Centrální limitní věta

- x_i jsou nezávislé náhodné proměnné s hustotami pravděpodobnosti $f_i(x_i)$
- očekávané hodnoty $E[x_i] = \mu_i$ a rozptyly $V[x_i] = \sigma_i^2$
- potom platí: $\boxed{y = \sum_{i=1}^{n} x_i \quad \text{pro } n \to \infty \quad \text{je} \quad y \in N\left(\sum_{i=1}^{n} \mu_i \, , \sqrt{\sum_{i=1}^{n} \sigma_i^2}\right)}$
- neboli $\boxed{z = \frac{y - \sum_{i=1}^{n} \mu_i}{\sqrt{\sum_{i=1}^{n} \sigma_i^2}} \quad \text{pro } n \to \infty \quad \text{je} \quad z \in N(0,1)}$

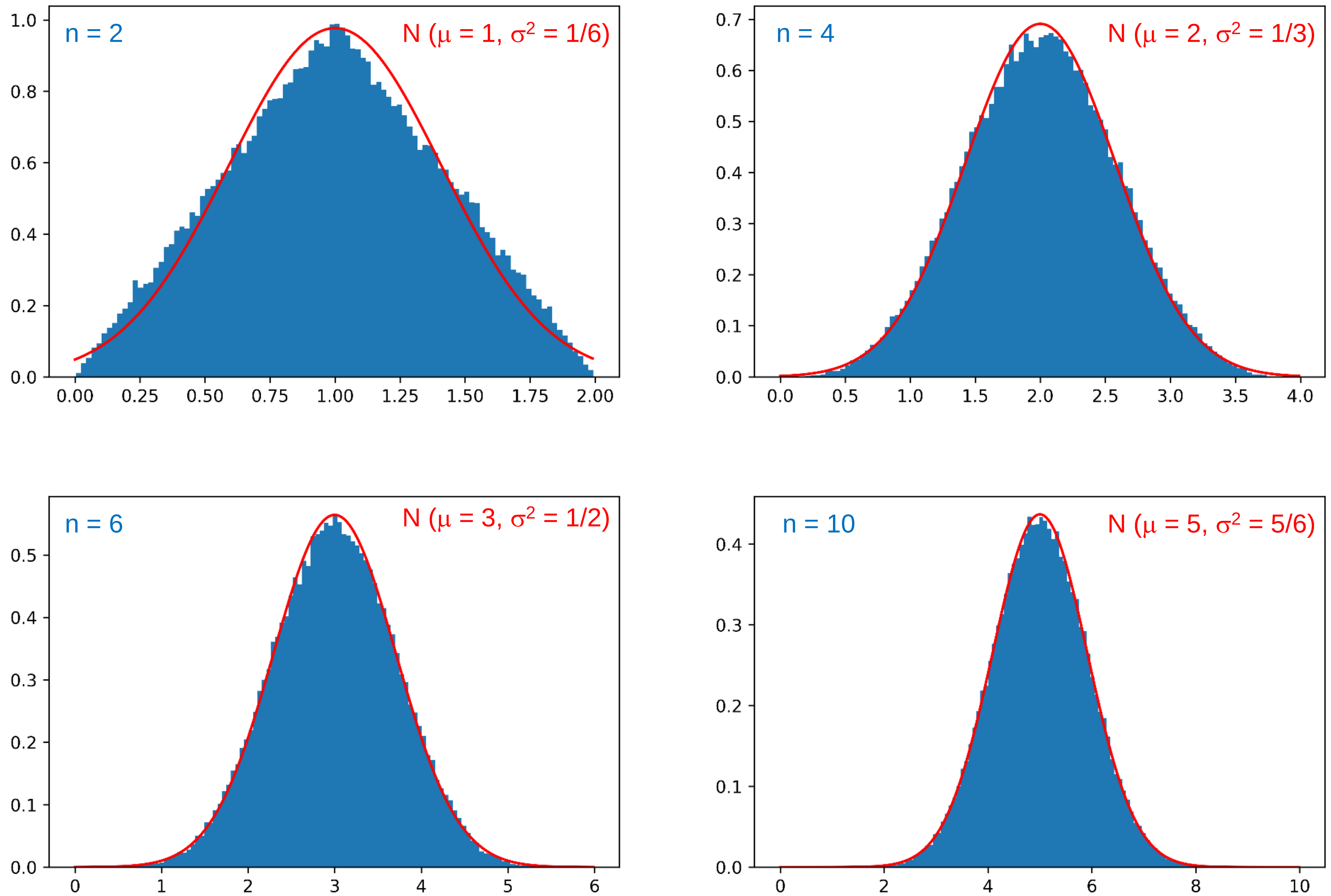
Centrální limitní věta

$$x_i \in U(0,1)$$

$$y = \sum_{i=1}^{n} x_i$$

$$\mu = \sum_{i=1}^{n} \mu_i = \frac{n}{2}$$

$$\sigma^2 = \sum_{i=1}^{n} \sigma_i^2 = \frac{n}{12}$$



Hustota pravděpodobnosti – případ dvou proměnných

- **marginální hustoty pravděpodobnosti**

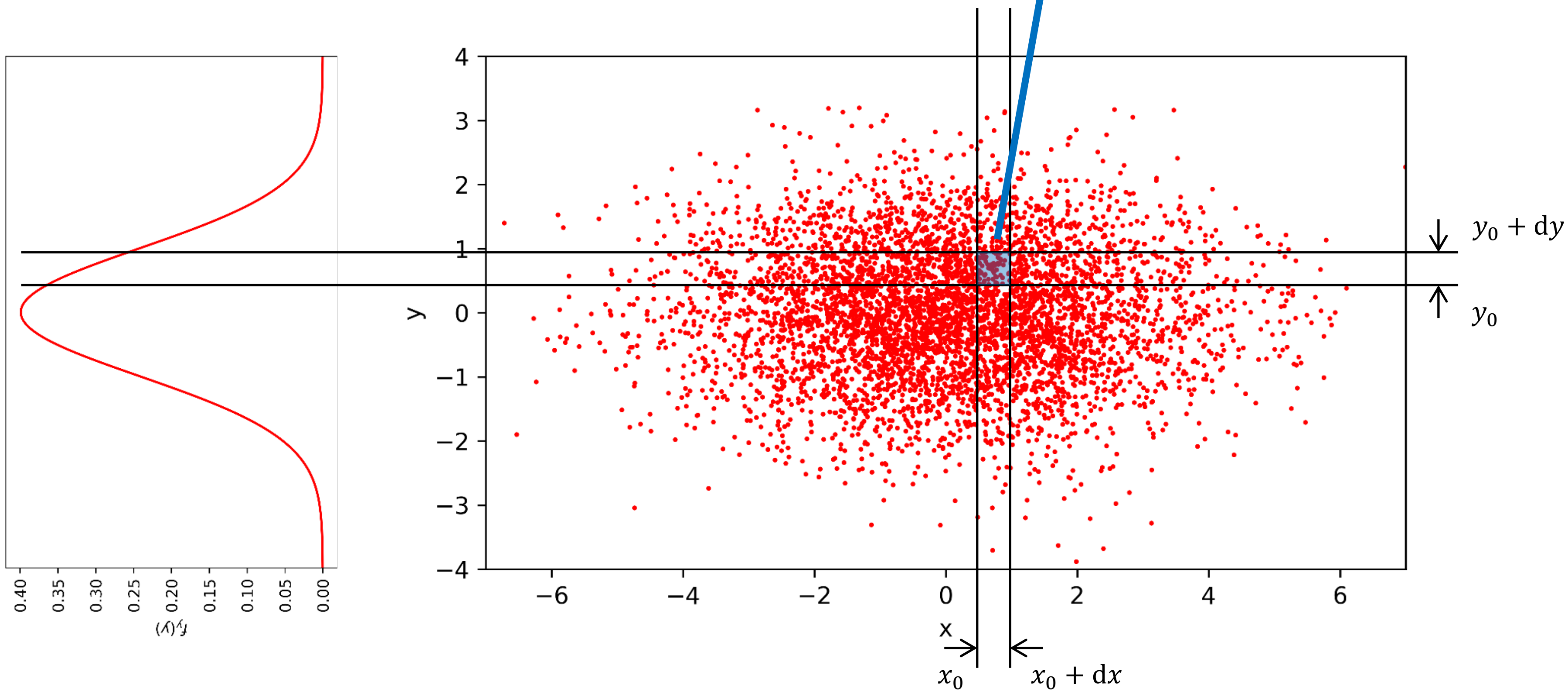
$$f_x(x) = \int_{-\infty}^{\infty} f(x,y)\mathrm{d}y$$

