Centrální limitní věta

- x_i náhodná proměnná s hustotou pravděpodobnosti $f_i(x)$
- x_i nezávislé

$$E[x_i] = \mu_i \qquad V[x_i] = \sigma_i^2$$

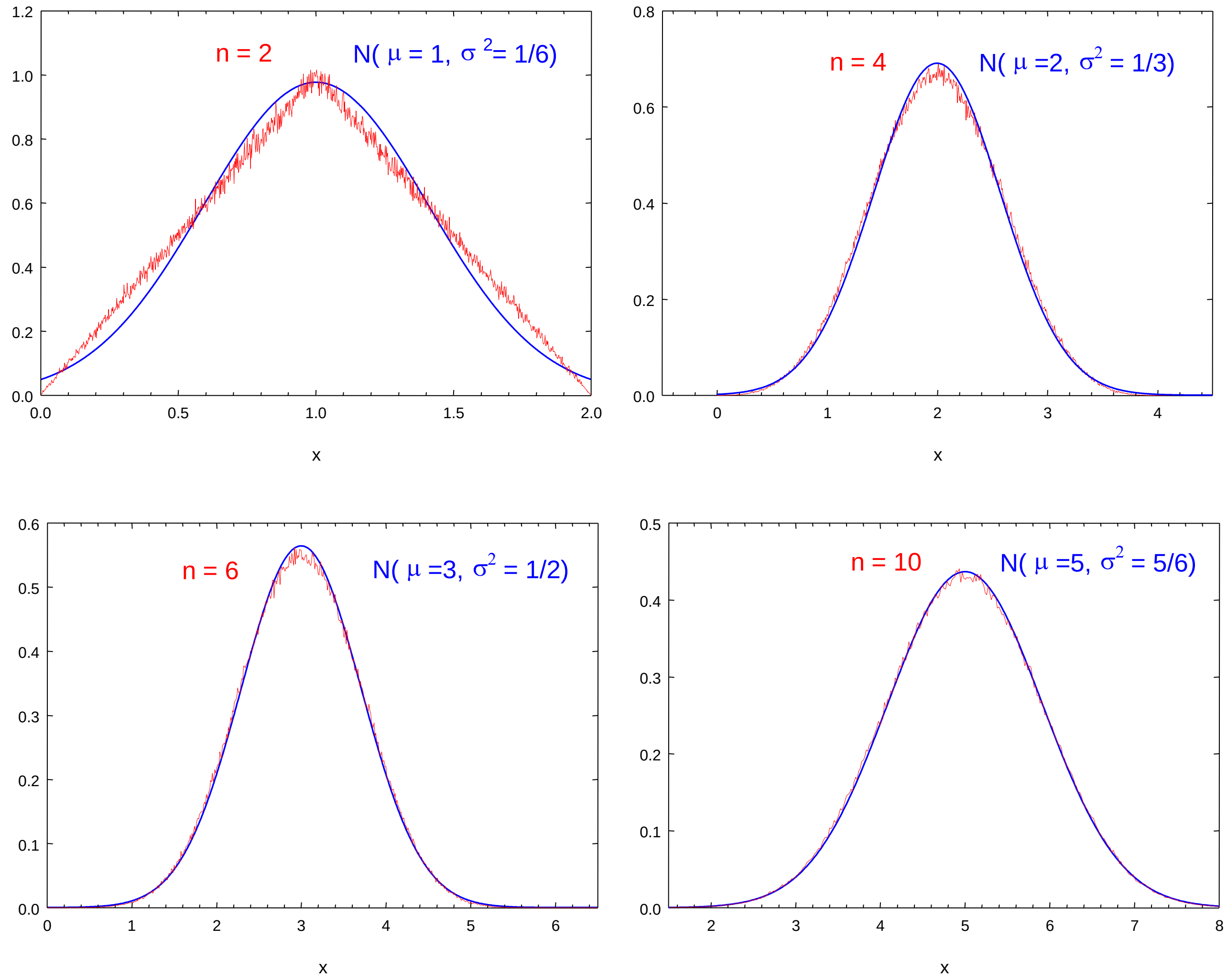
$$y = \sum_{i=1}^{N} x_i \qquad \text{pro } N \to \infty \text{ je } y \in N\left(\sum_{i=1}^{N} \mu_i, \sum_{i=1}^{N} \sigma_i^2\right)$$

$$\frac{y - \sum_{i=1}^{N} \mu_i}{\sqrt{\sum_{i=1}^{N} \sigma_i^2}} \to N(0,1)$$

Centrální limitní věta

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

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



Hustota pravděpodobnosti – případ dvou proměnných

marginální hustoty pravděpodobnosti:

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

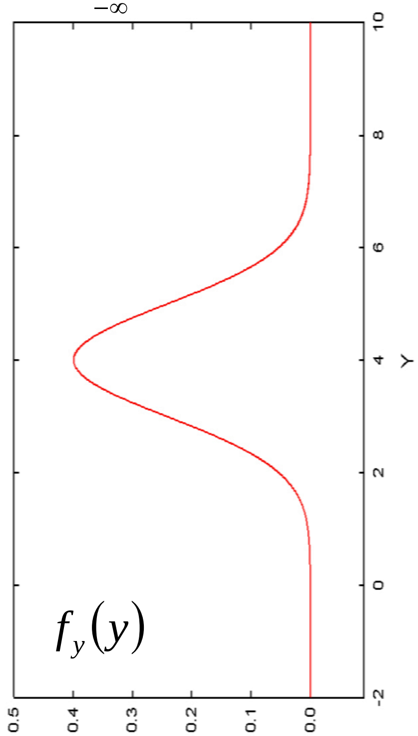
$$f_x(x) = \int_{-\infty}^{\infty} f(x,y)\,\mathrm{d}y$$

