Pravděpodobnostní metody a matematická statistika

Prof. RNDr. Gejza Dohnal, CSc.

III. Náhodná veličina

https://sms.nipax.cz/pas

III. Náhodná veličina

Náhodná veličina je funkce $X : \Omega \to \mathbf{R}$

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

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

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

Distribuční funkce náhodné veličiny:

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

Distribuční funkce odpovídá na otázku: S jakou prav-
děpodobností náhodná veličina nepřekročí hodnotu x?

III. Náhodná veličina

Náhodná veličina je funkce z Ω do

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

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

diskrétní náhodná veličina

spojitá náhodná veličina

Diskrétní náhodná veličina:

obor hodnot = $\{x_1, x_2, \ldots, x_n, \ldots\}$

pravděpodobnostní funkce: $P(x_i) = p_i,\ i=1, 2, \ldots$

vždycky musí platit, že $p_i \in \langle 0, 1 \rangle, \quad \sum_{i=1}^{\infty} p_i = 1$

potom je $F(x) = P(X \leq x) = \sum_{i=:x_i \leq x} p_i, \quad x \in \mathbf{R}$

III. Náhodná veličina

Spojitá náhodná veličina:

obor hodnot = (a,b) či $\langle a, \infty)$ nebo celé

hustota pravděpodobnosti

nutně platí: $P(x) = 0, \ \forall x \in (a, b)$

zavádíme nezápornou, zpravidla spojitou funkci $f(x)$ tak, že platí:

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

vždycky musí platit, že $\displaystyle\int_{-\infty}^{+\infty} f(t)\mathrm{d}t = 1$

III. Charakteristiky náhodné veličiny

Kvantily:

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

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

α - kvantil

je-li $F(x)$ spojitá a prostá, potom je $x_\alpha = F^{-1}(\alpha)$

v případě diskrétní (po částech konstantní) $F(x)$ je to taková maximální hodnota veličiny X, pro niž platí $P(X \leq x_\alpha) \geq \alpha$ a zároveň $P(X < x_\alpha) \leq \alpha$.

Patří sem: minimum, medián, maximum, dolní a horní decily, horní a dolní kvartily, ….

III. Charakteristiky náhodné veličiny

Momenty:

střední hodnota

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

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

Vlastnosti střední hodnoty:

- aditivita: $E(X+Y) = E(X) + E(Y)$
- homogenita: $E(kX) = kE(X), \quad k \in \mathbf{R}$

linearita

Střední hodnota tvoří jakési “těžiště” náhodné veličiny.

III. Charakteristiky náhodné veličiny

Momenty:

k-tý obecný moment náhodné veličiny $m_k = E(X^k)$

k-tý centrální moment náhodné veličiny $\nu_k = E(X - EX)^k$

rozptyl (míra variability) $Var(X) = E(X - EX)^2$

šikmost (míra symetrie) $Skew(X) = \dfrac{E(X - EX)^3}{(Var(X))^{3/2}}$

špičatost (míra koncentrace) $Kurt(X) = \dfrac{E(X - EX)^4}{(Var(X))^2}$

III. Charakteristiky náhodné veličiny

Další charakteristiky

modus, rozpětí, typ rozdělení, …

Pravděpodobnostní metody a matematická statistika

Prof. RNDr. Gejza Dohnal, CSc.

IV. Pravděpodobnostní modely

https://sms.nipax.cz/pas

Alternativní rozdělení

$$X \begin{array}{l} \xrightarrow{\;p\;} 1 \\ \xrightarrow[\;1-p\;]{} 0 \end{array}$$

přibližně v 100.p% případů nastane výsledek 1

přibližně v 100.(1-p)% případů nastane výsledek 0

střední hodnota X: $E(X) = p.1 + (1-p).0 = p$

rozptyl X: $Var(X) = E\left(X^2\right) - \left(E(X)\right)^2 = p.1 + (1-p).0 - p^2 = p(1-p)$

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 $\bar{X} = \frac{1}{n}\sum_{i=1}^{n} X_i$

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

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

Výběr bez vracení

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

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

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?

Obecně: N prvků, mezi nimiž je M s určitou sledovanou vlastností.
Jaká je pravděpodobnost, že při výběru n prvků bez vracení vybereme k prvků se sledovanou vlastností?