$$f_y(y) = \int_{-\infty}^{\infty} f(x,y)\mathrm{d}x$$

A: $x \in [x_0, x_0 + \mathrm{d}x]$

B: $y \in [y_0, y_0 + \mathrm{d}y]$

$$P(A \cap B) = f(x_0, y_0)\mathrm{d}x\mathrm{d}y$$



Očekávaná hodnota – případ dvou proměnných

- operátor **očekávané hodnoty**

očekávaná hodnota náhodné proměnné x

$$\boxed{\begin{aligned}\mu_x \equiv E[x] &= \int_{-\infty}^{\infty}\int_{-\infty}^{\infty} x f(x,y)\,\mathrm{d}x\mathrm{d}y \\ \mu_y \equiv E[y] &= \int_{-\infty}^{\infty}\int_{-\infty}^{\infty} y f(x,y)\,\mathrm{d}x\mathrm{d}y\end{aligned}}$$

očekávaná hodnota náhodné proměnné y

obecně: $$E[g(x,y)] = \int_{-\infty}^{\infty}\int_{-\infty}^{\infty} g(x,y) f(x,y)\,\mathrm{d}x\mathrm{d}y$$

Rozptyl – případ dvou proměnných

- operátor **rozptylu**

rozptyl náhodné proměnné x

$$\boxed{\begin{aligned}\sigma_x^2 \equiv V[x] &= E[(x-\mu_x)^2] = \int_{-\infty}^{\infty}\int_{-\infty}^{\infty}(x-\mu_x)^2 f(x,y)\,\mathrm{d}x\mathrm{d}y \\ \sigma_y^2 \equiv V[y] &= E\left[\left(y-\mu_y\right)^2\right] = \int_{-\infty}^{\infty}\int_{-\infty}^{\infty}\left(y-\mu_y\right)^2 f(x,y)\,\mathrm{d}x\mathrm{d}y\end{aligned}}$$

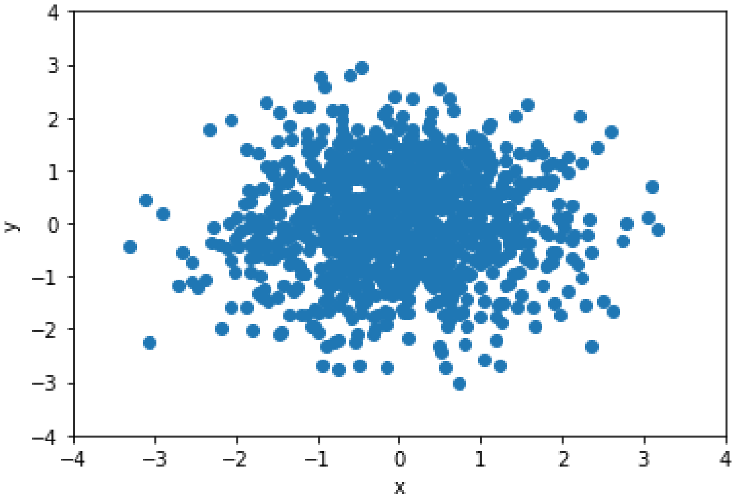
rozptyl náhodné proměnné y

Kovariance a korelace náhodných proměnných

- **kovariance** a **korelace**

minimální smysluplná informace

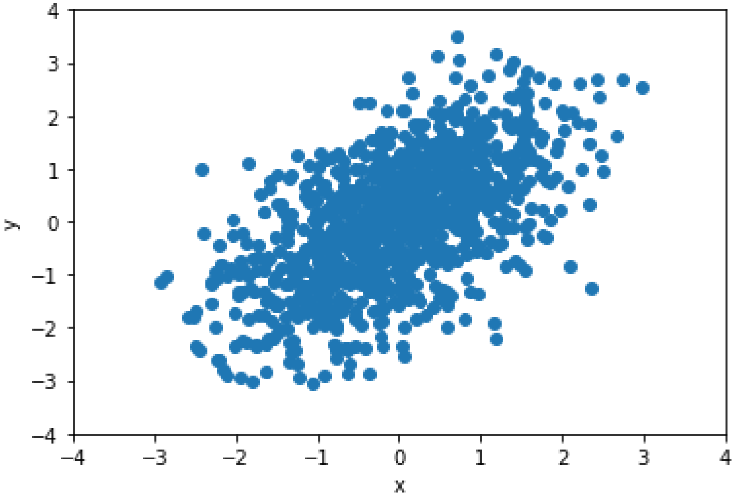
o náhodných proměnných *x* a *y*:



míra polohy: μ_x a μ_y

míra disperze: σ_x a σ_y

4 čísla nestačí!



