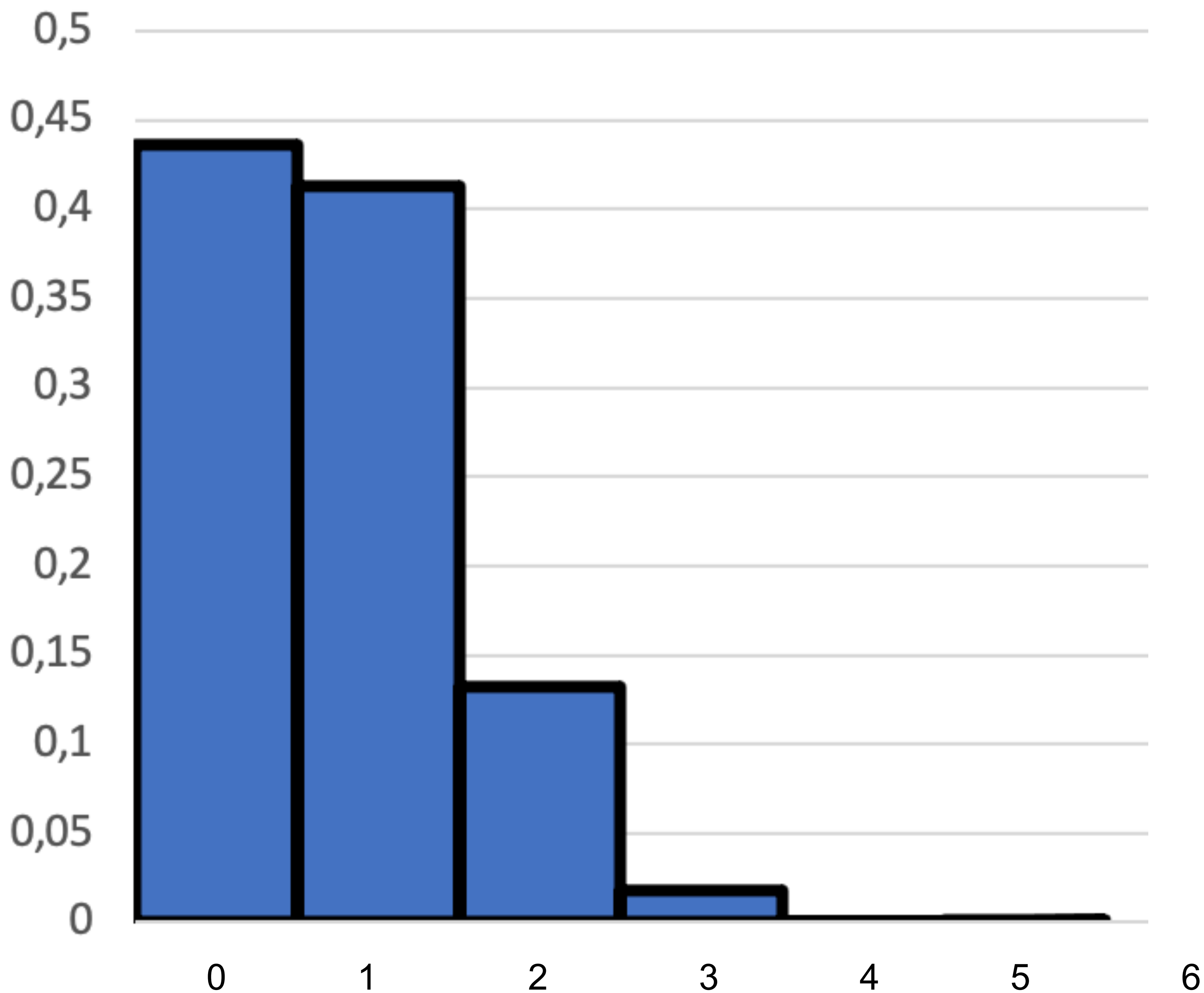
4) Hypergeometrický model (model výběru bez vracení)

Příklad 4: Sportka: 49 čísel, ze kterých 6 vyhrává (jsou vytaženy). Jaká je pravděpodobnost, že při výběru 6ti čísel vybereme 4 z tažených?