počet k-tic v M prvcích

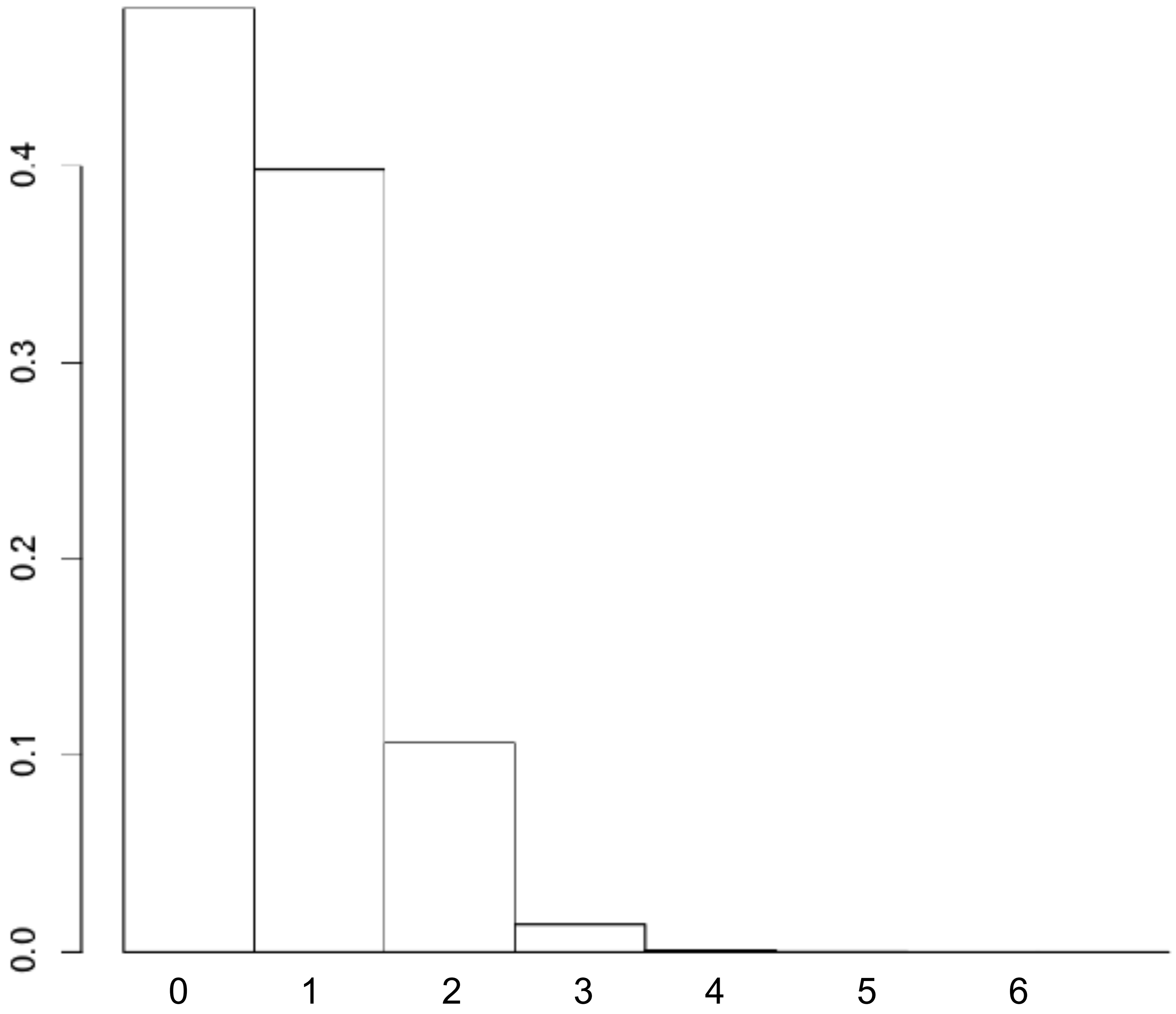
$$\boxed{P(k;n,N,M) = \frac{\binom{M}{k}\binom{N-M}{n-k}}{\binom{N}{n}}}$$

počet zbylých (n-k)-tic z ostatních (N-M) prvků

počet všech možností = počet n-tic z N prvků

Výběr bez vracení

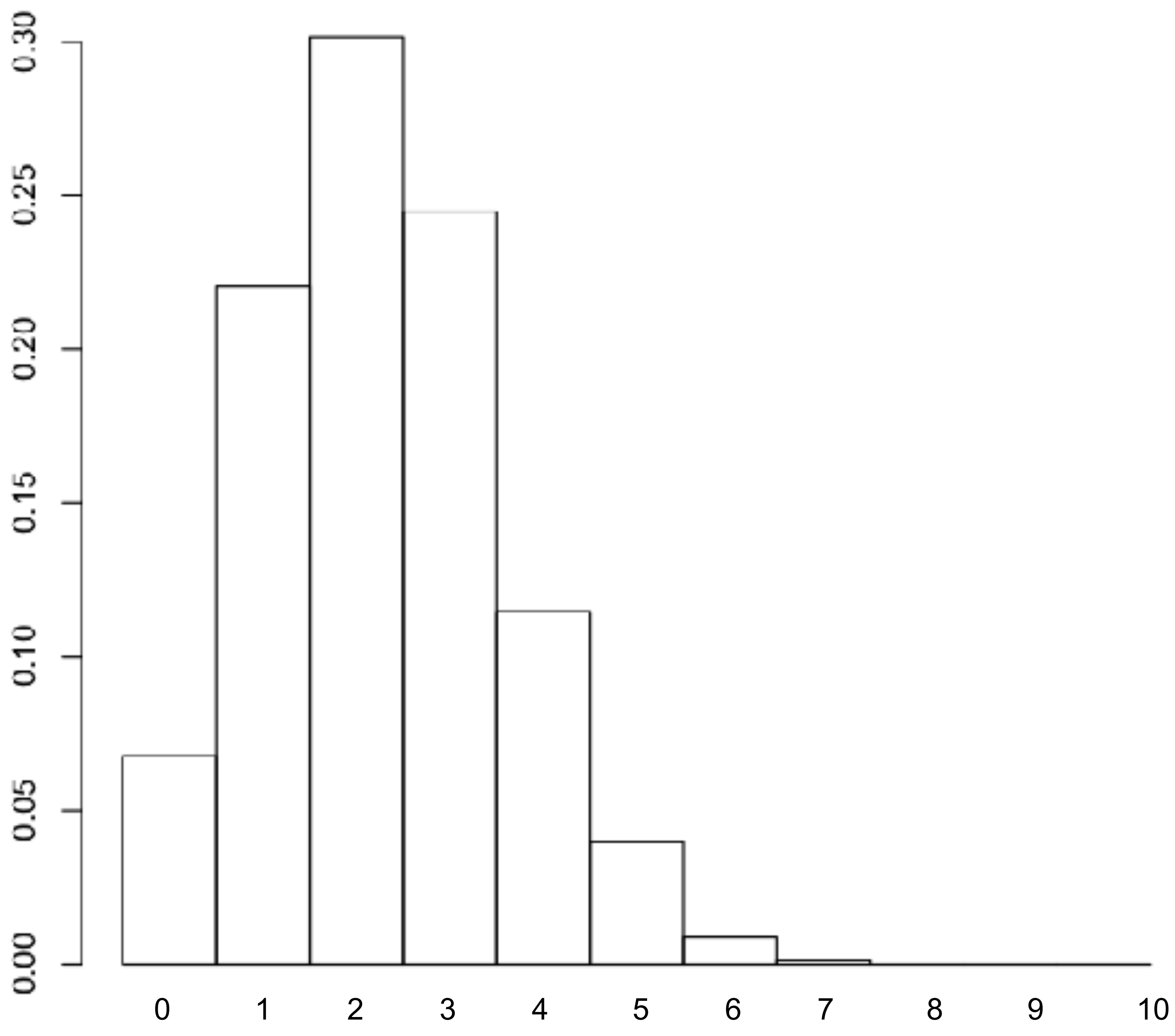
Sportka: 49 čísel, ze kterých 6 vyhrává (jsou vytaženy).
Jaká je pravděpodobnost, že při výběru 6ti čísel vybereme 4 z tažených?