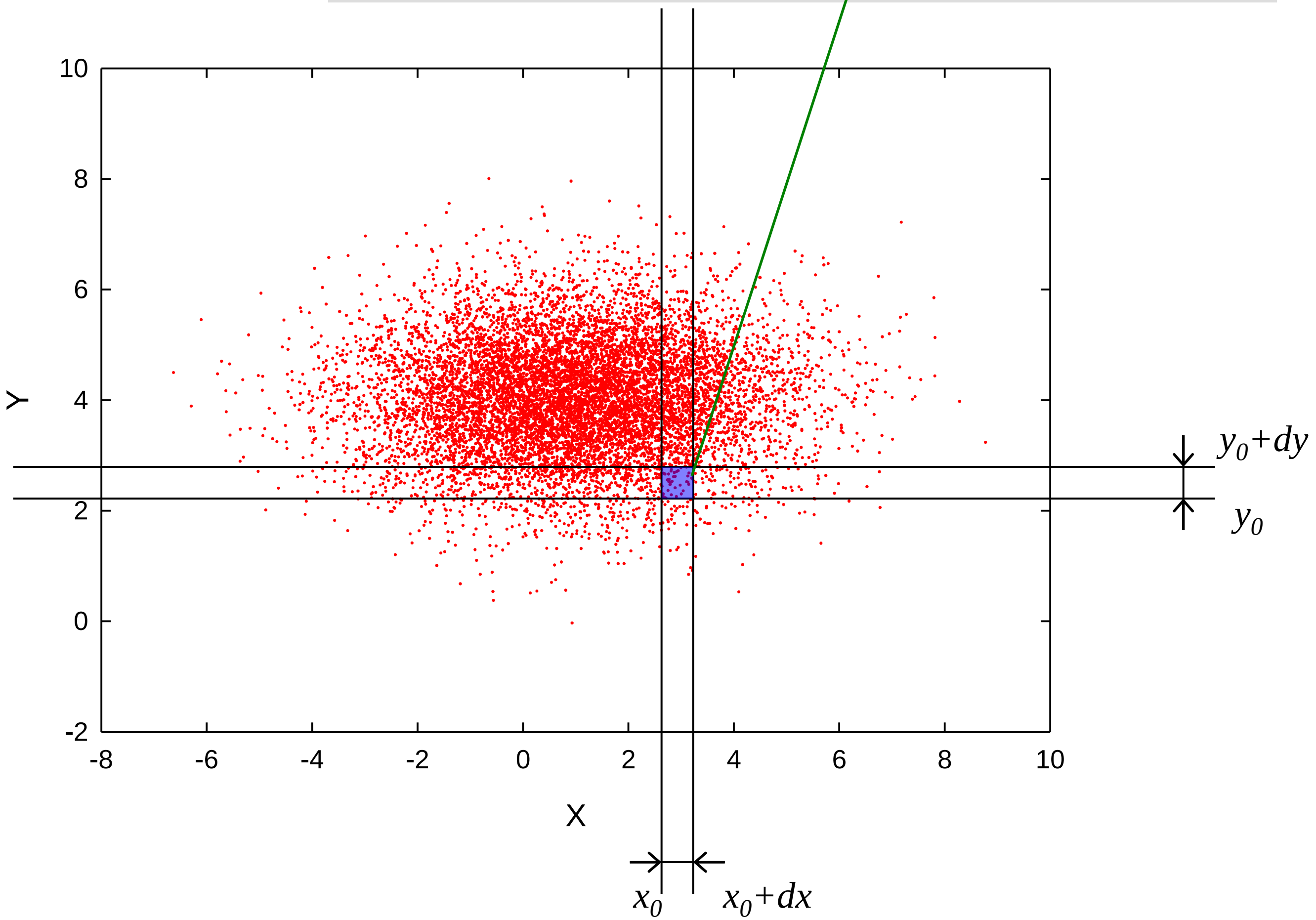
$A: \quad x \in \langle x_0, x_0 + dx \rangle$