Příklad 5: Kontrola jakosti: 1000 výrobků, mezi nimi jsou 3% vadných. Jaká je pravděpodobnost, že při výběru 10 výrobků vybereme alespoň 1 zmetek?

Příklad 6: Výběr uchazečů o práci: z 15ti uchazečů o zaměstnání, mezi kterými je 10 žen, vybíráme anonymně podle výsledku testu 5 osob. Jaká je pravděpodobnost, že to budou samé ženy?

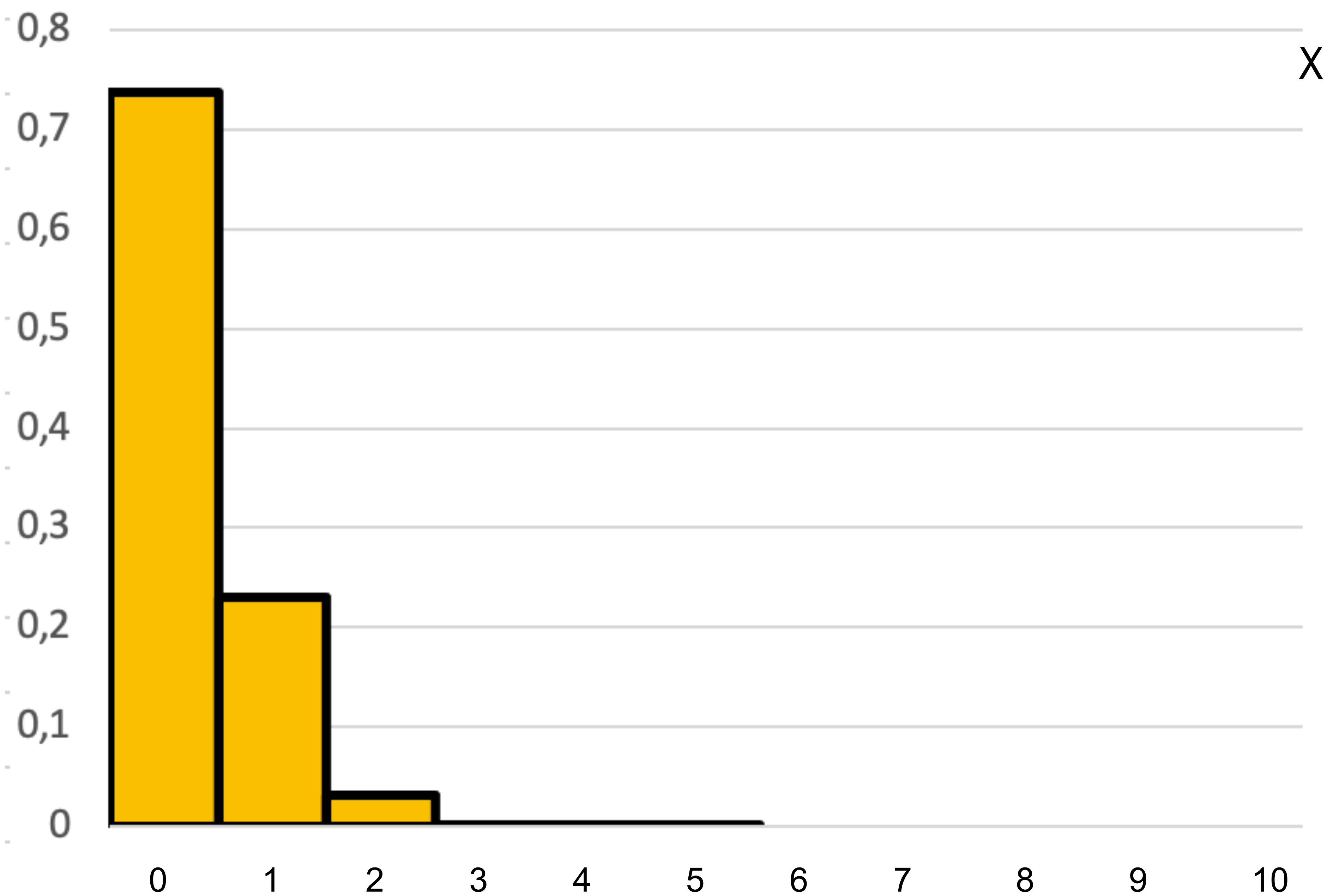
Příklad 4: Sportka: 49 čísel, ze kterých 6 vyhrává (jsou vytaženy). Jaká je pravděpodobnost, že při výběru 6ti čísel vybereme 4 z tažených?