Výběr bez vracení

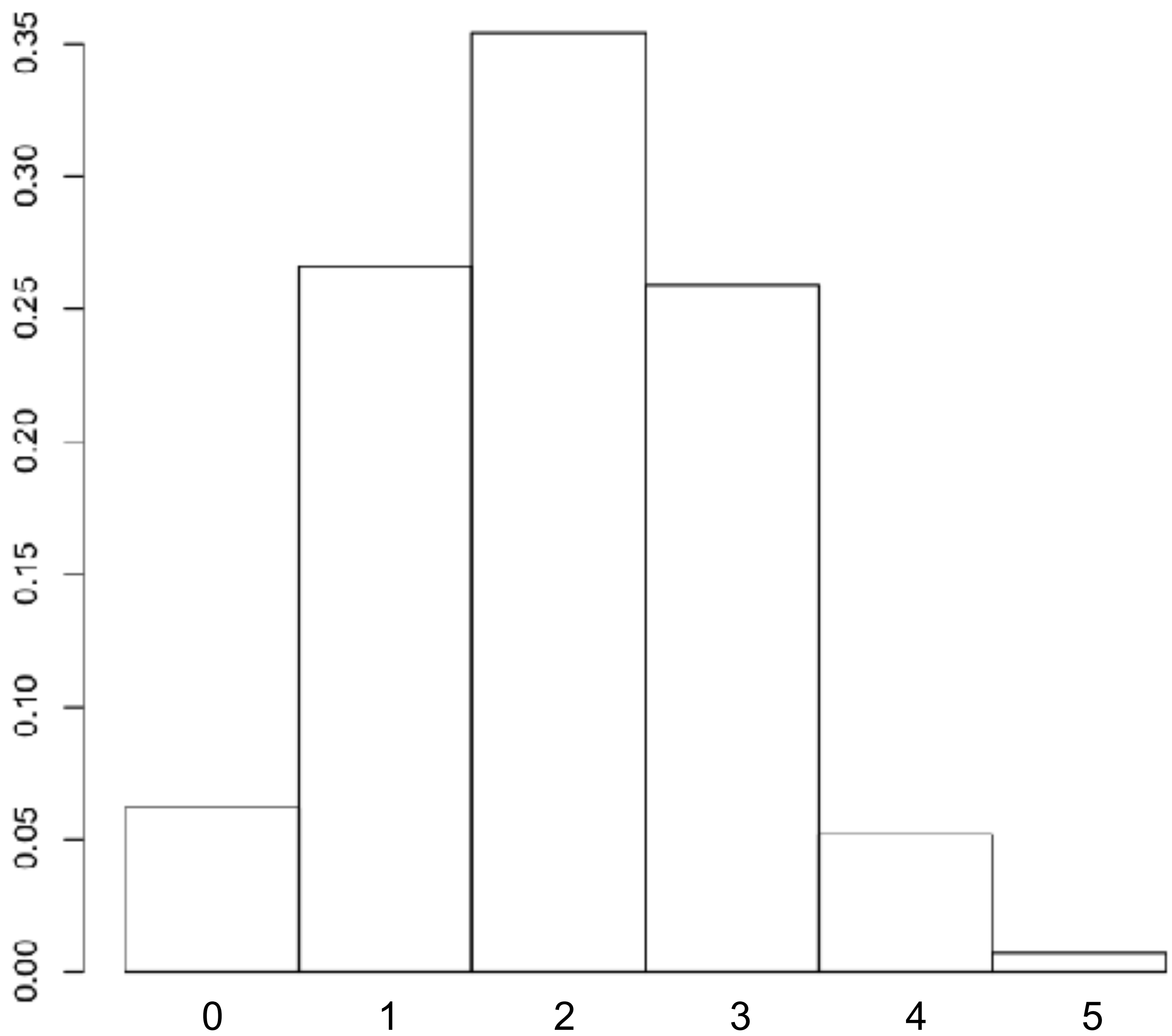
Kontrola jakosti: 1000 výrobků, mezi nimi jsou 3% vadných.

Jaká je pravděpodobnost, že při výběru 10 výrobků vybereme alespoň 1 zmetek?