$B: \quad y \in \langle y_0, y_0 + dy \rangle$

$$P(A \cap B) = f(x_0, y_0)\,dxdy$$

celková hustota pravděpodobnosti

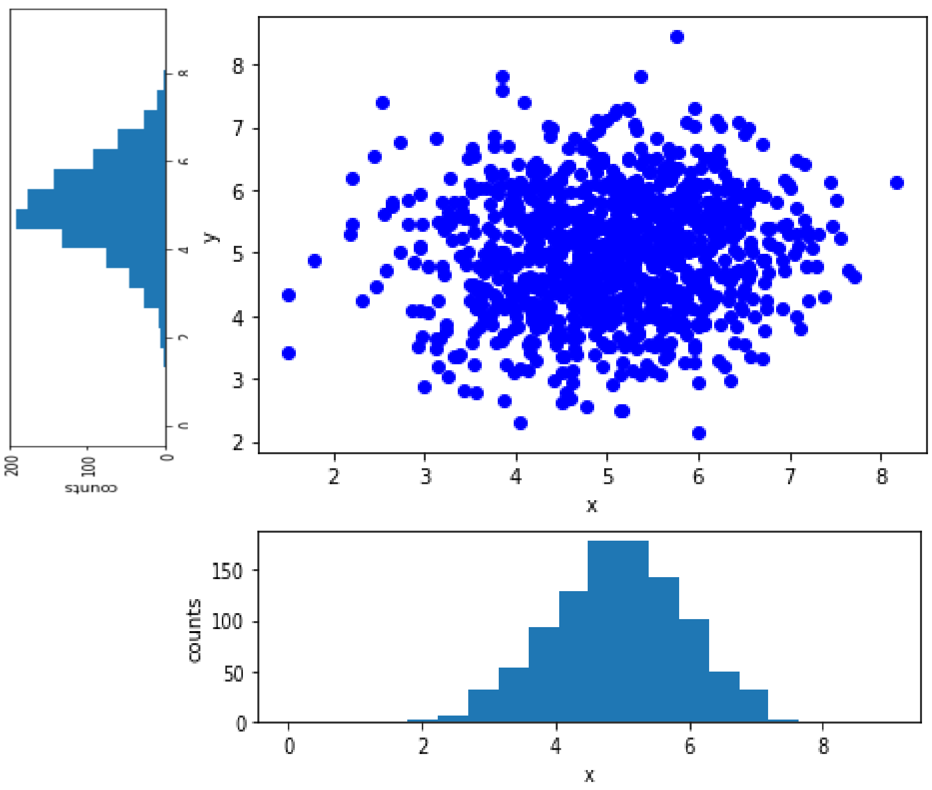


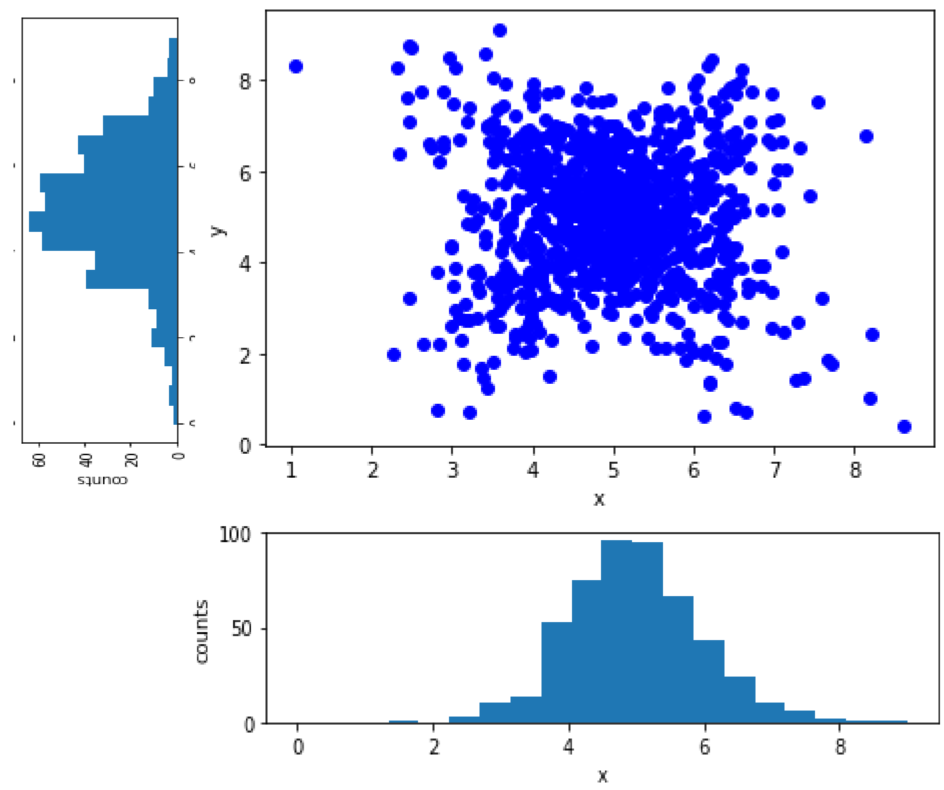


Hustota pravděpodobnosti – případ dvou proměnných

nezávislé náhodné proměnné

x, y jsou nezávislé $\Leftrightarrow$ $f(x,y) = f_x(x) f_y(y)$





Hustota pravděpodobnosti – případ dvou proměnných

operátor **očekávané hodnoty** : $E[\ ]$

očekávaná hodnota náhodné proměnné x : $\mu_x \equiv E[x] = \iint x\ f(x,y)\,\mathrm{d}x\,\mathrm{d}y$

očekávaná hodnota náhodné proměnné y : $\mu_y \equiv E[y] = \iint y\ f(x,y)\,\mathrm{d}x\,\mathrm{d}y$

obecně : $E[g(x,y)] = \iint g(x,y)\ f(x,y)\,\mathrm{d}x\,\mathrm{d}y$

Hustota pravděpodobnosti – případ dvou proměnných

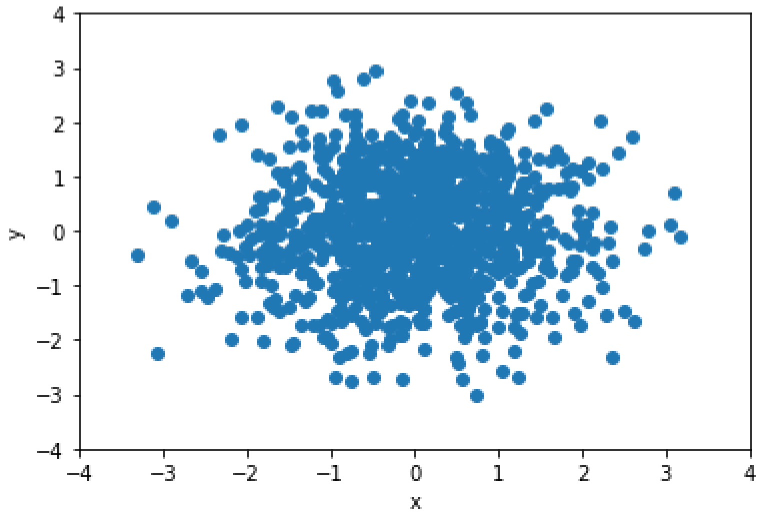
rozptyl náhodné proměnné x : $\sigma_x^2 \equiv V[x] = E\left[(x-\mu_x)^2\right] = \iint (x-\mu_x)^2 f(x,y)\,\mathrm{d}x\,\mathrm{d}y$

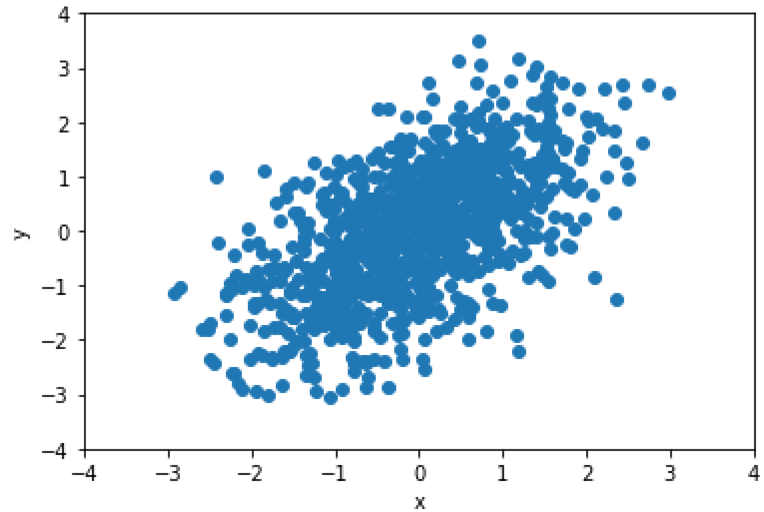
rozptyl náhodné proměnné y : $\sigma_y^2 \equiv V[y] = E\left[(y-\mu_y)^2\right] = \iint (y-\mu_y)^2 f(x,y)\,\mathrm{d}x\,\mathrm{d}y$

minimální smysluplná informace o náhodných proměnných x, y : μ_x, μ_y (míra polohy)

σ_x, σ_y (míra disperze)