X = počet uhodnutých tažených čísel

k	P(X=k)
0	0,43596498
1	0,41301945
2	0,13237803
3	0,0176504
4	0,00096862
5	1,845E-05
6	7,1511E-08

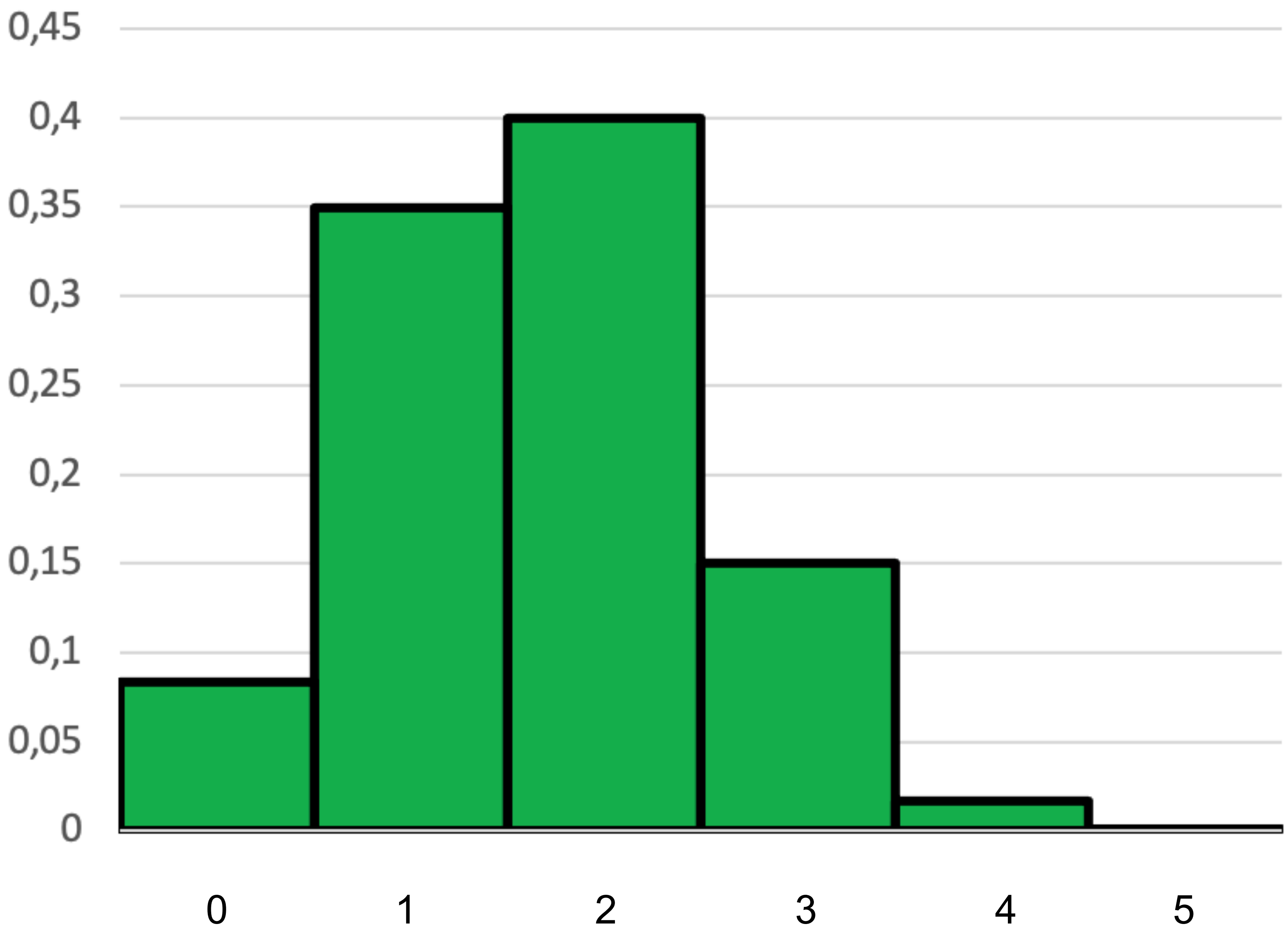
Příklad 5: Kontrola jakosti: 1000 výrobků, mezi nimi jsou 3% vadných. Jaká je pravděpodobnost, že při výběru 10 výrobků vybereme alespoň 1 zmetek?