kovariance: $\mathrm{cov}(x, y)$

Kovariance a korelace náhodných proměnných

- **kovariance** náhodných proměnných x a y

$$\boxed{\operatorname{cov}(x,y) = E\big[(x-\mu_x)(y-\mu_y)\big] = E[xy] - E[x]E[y]}$$

kovarianční matice

$$V = \begin{pmatrix} V[x] & \operatorname{cov}(x,y) \\ \operatorname{cov}(y,x) & V[y] \end{pmatrix} = \begin{pmatrix} \sigma_x^2 & \operatorname{cov}(x,y) \\ \operatorname{cov}(y,x) & \sigma_y^2 \end{pmatrix}$$

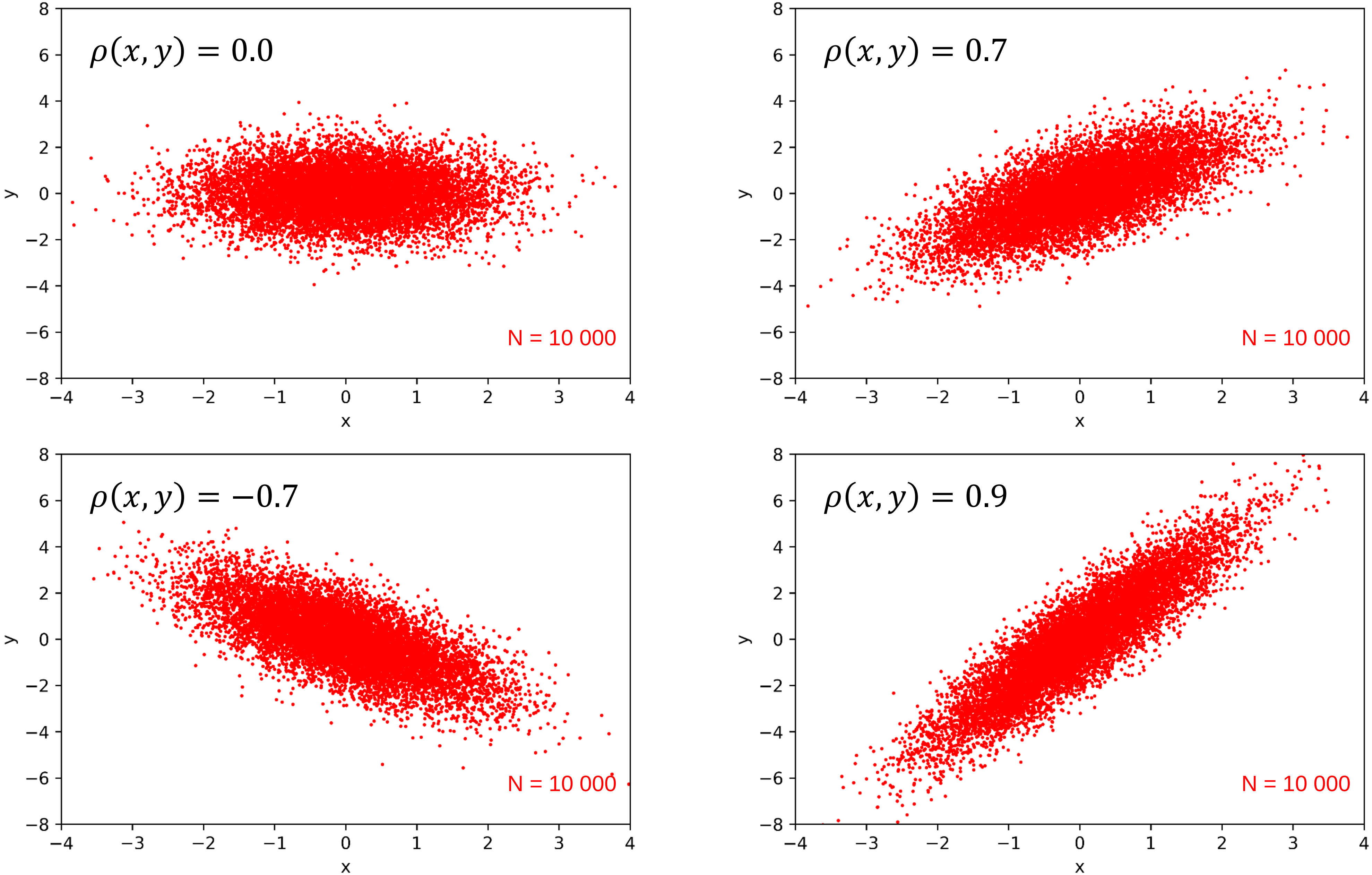
- **korelace** náhodných proměnných x a y

