Pravděpodobnost a rozdělení pravděpodobnosti

- motivace: potřebujeme formalismus pro práci s neúplnými čísly

$$x = x_0 \pm \Delta x$$

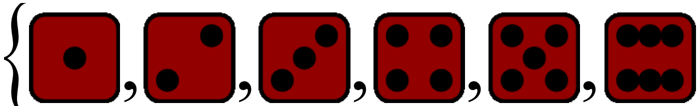
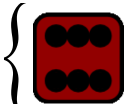
- pravděpodobnost – zavedení a vlastnosti

- rozdělení pravděpodobnosti – zavedení, vlastnosti a příklady
  - diskrétní veličiny (rozdělení pravděpodobnosti)
  - spojité veličiny (hustota pravděpodobnosti)

Náhodný jev, náhodná proměnná, pravděpodobnost

• **náhodný jev** A_E na statistickém experimentu E

- je určen vybranou množinou výsledků experimentu:

např. hrací kostka: $\{V_E\} = \{$  $\}$ $\{A_E\} = \{$  $\}$

• výsledku experimentu lze přiřadit číslo, **náhodnou proměnnou** x_E

hrací kostka: $x_E = \{1, 2, 3, 4, 5, 6\}$ $\{A_E\} = \{6\}$

• experiment typu **náhodný výběr**, N $\{V_N\}$ je konečná

- každý z výsledků experimentu N je stejně pravděpodobný

• **pravděpodobnost** jevu A_N na experimentu N (klasická definice):

Klasická definice pravděpodobnosti: $p_{A_N} = \dfrac{n_{A_N}}{n_N}$ (Laplace, 1814)

Nezávislé experimenty, pravděpodobnost

- **nezávisle** n-krát opakujeme experiment:

$$\mathrm{E} \equiv \mathrm{E}_1 \cdot \mathrm{E}_2 \cdot \ldots \cdot \mathrm{E}_n$$

$$\Omega = \mathrm{V}_{\mathrm{E}_1} \times \mathrm{V}_{\mathrm{E}_2} \times \ldots \times \mathrm{V}_{\mathrm{E}_n}$$

n_{A} = počet výskytů náhodného jevu A

relativní četnost jevu A: $X_{\mathrm{A}} = \dfrac{n_{\mathrm{A}}}{n}$

- (statistická) definice pravděpodobnosti: $p(\mathrm{A}) = \lim\limits_{n\to\infty} X_{\mathrm{A}}$

obecnější axiomatické zavedení: (jev A, možina výsledků Ω)

1. $p(\mathrm{A}) \geq 0$ pravděpodobnost libovolného jevu je nezáporná.
2. $p(\Omega) = 1$ pravděpodobnost jistého jevu = 1.
3. pro diskjunktní jevy A a B: $p(\mathrm{A} \cup B) = p(\mathrm{A}) + p(\mathrm{B})$

... a z nich plynoucí vlastnosti

Rozdělení pravděpodobnosti

diskrétní náhodná proměnná

$\Omega = \{x_1, x_2, x_3, ..., x_n\}$ konečná

$\Omega = \{x_1, x_2, x_3, ...\}$ nekonečná

Rozdělení pravděpodobnosti udává pravděpodobnost p_i, že nastane výsledek x_i

$$p(x_i) = p_i$$

Normalizační podmínka:

$$\sum_{i=1}^{n} p_i = 1$$ konečná

$$\sum_{i=1}^{\infty} p_i = 1$$ nekonečná

Pravděpodobnost, že náhodná proměnná X bude nalezena v intervalu $(-\infty, x\rangle$

$$F(x) \equiv p(X \leq x)$$ ... **distribuční funkce**

Hustota pravděpodobnosti spojité proměnné

spojitá náhodná proměnná $x \in \Omega$ Ω je nespočetná !