X = počet vybraných zmetků

k	P(X=k)
0	0,73639189
1	0,229883
2	0,03118475
3	0,00241792
4	0,00011851
5	3,8317E-06
6	8,2637E-08
7	1,172E-09
8	1,0443E-11
9	5,2686E-14
10	1,1406E-16

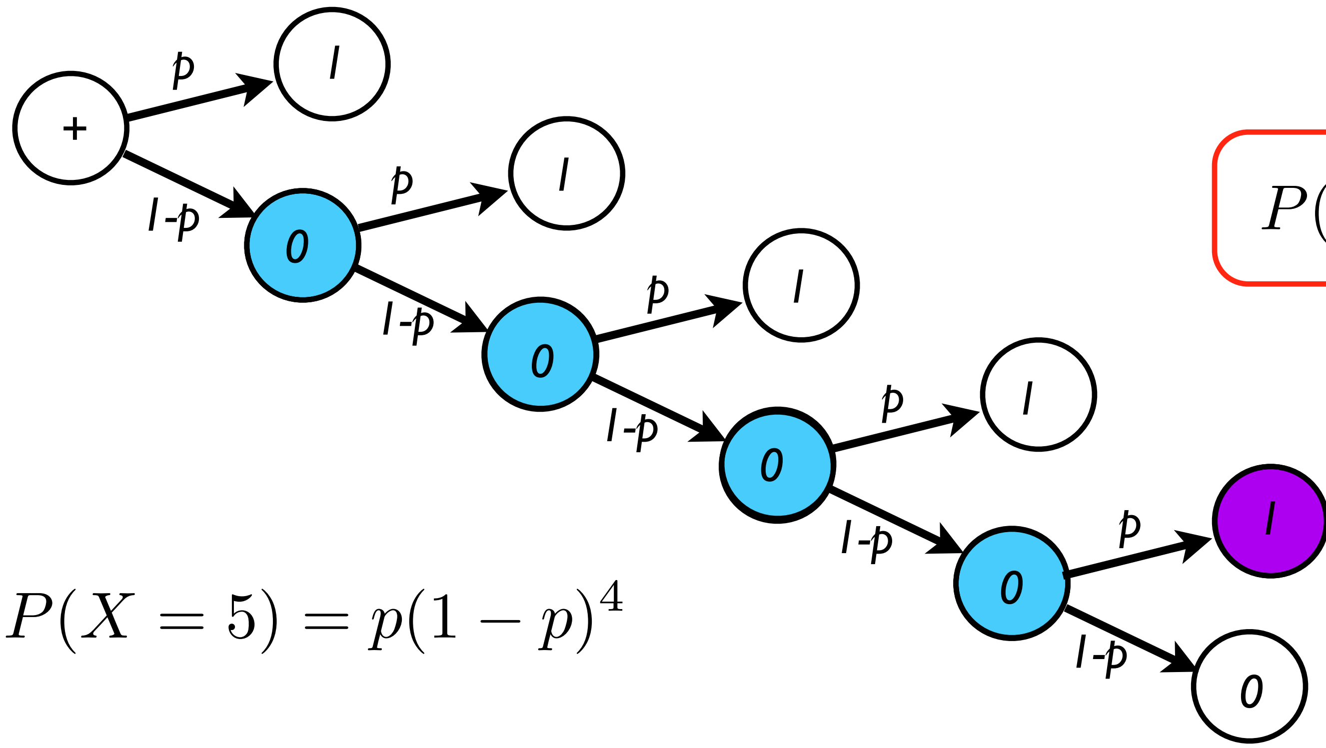
Příklad 6: Výběr uchazečů o práci: z 15ti uchazečů o zaměstnání, mezi kterými je 10 žen, vybíráme anonymně podle výsledku testu 5 osob. Jaká je pravděpodobnost, že to budou samé ženy?