Výběr bez vracení

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?



Hypergeometrické rozdělení

$$\boxed{p(N, M, n, k) = \frac{\binom{M}{k}\binom{N-M}{n-k}}{\binom{N}{n}}}$$

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

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

Výběr s vracením

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

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?

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?

Obecně: N prvků, mezi nimiž je M s určitou sledovanou vlastností.
Jaká je pravděpodobnost, že při výběru n prvků s vracením vybereme k prvků se sledovanou vlastností?

počet k-tic v n prvcích

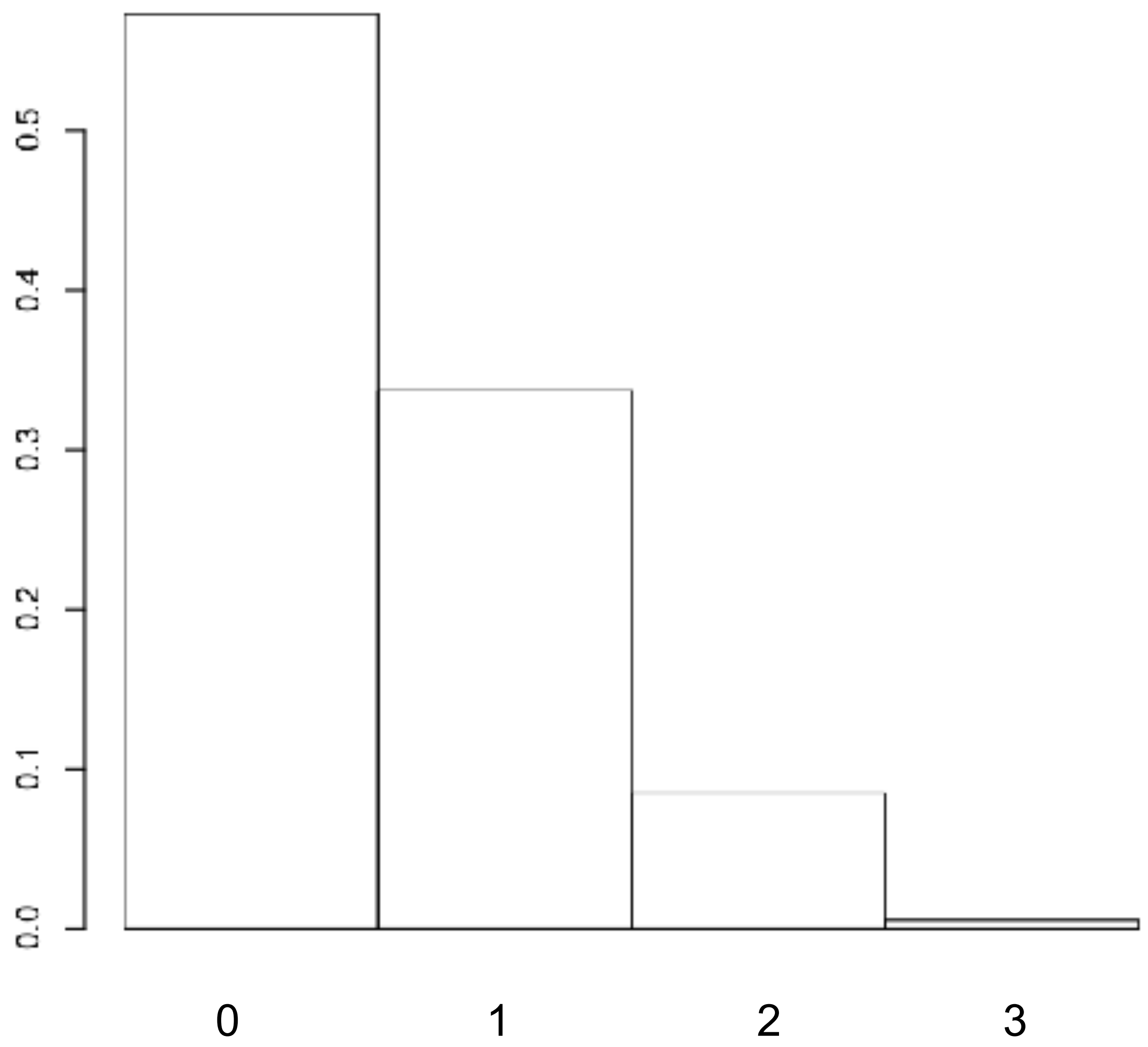
(*n*-*k*)-krát vybereme prvek s pravděpodobností (1- $\frac{M}{N}$)

$$\boxed{P(k;n,N,M) = \binom{n}{k}\left(\frac{M}{N}\right)^k\left(\frac{N-M}{N}\right)^{n-k}}$$