4 čísla nestačí





Hustota pravděpodobnosti – případ dvou proměnných

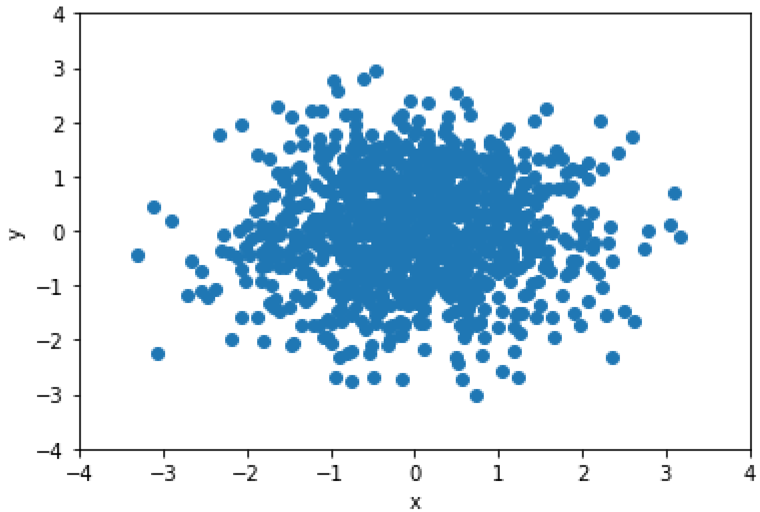
rozptyl náhodné proměnné *x* : $\sigma_x^2 \equiv V[x] = E\left[(x-\mu_x)^2\right] = \iint (x-\mu_x)^2 f(x,y)\,\mathrm{d}x\,\mathrm{d}y$

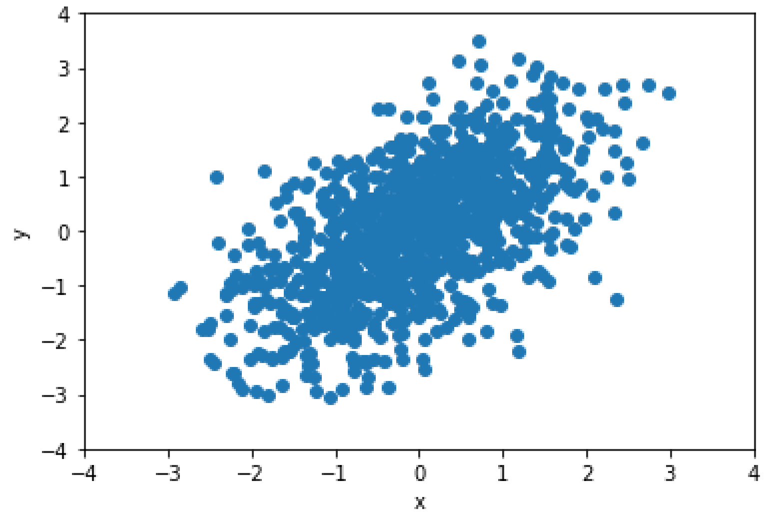
rozptyl náhodné proměnné *y* : $\sigma_y^2 \equiv V[y] = E\left[(y-\mu_y)^2\right] = \iint (y-\mu_y)^2 f(x,y)\,\mathrm{d}x\,\mathrm{d}y$

minimální smysluplná informace o náhodných proměnných *x*, *y* : μ_x, μ_y (míra polohy)

σ_x, σ_y (míra disperze)

kovariance cov (x,y)





Hustota pravděpodobnosti – případ dvou proměnných

kovariance náhodných proměnných *x* a *y* :

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

$\mathrm{cov}(x,x) = \mathrm{E}\,[(x-\mu_x)^2] = \sigma_x^2$

$\mathrm{cov}(y,y) = \mathrm{E}\,[(\mathrm{y}-\mu_y)^2] = \sigma_y^2$

kovarianční matice $V = \begin{pmatrix} \mathrm{cov}(x,x) & \mathrm{cov}(x,y) \\ \mathrm{cov}(y,x) & \mathrm{cov}(y,y) \end{pmatrix} = \begin{pmatrix} \sigma_x^2 & \mathrm{cov}(x,y) \\ \mathrm{cov}(x,y) & \sigma_y^2 \end{pmatrix}$

korelace náhodných proměnných *x* a *y* : $\rho(x,y) = \dfrac{\mathrm{cov}(x,y)}{\sigma_x\,\sigma_y}$