Rozdělení pravděpodobnosti → **Hustota pravděpodobnosti** $f(x)$

... udává pravděp. p, že se výsledek nachází v infinitezimálním intervalu $\langle x_0, x_0 + \mathrm{d}x \rangle$

$$p \equiv f(x_0)\mathrm{d}x$$

srov. s diskrétní: $p(x_i) = p_i$

Distribuční funkce $F(x) \equiv \int_{-\infty}^{x} f(t)\mathrm{d}t$

srov. s diskrétní: $F(x) \equiv p(X \leq x)$

Pravděpodobnost že $x \in \langle a, b \rangle$ je:

$$p = \int_{a}^{b} f(x)\mathrm{d}x = F(b) - F(a)$$

srov. s diskrétní:

$$p(\mathrm{x} \in \langle a, b \rangle) = \sum_{x \geq a}^{x \leq b} p_i = F(b) - F(a)$$

Normalizační podmínka: $\int_{-\infty}^{\infty} f(x)\mathrm{d}x = 1$

srov. s diskrétní: $\sum_{i=1}^{\infty} p_i = 1$

Rovnoměrné rozdělení

- množina výsledků: $\Omega = \{x_1, x_2, ..., x_n\}$
- každý výsledek x_i je stejně pravděpodobný:

$$p(\mathrm{A}_i) = p \qquad \forall i \in \langle 1, \dots, n \rangle$$

- normalizační podmínka: $1 = \sum_{i=1}^{n} p(\mathrm{A}_i) = \sum_{i=1}^{n} p = np \quad \Rightarrow \quad p = \frac{1}{n}$

- obecně pro interval $\langle a, b \rangle$:

$$\mathrm{p}_i = \frac{1}{n} = \frac{1}{b - a + 1}$$

- distribuční funkce:

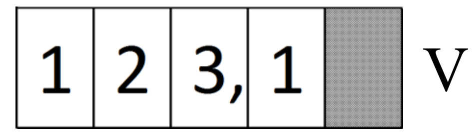
i < a	0
a ≤ i ≤ b	$\frac{\lvert i \rvert - a + 1}{n}$
i > b	1

- je-li náhodná veličina spojitá:

$$f(x) = \begin{cases} \frac{1}{b-a} & x \in \langle a, b \rangle \\ 0 & x \notin \langle a, b \rangle \end{cases} \xrightarrow{\int f(x)} F(x) = \begin{cases} 0 & x < a \\ \frac{x-a}{b-a} & x \in \langle a, b \rangle \\ 1 & x > b \end{cases}$$

Rovnoměrné rozdělení – příklad: zaokrouhlovací chyba

• 5-místný displej, poslední pozice „rozbitá“:
Jaké se dopouštíme (dodatečné) chyby?



Histogram

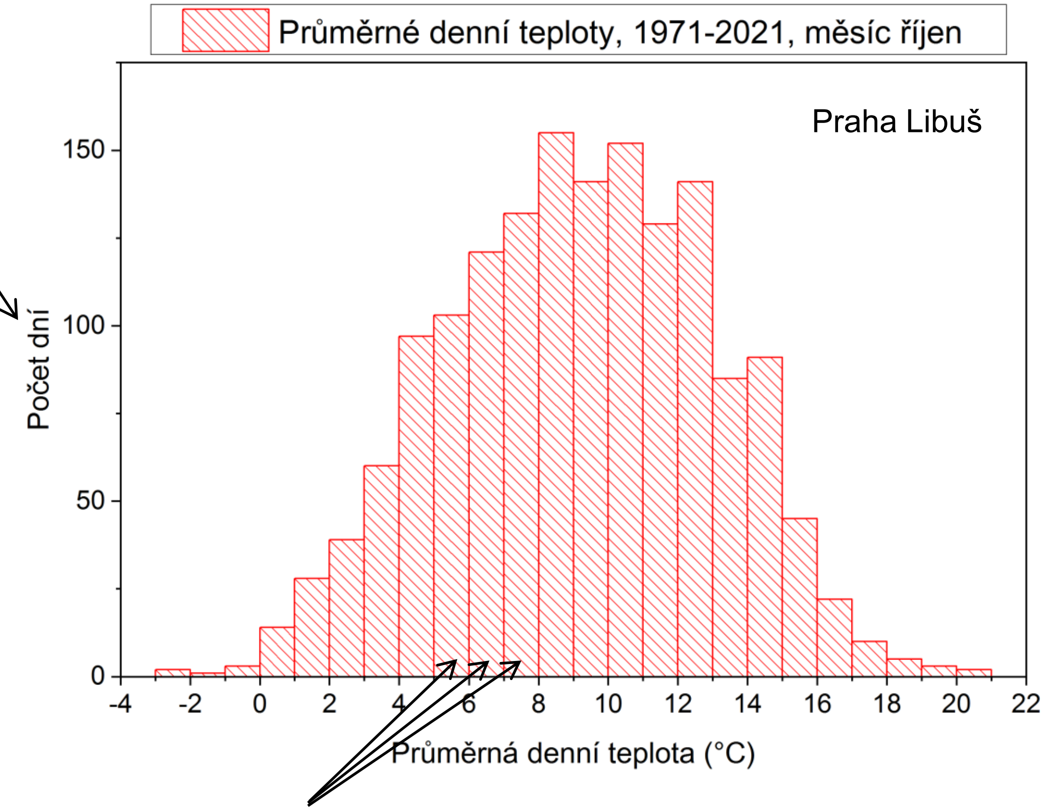
Jak získat hustotu pravděpodobnosti $f(x)$ z dat?