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

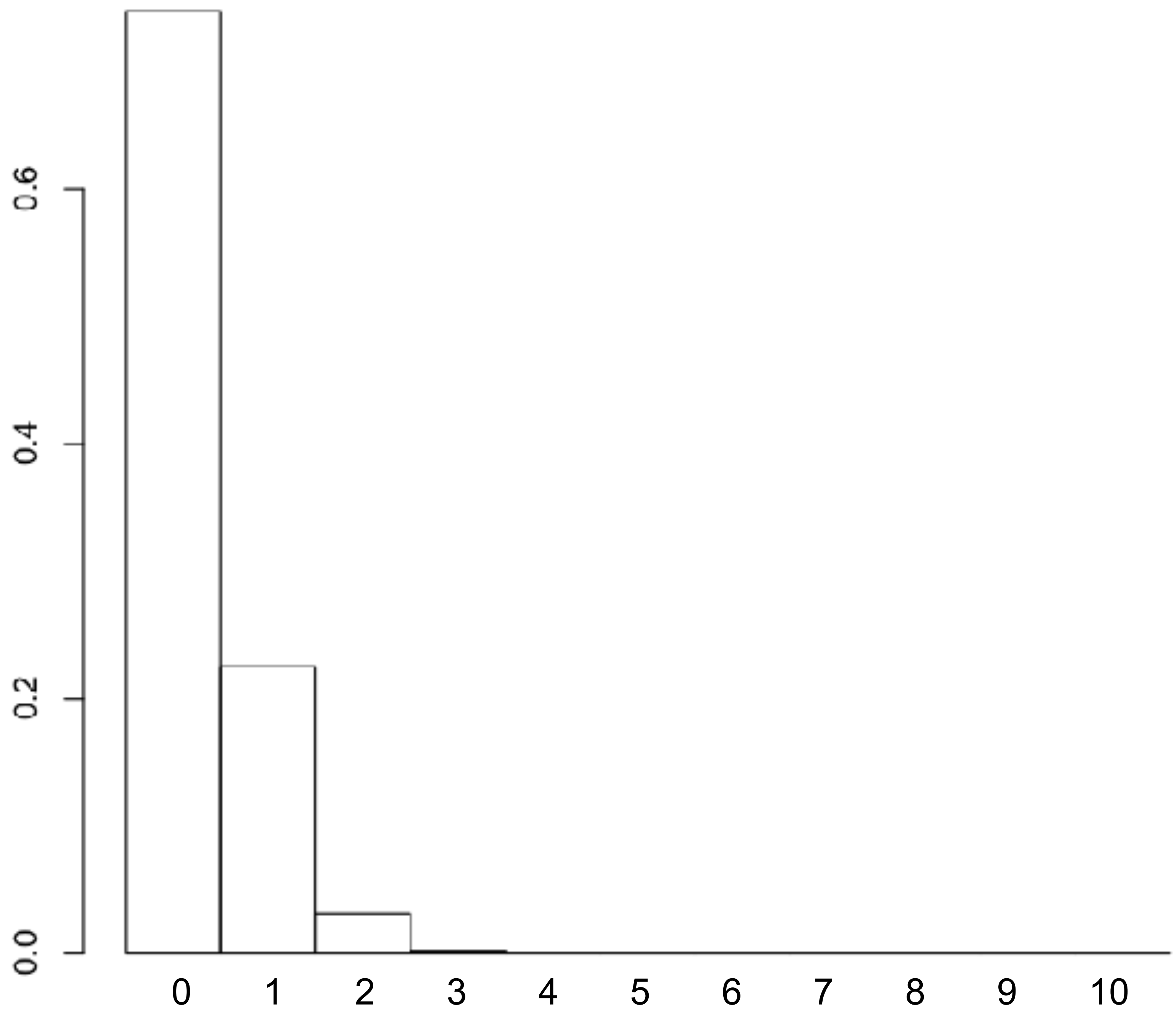
Výběr s vracením

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



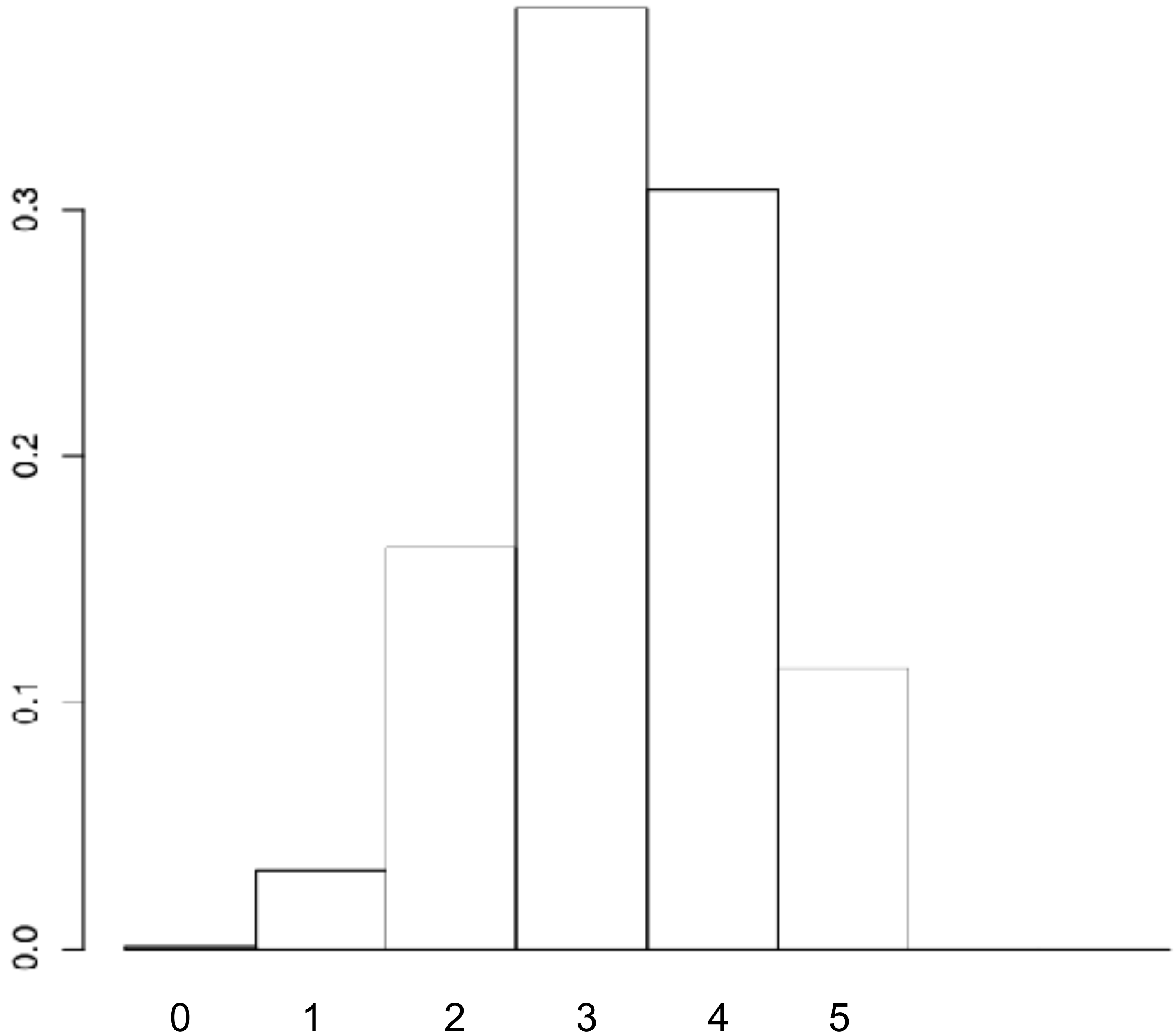
Výběr s vracením

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?



Výběr s vracením