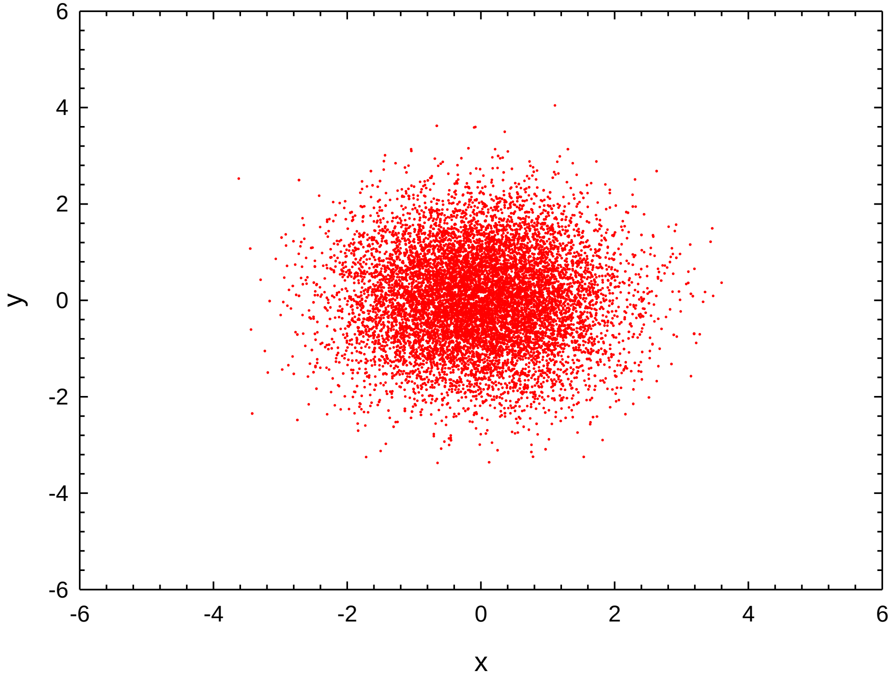
$\rho(x,x) = 1$

$\rho(y,y) = 1$

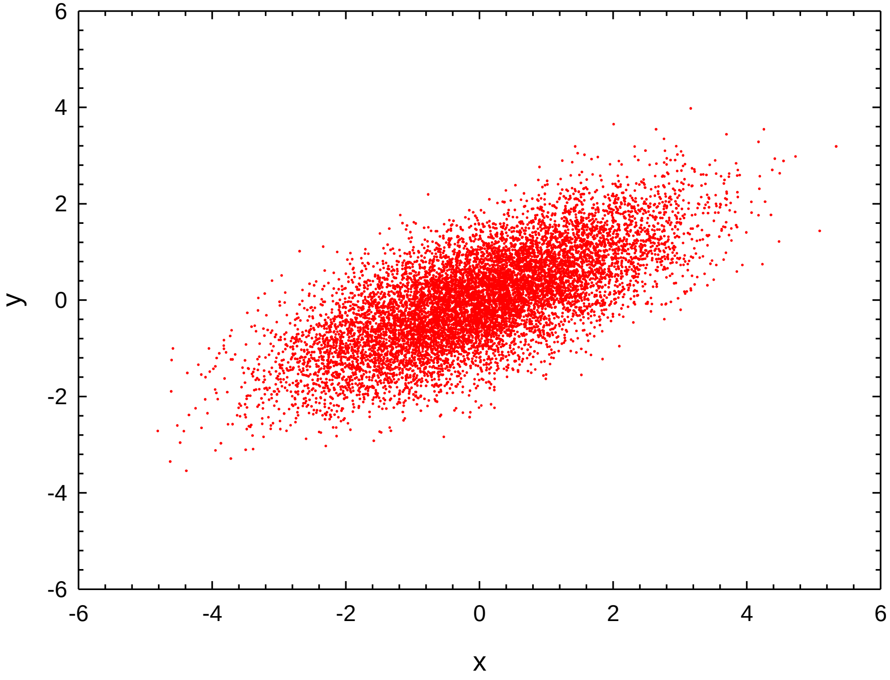
korelační matice $R = \begin{pmatrix} \varrho(x,x) & \varrho(x,y) \\ \varrho(y,x) & \varrho(y,y) \end{pmatrix} = \begin{pmatrix} 1 & \varrho(x,y) \\ \varrho(x,y) & 1 \end{pmatrix}$

Korelace náhodných proměnných

$\rho(x,y) = 0.0$



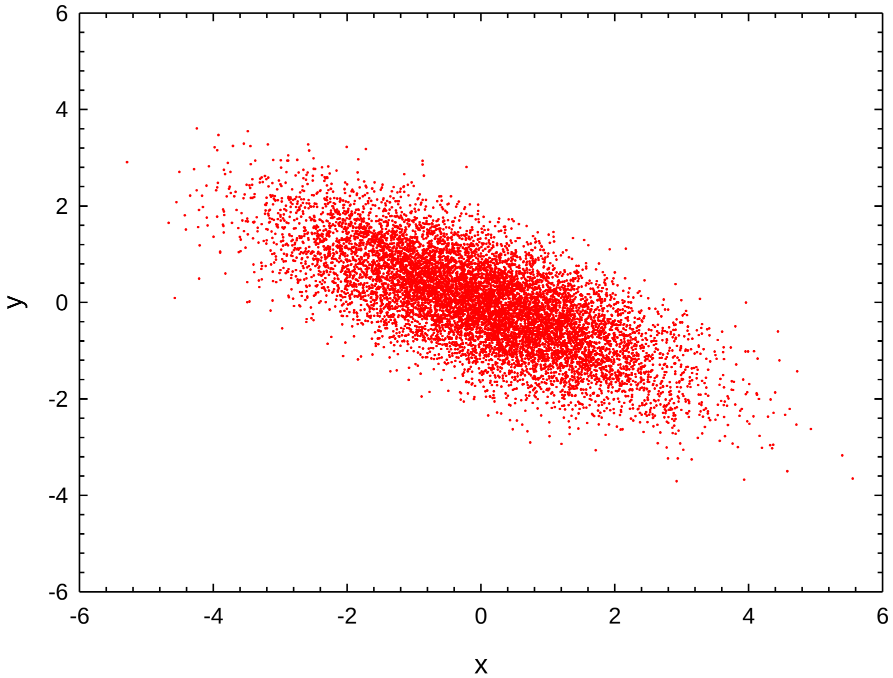
$\rho(x,y) = 0.7$



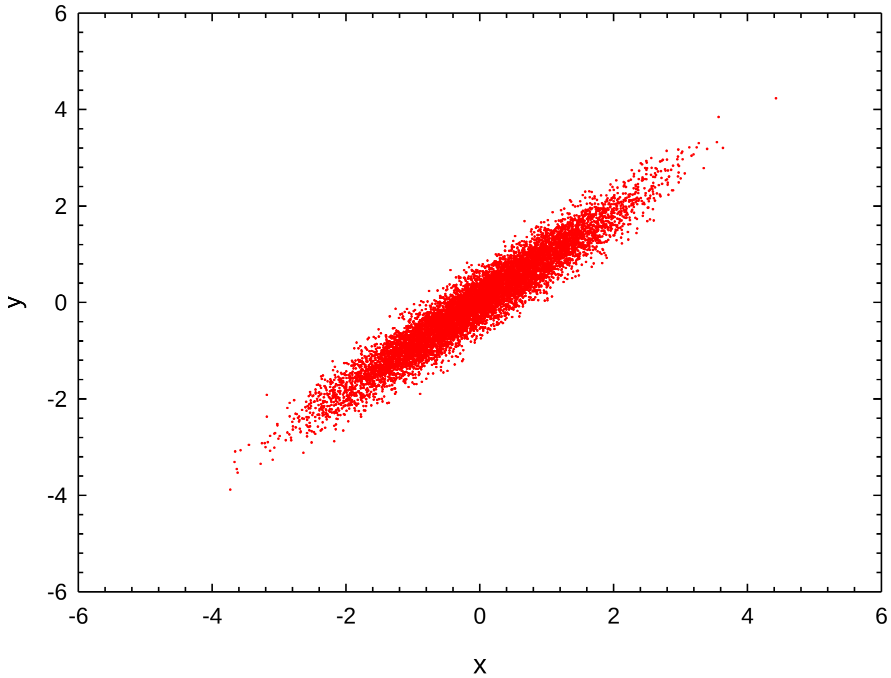
N = 10000

Korelace náhodných proměnných

$\rho(x,y) = -0.7$



$\rho(x,y) = 0.96$



N = 10000

Hustota pravděpodobnosti – případ dvou proměnných

nezávislé náhodné proměnné *x,y* : $f(x,y) = f_x(x) f_y(y)$

$$\mathrm{cov}(x,y) = E\left[(x-\mu_x)(y-\mu_y)\right] = \int_{-\infty}^{\infty}\int_{-\infty}^{\infty}(x-\mu_x)(y-\mu_y)f(x,y)\,dx\,dy =$$

$$= \int_{-\infty}^{\infty}(x-\mu_x)f_x(x)\,dx \int_{-\infty}^{\infty}(y-\mu_y)f_y(y)\,dy = E\left[(x-\mu_x)\right]E\left[(y-\mu_y)\right] = 0$$