Český hydrometeorologický ústav, www.chmi.cz

Základní informace o stanici

zeměpisná šířka	50° 00' 28'' N
zeměpisná délka	14° 26' 49'' E
nadmořská výška	302 m
počátek pravidelných měření	1. 1. 1971

četnost



biny $x_i, \ i = 1 \dots m$

relativní četnost $\quad r_i = \dfrac{n_i}{N} \qquad N = \sum_{i=1}^{m} n_i$

šířka binu $\quad \Delta_i = x_{i+1} - x_i$

Normovaný histogram

- relativní četnosti se vydělí šířkou binu:

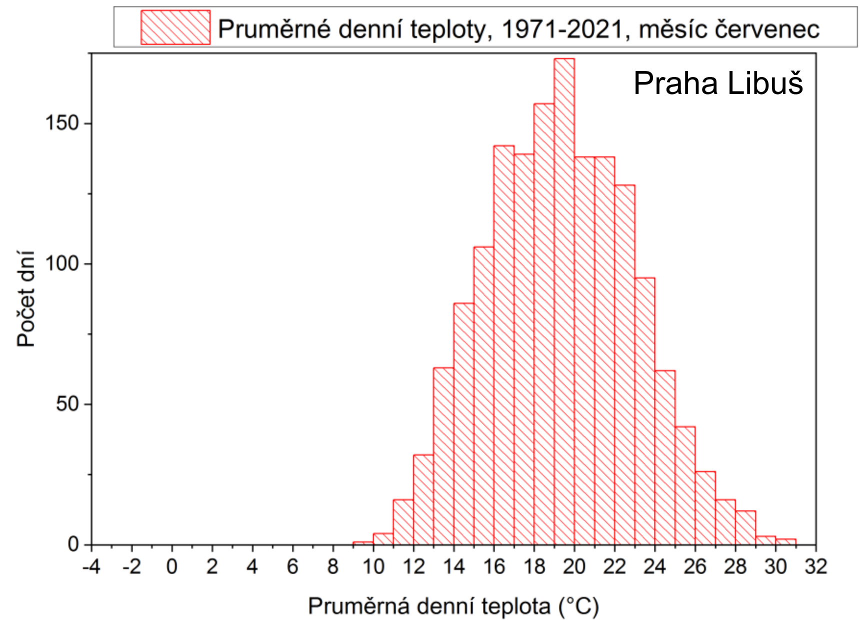
$$\xi_i = \frac{r_i}{\Delta_i}$$

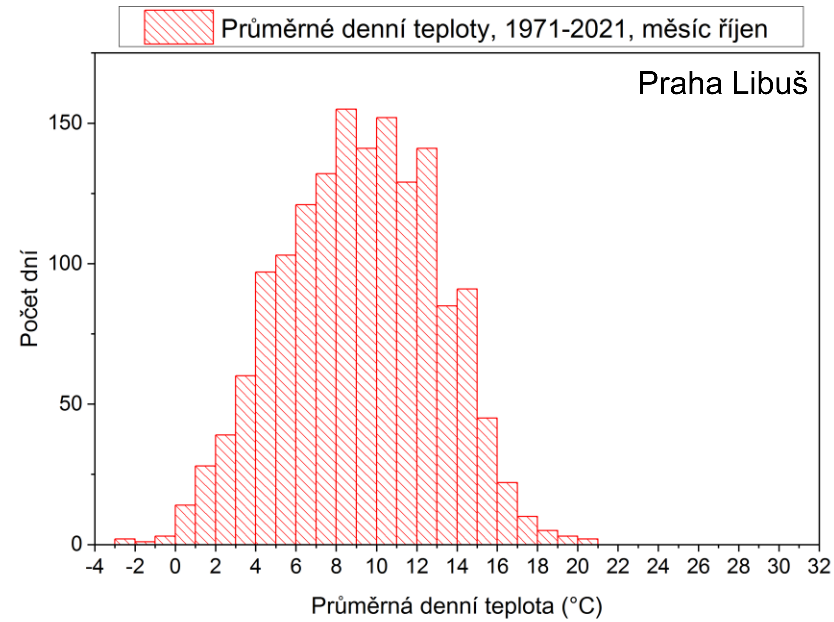
- plocha je normována: $\quad \sum_{i=1}^{m} \xi_i \Delta_i = \sum_{i=1}^{m} r_i = 1$

Hustota pravděpodobnosti: $\quad f(x) = \lim_{N\to\infty,\, \Delta_i \to 0} \xi_i$

Histogram

Jak získat hustotu pravděpodobnosti $f(x)$ z dat?