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?



Binomické rozdělení

$$P(k;n,N,M) = \binom{n}{k}\left(\frac{M}{N}\right)^k\left(\frac{N-M}{N}\right)^{n-k}$$

$$N = 1,2,\ldots, \quad M \le N, \quad n \le N, \quad max(0,n+M-N) \le k \le min(n,M)$$

Obvyklejší je tvar

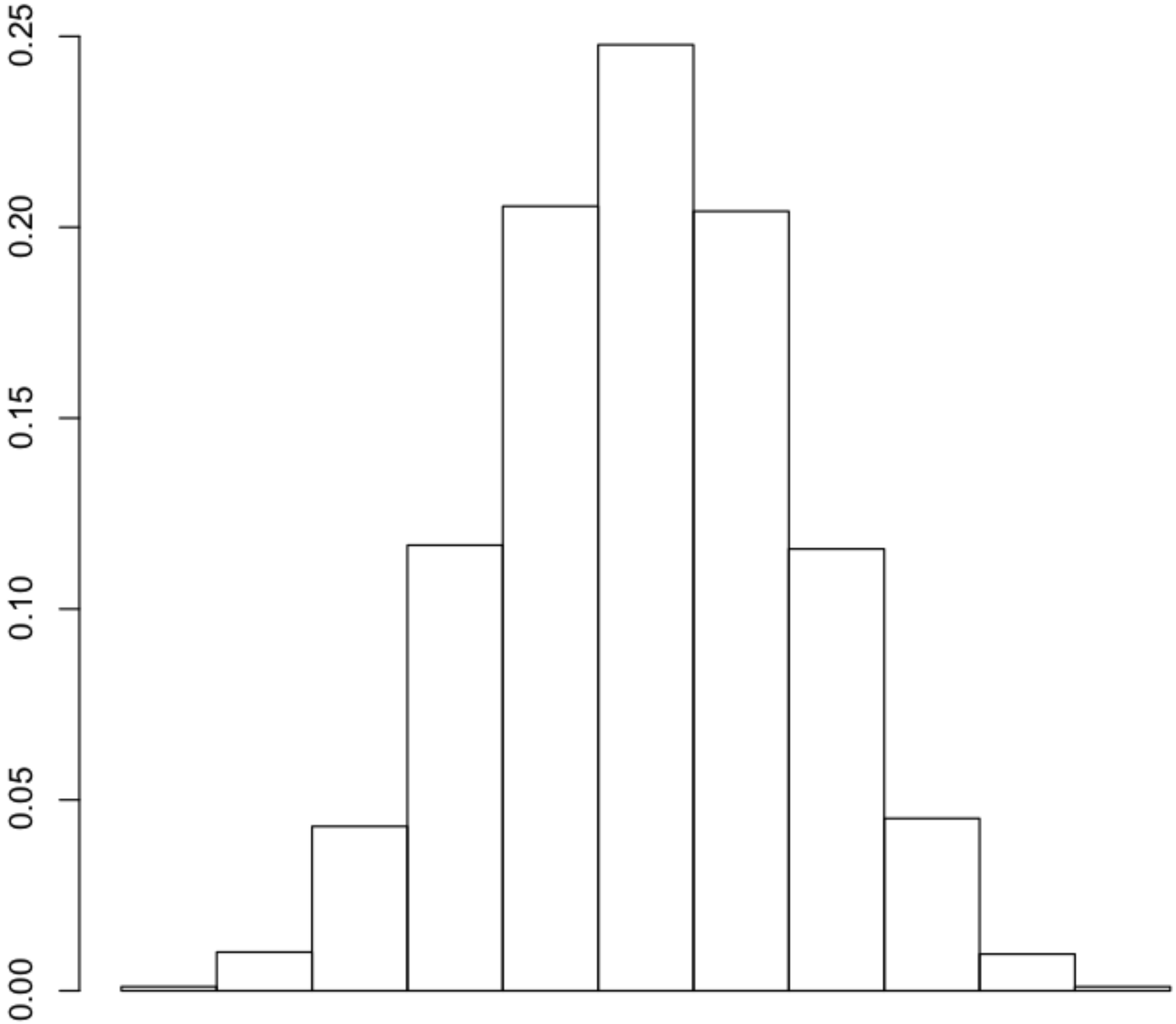
$$P(k;n,p) = \binom{n}{k}p^k(1-p)^{n-k}$$