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

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

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

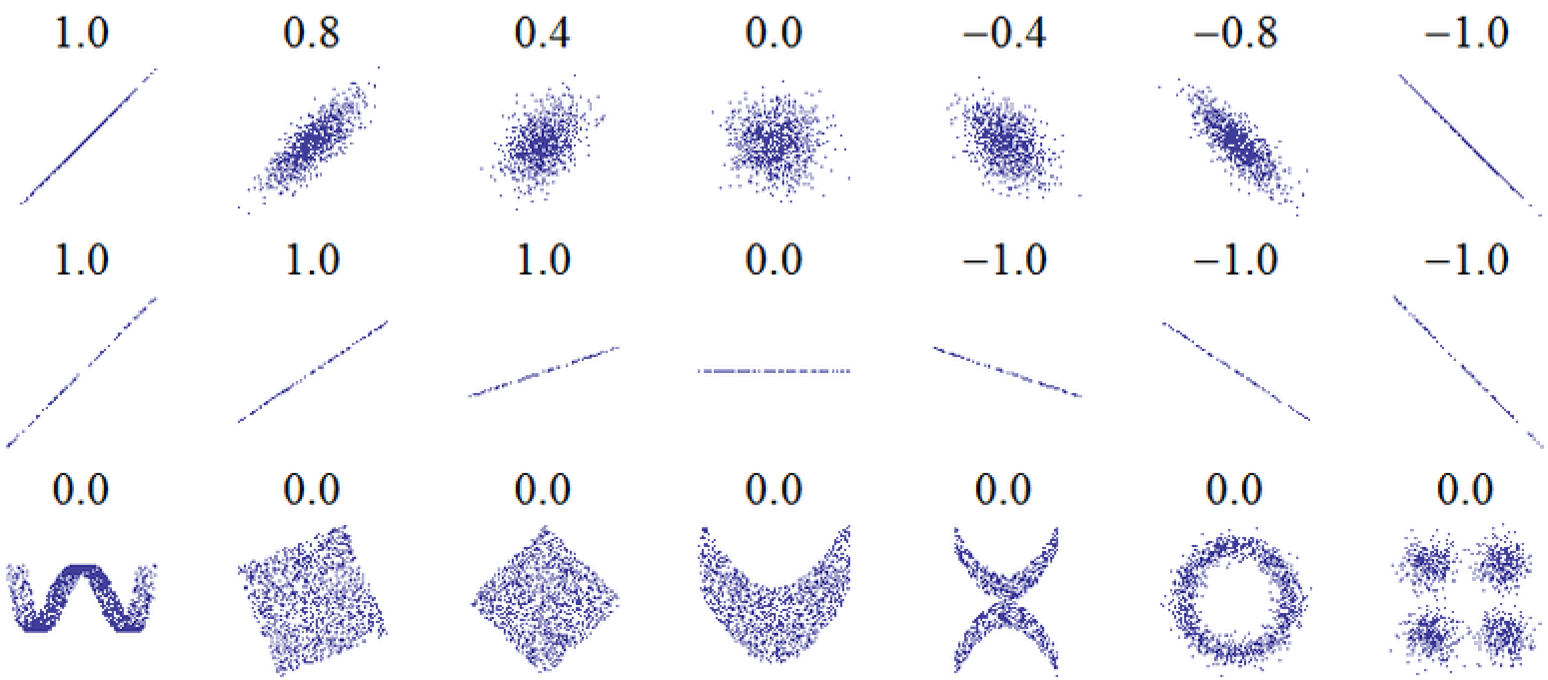
korelační matice: $R = \begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix}$

Hustota pravděpodobnosti – případ dvou proměnných

nezávislé náhodné proměnné $x,y \Rightarrow$ cov $(x,y) = 0$, $\rho\,(x,y) = 0$

obráceně to neplatí

Nulová korelace je nutná, ale ne postačující podmínka nezávislosti.



Odhad kovariance a korelace

náhodné proměnné *x*, *y*

- naměříme $x_1, x_2, \dots x_N$; $y_1, y_2, \dots y_N$

- cov (*x*, *y*) ???

$$\mathrm{cov}(x,y) = E[xy] - E[x]E[y]$$

$$\hat{\varrho}(x,y) = \frac{\langle xy\rangle - \langle x\rangle\langle y\rangle}{\sigma_x \sigma_y}$$

$$\boxed{\hat{\mathrm{cov}}(x,y) = \langle xy\rangle - \langle x\rangle\langle y\rangle}$$

$$\boxed{\hat{\varrho}(x,y) = \frac{\langle xy\rangle - \langle x\rangle\langle y\rangle}{\hat{\sigma}_x \hat{\sigma}_y} \quad \sigma_{\hat{\varrho}} \approx \frac{1-\hat{\varrho}^2}{\sqrt{N-1}}}$$

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

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

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

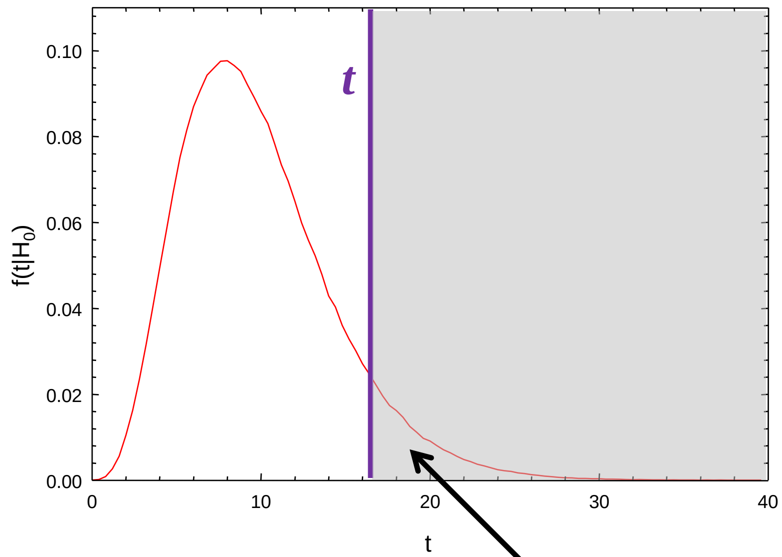
$$\hat{\sigma}_y = \sqrt{\frac{1}{N-1}\sum_{i=1}^{N}(y_i - \langle y\rangle)^2}$$