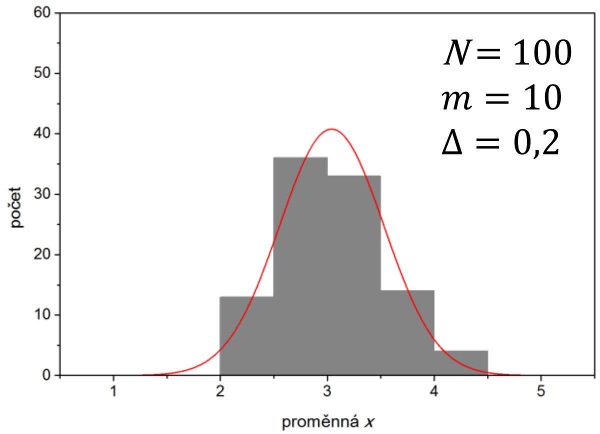


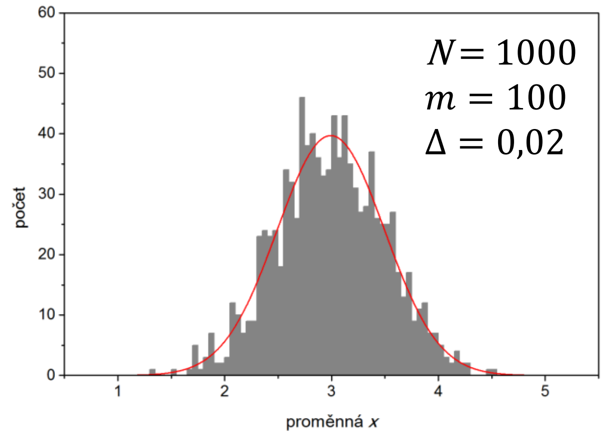


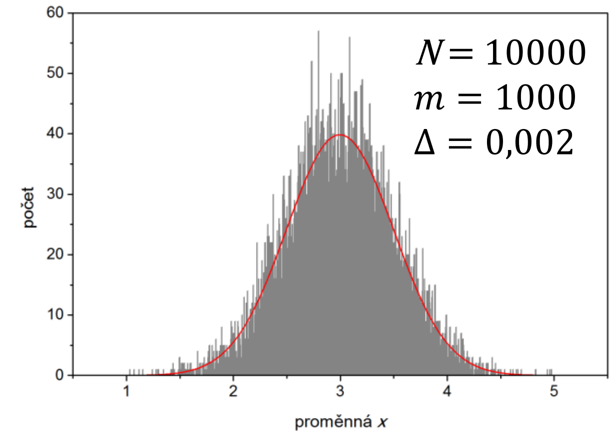
Histogram – hustota pravděpodobnosti

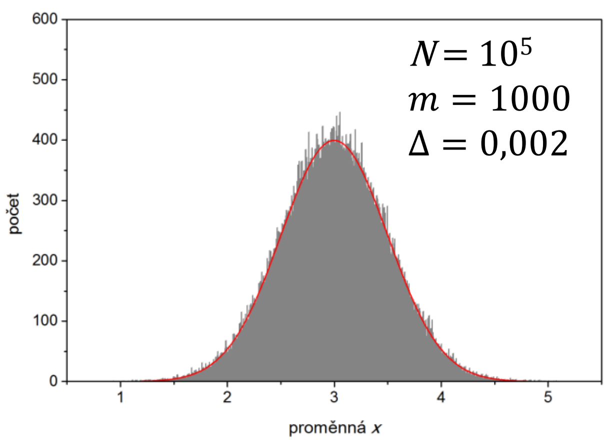
Náhodná proměnná *x* s normálním rozdělením: $f(x) = \frac{1}{\sqrt{2\pi}\sigma} e^{-\frac{(x-\mu)^2}{2\sigma^2}}$ $\mu = 3, \sigma = 0{,}5$ $N(3, 0{,}5)$