X = počet mužů ve výběru

k	P(X=k)
0	0,08391608
1	0,34965035
2	0,3996004
3	0,14985015
4	0,01665002
5	0,000333

5) Geometrický model (diskrétní doba života)



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

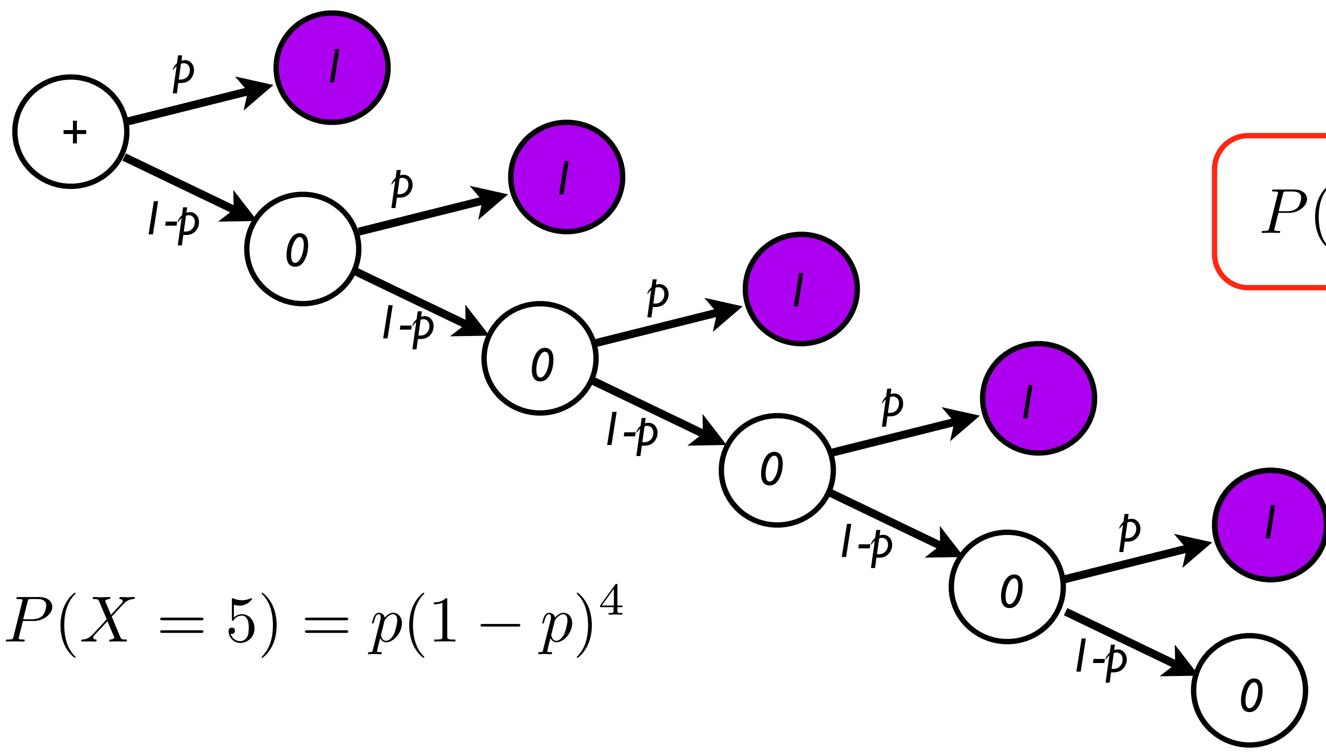
$$k = 1, 2, \ldots$$

$$P(X = 5) = p(1-p)^4$$

X je počet kroků, které je třeba učinit, aby nastal první výskyt sledovaného jevu

$$E(X) = \frac{1}{p} \qquad Var(X) = \frac{1-p}{p^2}$$

5) Geometrický model (diskrétní doba života)