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

Test korelace – Fisherova transformace

náhodné proměnné x, y

- naměříme $x_1, x_2, \ldots x_N$; $y_1, y_2, \ldots y_N$ $\rightarrow$ dostaneme odhad korelace $\hat{\varrho}$
- je korelace statisticky významná?
- nulová hypotéza H_0 (zde nulová korelace)
- testovací statistika t
- známé rozdělení $f(t|H_0)$
- **Fisherova transformace** $F(\varrho) = \dfrac{1}{2} \ln \dfrac{1+\varrho}{1-\varrho}$



Pokud platí nulová hypotéza, tak má normální rozdělení s očekávanou hodnotou 0 a standardní odchylkou $\sigma = \dfrac{1}{\sqrt{N-3}}$

P - hodnota
pokud je $P < P_\alpha$
odmítneme H_0

- z - proměnná $z = \dfrac{F(\varrho)}{\sigma}$

P_α hladina signifikance, buď 5% nebo 1%

- pokud platí nulová hypotéza $z \in N(0, 1$

$$\boxed{P = 2\left(1 - F_{0,1}(z)\right) = 2\left(1 - \frac{1}{2}\left[1 + \mathrm{erf}\left(\frac{z}{\sqrt{2}}\right)\right]\right)}$$

Test korelace – studentovo rozdělení

náhodné proměnné *x*, *y*

- naměříme $x_1, x_2, \ldots x_N$; $y_1, y_2, \ldots y_N$ $\rightarrow$ dostaneme odhad korelace $\hat{\varrho}$
- je korelace statisticky významná?
- nulová hypotéza H_0 (zde nulová korelace)
- testovací statistika $t = \varrho \sqrt{\dfrac{N-2}{1-\varrho^2}}$
- pokud platí nulová hypotéza $t \in T(x|N-2)$

Studentovo rozdělení s počtem stupňů volnosti $\nu = N - 2$

Studentovo *t* rozdělení

William Sealy Gosset

studentovo rozdělení s ν stupni volnosti

$$f(x|\nu) = \frac{\Gamma\left(\frac{\nu+1}{2}\right)}{\sqrt{\nu\pi}\,\Gamma\left(\frac{\nu}{2}\right)}\left(1+\frac{x^2}{\nu}\right)^{-\frac{\nu+1}{2}}$$

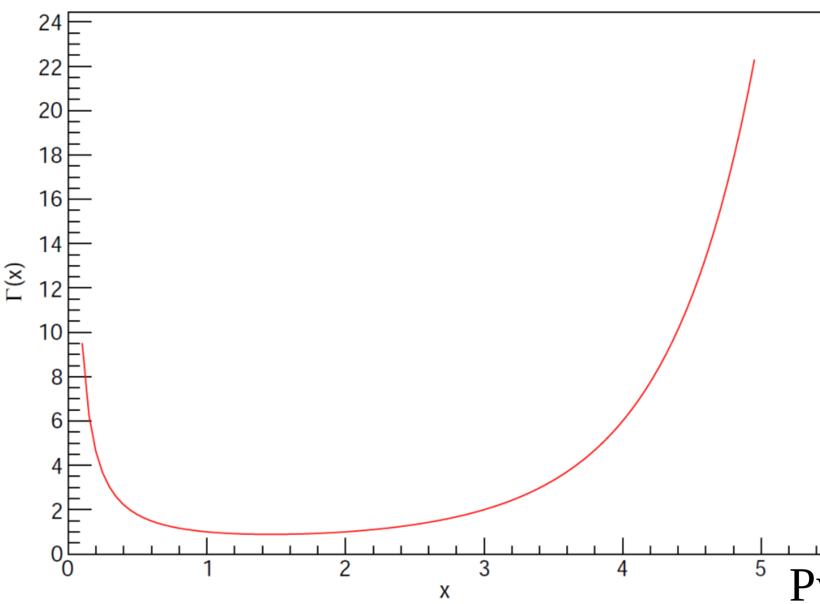
ν – počet stupňů volnosti

Gama funkce $\Gamma(x) = \int_0^\infty t^{x-1}\,e^{-t}\,dt \qquad x \geq 0$

$$\Gamma\left(\frac{1}{2}\right) = \sqrt{\pi}$$

$$\Gamma(n) = (n-1)! \qquad n \in \mathcal{N}$$

$$\Gamma(x+1) = x\Gamma(x) \qquad x \in \mathcal{R}$$



ROOT: `ROOT::Math::tgamma(x)`

Excel: `EXP(GAMMALN(x))`

Gnuplot: `gamma(x)`

Python: `from scipy.special import gamma`

Studentovo *t* rozdělení

Python:

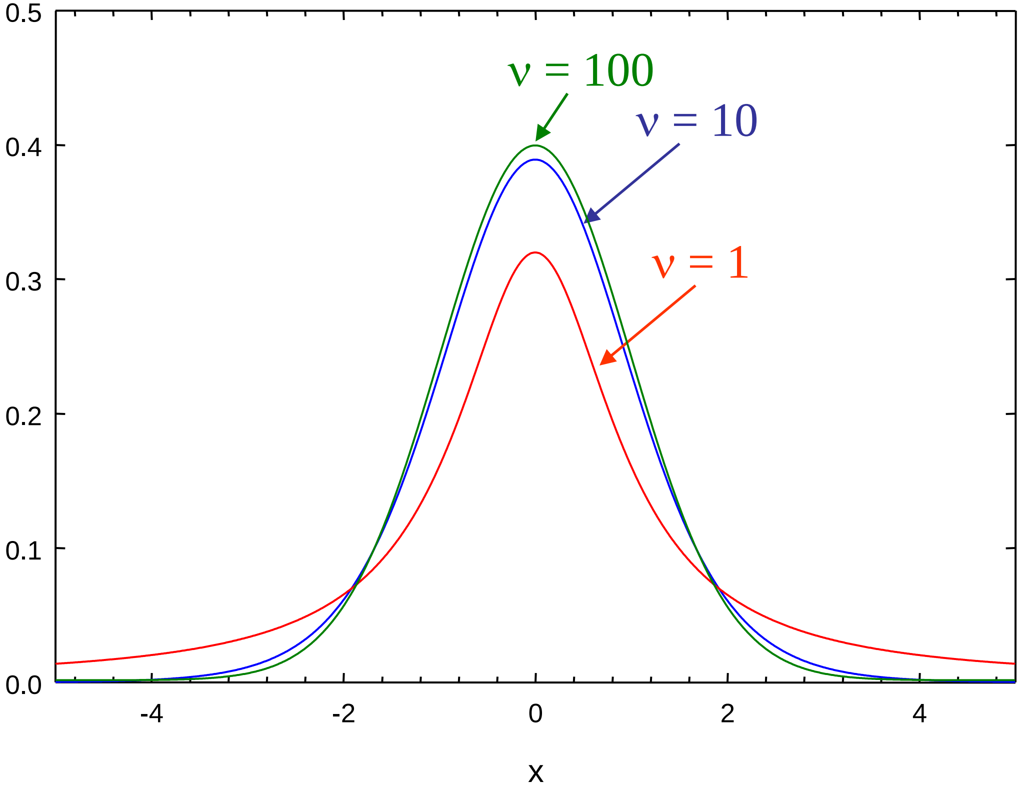
```
from scipy.stats import t
t.pdf(x,ν)
```

ROOT:

```
ROOT::Math::tdistribution_pdf(x,ν)
```

studentovo rozdělení s ν stupni volnosti

$$f(x|\nu) = \frac{\Gamma\left(\frac{\nu+1}{2}\right)}{\sqrt{\nu\pi}\,\Gamma\left(\frac{\nu}{2}\right)}\left(1+\frac{x^2}{\nu}\right)^{-\frac{\nu+1}{2}}$$



ν – počet stupňů volnosti

$$f(x|\nu=1) = \frac{1}{\pi}\frac{1}{1+x^2}$$

$$\nu \to \infty \quad f(x|\nu) \to N(0,1)$$

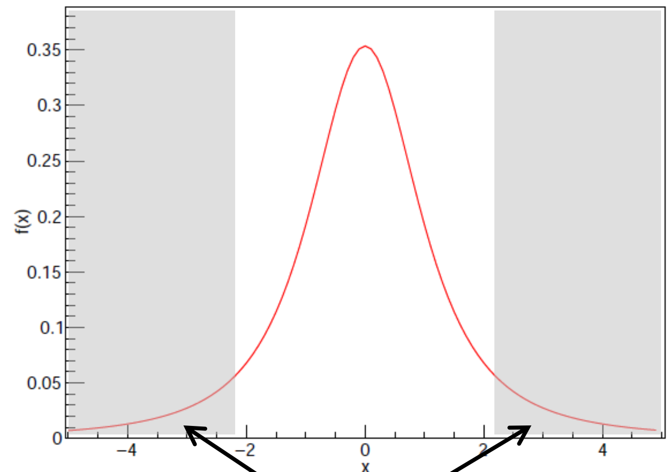
$$E[x] = \mu = 0$$

$$V[x] = \sigma^2 = \frac{\nu}{\nu-2} \quad \nu > 2$$

Test korelace – studentovo rozdělení

náhodné proměnné *x*, *y*

- naměříme $x_1, x_2, \ldots x_N$; $y_1, y_2, \ldots y_N$ $\rightarrow$ dostaneme odhad korelace $\hat{\varrho}$
- je korelace statisticky významná?
- nulová hypotéza H_0 (zde nulová korelace)
- testovací statistika $t = \varrho\sqrt{\dfrac{N-2}{1-\varrho^2}}$
- pokud platí nulová hypotéza $t \in T(N-2)$



Studentovo rozdělení s počtem stupňů volnosti $\nu = N - 2$

P_α

oboustranná *P*-hodnota

- *P*-hodnota: $P = 2(1 - T_\nu(t))$

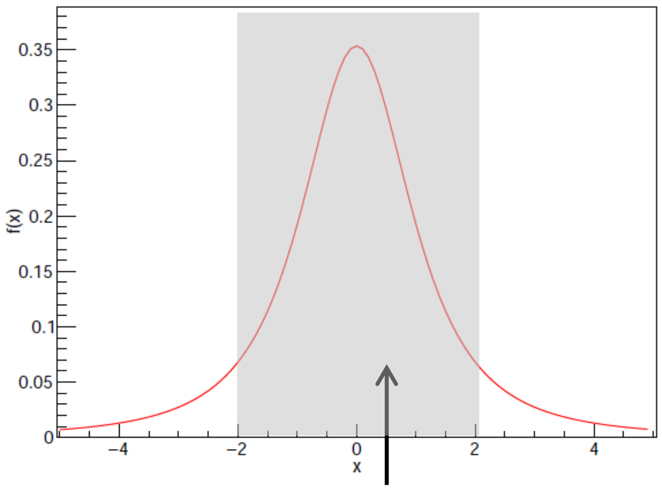
distribuční funkce studentova rozdělení

Test korelace – studentovo rozdělení

náhodné proměnné *x*, *y*

- naměříme $x_1, x_2, \dots x_N$; $y_1, y_2, \dots y_N$ $\rightarrow$ dostaneme odhad korelace $\hat{\varrho}$
- je korelace statisticky významná?
- nulová hypotéza H_0 (zde nulová korelace)
- testovací statistika $t = \varrho\sqrt{\dfrac{N-2}{1-\varrho^2}}$
- pokud platí nulová hypotéza $t \in T(N-2)$

  Studentovo rozdělení s počtem stupňů volnosti $\nu = N - 2$



$1\text{-}P_\alpha$
konfidenční interval

- P-hodnota: $P = 2(1 - T_\nu(t))$

  distribuční funkce studentova rozdělení

- konfidenční interval: $\left(-T_\nu^{-1}\left(1-\frac{P}{2}\right), T_\nu^{-1}\left(1-\frac{P}{2}\right)\right)$

  inverzní funkce k distribuční funkci studentova rozdělení