$$\boxed{\rho(x,y) = \frac{\operatorname{cov}(x,y)}{\sigma_x \sigma_y}}$$

korelační matice

$$R = \begin{pmatrix} \rho(x,x) & \rho(x,y) \\ \rho(y,x) & \rho(y,y) \end{pmatrix} = \begin{pmatrix} 1 & \rho(x,y) \\ \rho(y,x) & 1 \end{pmatrix}$$

Korelace náhodných proměnných



Kovariance a korelace nezávislých náhodných proměnných

- **nezávislé proměnné** x a y

$$\boxed{f(x,y) = f_x(x) f_y(y)}$$

rozptyl náhodné proměnné x

$$V[x] = E[(x-\mu_x)^2] = \int_{\infty}^{\infty} (x-\mu_x)^2 f_x(x)\mathrm{d}x \int_{\infty}^{\infty} f_y(y)\mathrm{d}y = \int_{\infty}^{\infty} (x-\mu_x)^2 f_x(x)\mathrm{d}x = \sigma_x^2$$

rozptyl náhodné proměnné y

$$V[y] = E\left[\left(y-\mu_y\right)^2\right] = \int_{\infty}^{\infty} f_x(x)\mathrm{d}x \int_{\infty}^{\infty} \left(y-\mu_y\right)^2 f_y(y)\mathrm{d}y = \int_{\infty}^{\infty} \left(y-\mu_y\right)^2 f_y(y)\mathrm{d}y = \sigma_y^2$$

kovariance náhodných proměnných x a y

$$\mathrm{cov}(x,y) = E\left[(x-\mu_x)\left(y-\mu_y\right)\right] = \int_{\infty}^{\infty} (x-\mu_x) f_x(x)\mathrm{d}x \int_{\infty}^{\infty} \left(y-\mu_y\right) f_y(y)\mathrm{d}y$$