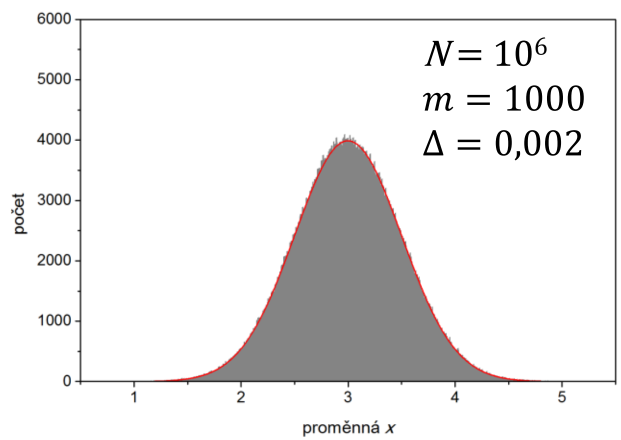
N ... počet vygenerovaných hodnot
Δ ... šířka binu
m ... počet binů

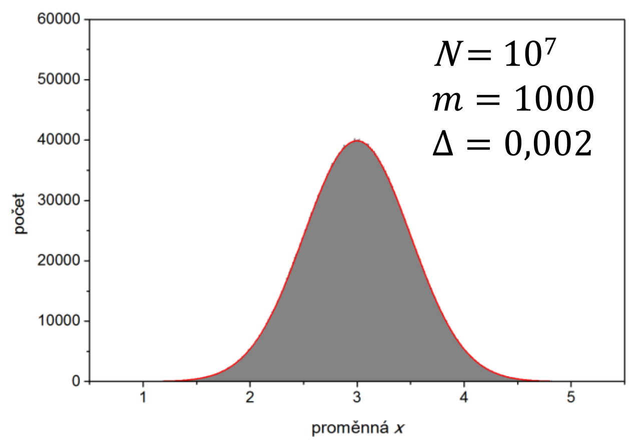












Histogram – optimální počet binů

$$\Delta = \frac{x_{\max} - x_{\min}}{m}$$

např.: Sturges (1926): $m_{\text{opt}} = \lceil \log_2 N \rceil + 1$ (zde $N = 1581$)