$$n = 1,2,\ldots, \quad p \in (0,1), \quad k = 0,1,\ldots,n$$

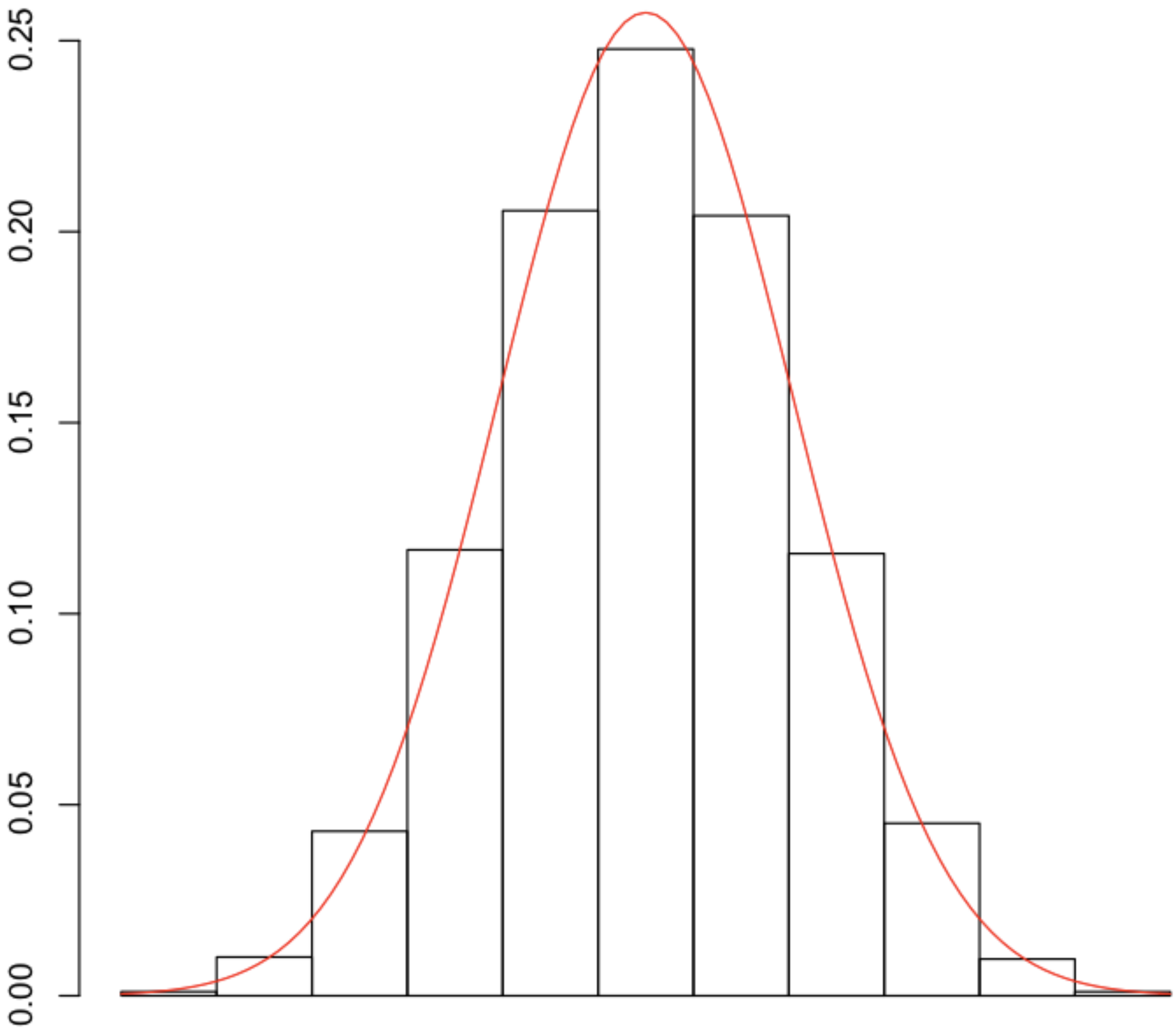
Náhodná veličina s binomickým rozdělením popisuje počet úspěchů při n nezávislých opakováních bernoulliovských pokusů s pravděpodobností úspěchu p.