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

$$k = 1, 2, \ldots$$

$$P(X = 5) = p(1-p)^4$$

X je počet kroků, které je třeba učinit, aby nastal první výskyt sledovaného jevu

$$E(X) = \frac{1}{p} \qquad Var(X) = \frac{1-p}{p^2}$$

Y je počet kroků, které předcházejí prvnímu výskytu sledovaného jevu

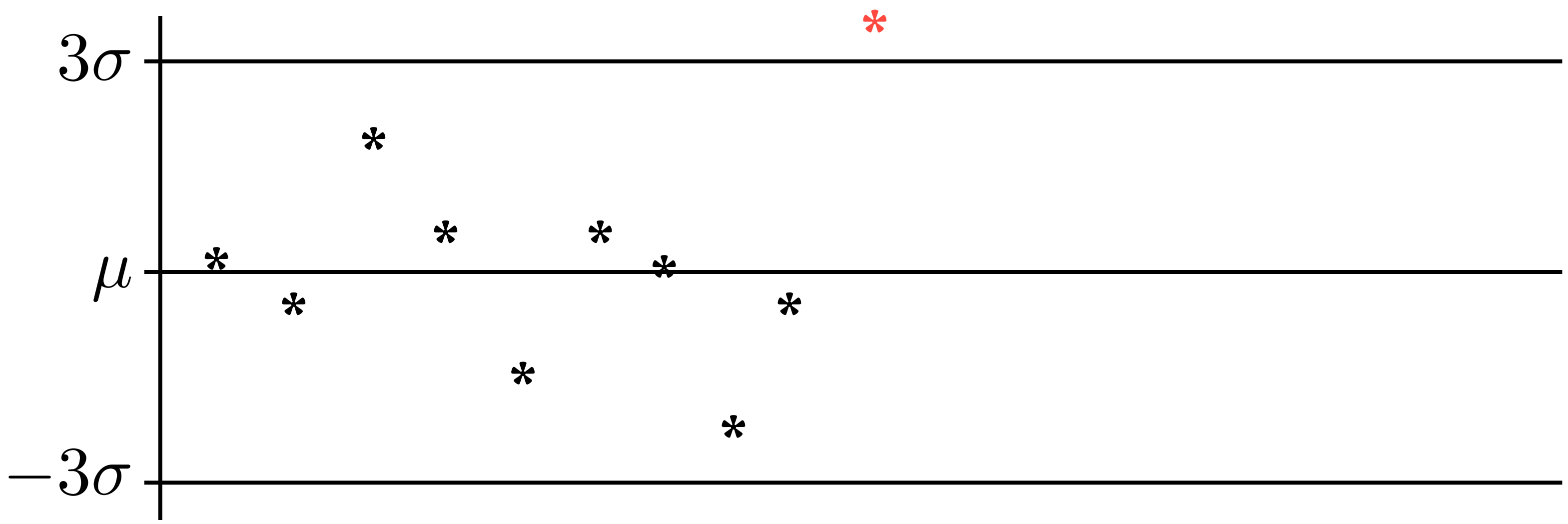
$$P(Y = k) = p(1-p)^k$$

$$k = 0, 1, \ldots$$

$$E(Y) = \frac{1-p}{p} \qquad Var(Y) = \frac{1-p}{p^2}$$

Je-li sledovaný jev porucha, potom se Y nazývá “diskrétní doba života”

5) Geometrický model (diskrétní doba života)



$$P(|X - \mu| \leq 3\sigma) = \Phi(3) - \Phi(-3) = 2\Phi(3) - 1 = 0,9973$$

$$P(|X - \mu| \geq 3\sigma) = 0,0027$$

N = počet inspekcí před signálem

$p = 0,0027$ $\quad\quad$ $E(N) = \frac{1}{p} = \frac{1}{0,0027} = 370$

Počet inspekcí před prvním falešným signálem (ARL = Average Run Length)