$$= E[(x-\mu_x)]E\left[\left(y-\mu_y\right)\right] = 0$$

Kovariance a korelace nezávislých náhodných proměnných

- **nezávislé proměnné** x a y

$$\boxed{f(x,y) = f_x(x) f_y(y)}$$

kovarianční matice

$$V = \begin{pmatrix} \sigma_x^2 & 0 \\ 0 & \sigma_y^2 \end{pmatrix}$$

$$\operatorname{cov}(x,y) = 0$$

korelační matice

$$R = \begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix}$$

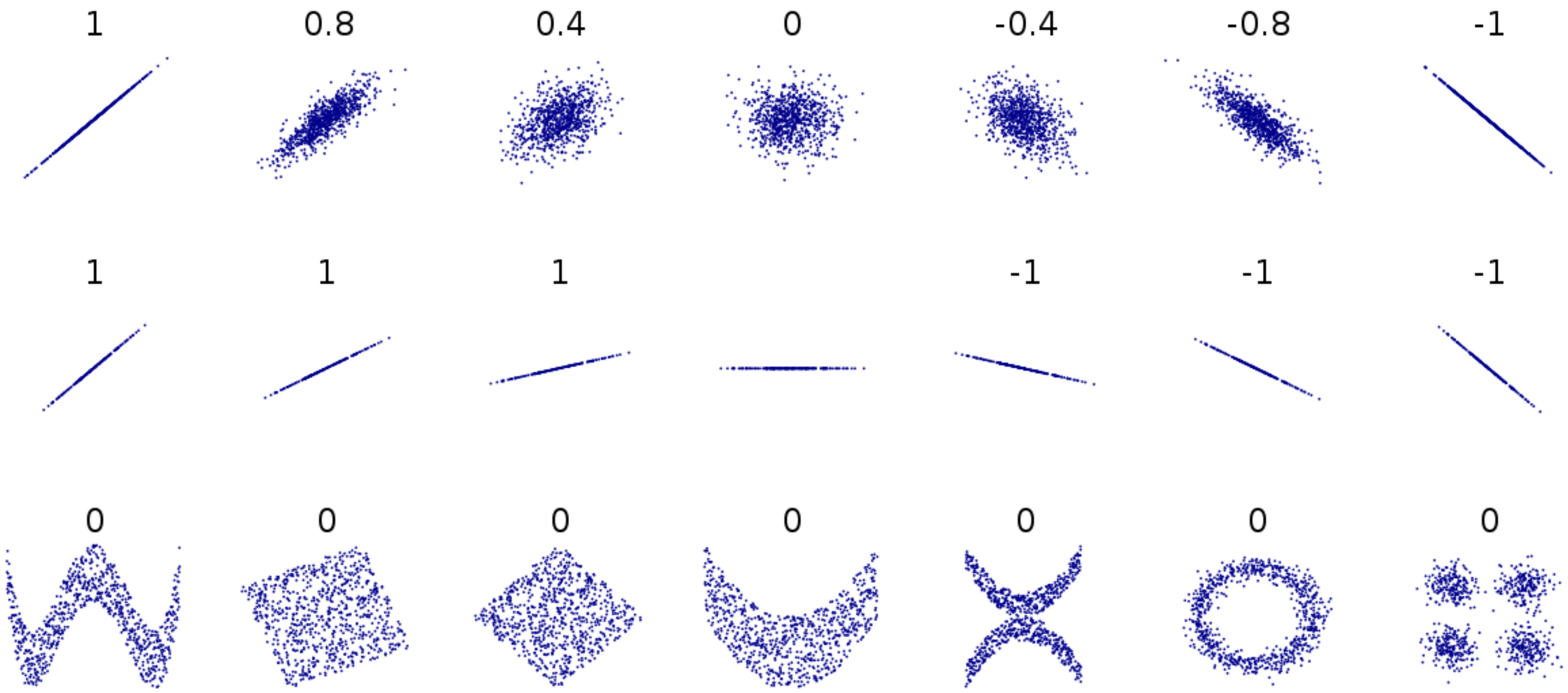
$$\rho(x,y) = 0$$

Kovariance a korelace nezávislých náhodných proměnných

- **nezávislé proměnné** x a $y \Rightarrow$ $\boxed{\mathrm{cov}(x, y) = 0 \qquad \rho(x, y) = 0}$

Obrácená implikace neplatí!

Nulová korelace je nutná, nikoli postačující podmínka nezávislosti proměnných.



Odhad kovariance a korelace

- máme náhodné proměnné x a y
- naměříme hodnoty $x_1, x_2, \dots, x_N$ a $y_1, y_2, \dots, y_N$

$$\boxed{\widehat{\mathrm{cov}}(x,y) = \langle xy\rangle - \langle x\rangle\langle y\rangle \qquad \hat{\rho}(x,y) = \frac{\langle xy\rangle - \langle x\rangle\langle y\rangle}{\hat{\sigma}_x\hat{\sigma}_y} \qquad \sigma_{\hat{\rho}} \approx \frac{1-\hat{\rho}^2}{\sqrt{N-1}}}$$

$$\langle xy\rangle = \frac{1}{N}\sum_{i=1}^{N} x_i y_i$$

$$\langle x\rangle = \frac{1}{N}\sum_{i=1}^{N} x_i$$

$$\langle y\rangle = \frac{1}{N}\sum_{i=1}^{N} y_i$$

$$\hat{\sigma}_x = \sqrt{\frac{1}{N-1}\sum_{i=1}^{N}(x_i - \langle x\rangle)^2}$$

$$\hat{\sigma}_y = \sqrt{\frac{1}{N-1}\sum_{i=1}^{N}(y_i - \langle y\rangle)^2}$$