$$E(X) = np \qquad\qquad Var(X) = np(1-p)$$

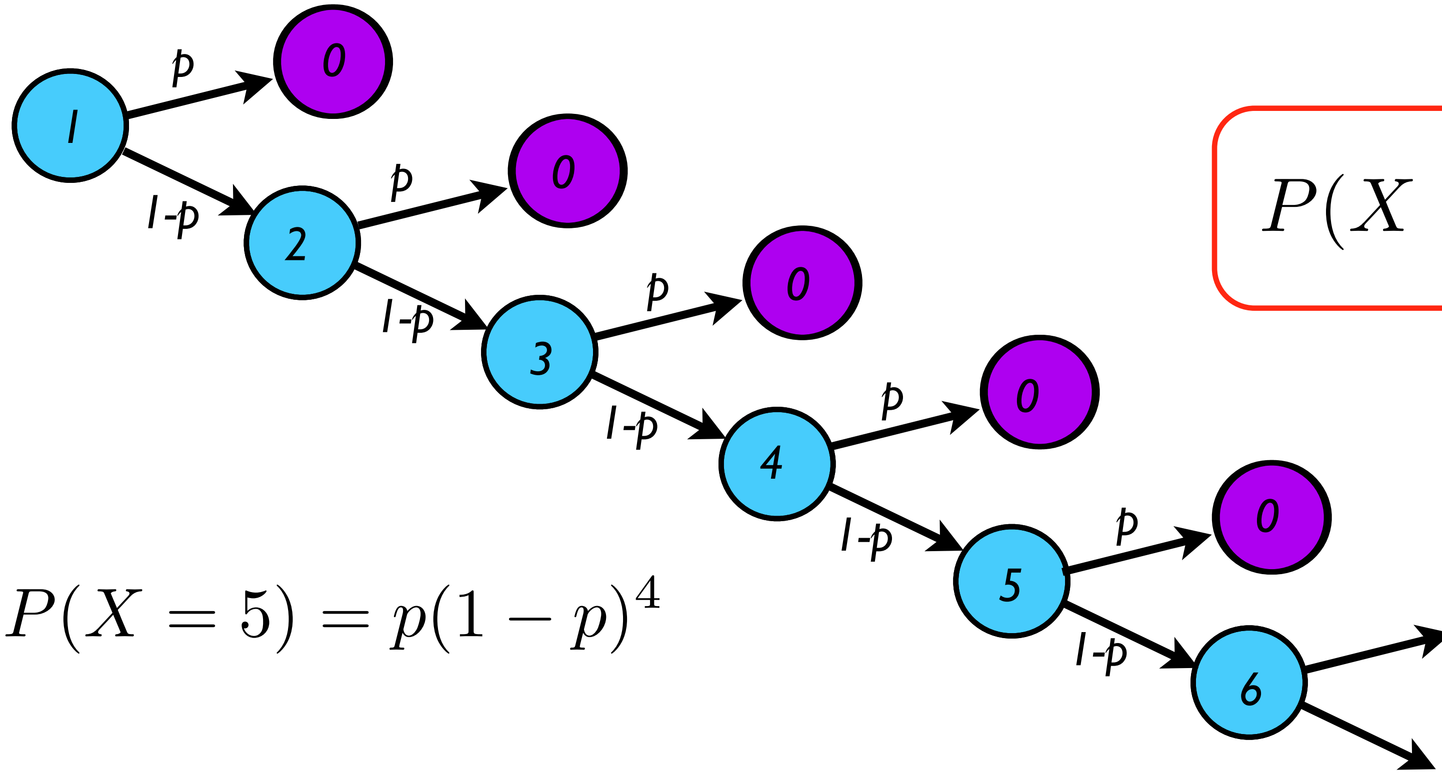
Binomické rozdělení



Binomické rozdělení



Geometrické rozdělení



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

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

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

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

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

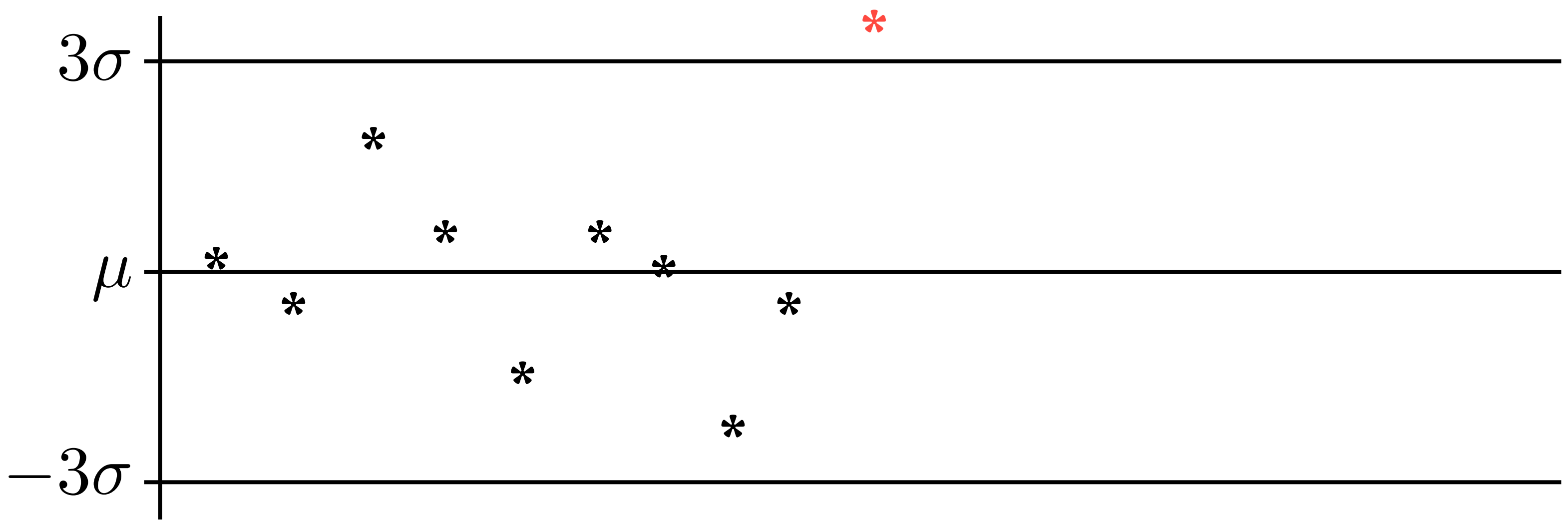
Y je počet kroků, které předcházejí prvnímu výskytu sledovaného jevu

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

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

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

Geometrické rozdělení



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

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

N = počet inspekcí před signálem

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

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