Centrální limitní věta

1. Ukažte, že součet N náhodných proměnných z rovnoměrným rozdělením U(0,1) konverguje k normálnímu rozdělení $N\left(\frac{N}{2},\sqrt{\frac{N}{12}}\right)$

CLT.py

```python
import numpy as np
import matplotlib.pyplot as plt

def gaussian(x,mu,sigma):
    return 1/(np.sqrt(2*np.pi)*sigma)*np.exp(-(x-mu)**2/(2*sigma**2))

N=10
Nsim=10000

x=np.empty(N)
y=np.empty(Nsim)

for i in range(Nsim):
    x=np.random.random_sample(N)
    y[i]=np.sum(x)


mu=N/2
sigma=np.sqrt(N/12)
xp=np.arange(0,N,0.01)
yp=gaussian(xp,mu,sigma)

plt.hist(y,bins=100,density='True')
plt.plot(xp,yp,c='red')
```

Centrální limitní věta

1. Ukažte, že součet N náhodných proměnných z rovnoměrným rozdělením U(0,1) konverguje k normálnímu rozdělení $N\left(\frac{N}{2}, \sqrt{\frac{N}{12}}\right)$

CLT.py

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

Předpověď CLT:

gaussián:

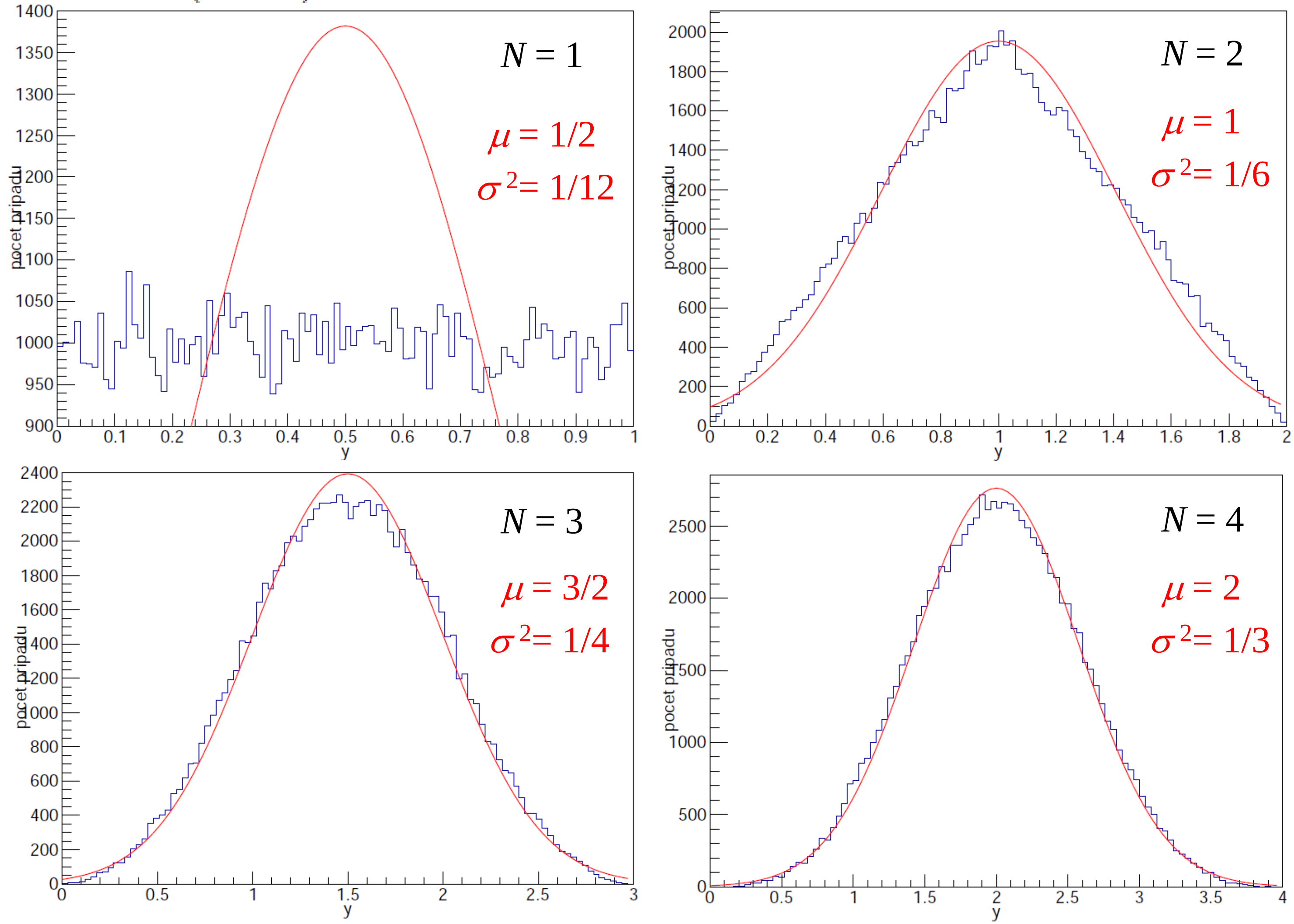
$$\frac{1}{\sqrt{2\pi}\sigma} \exp\left(-\frac{(y-\mu)^2}{2\sigma^2}\right)$$

očekávaná hodnota:

$$\mu = \frac{N}{2}$$

standardní odchylka:

$$\sigma = \sqrt{\frac{N}{12}}$$



Centrální limitní věta

1. Ukažte, že součet N náhodných proměnných z rovnoměrným rozdělením U(0,1) konverguje k normálnímu rozdělení $N\left(\frac{N}{2}, \sqrt{\frac{N}{12}}\right)$

CLT.py

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

Předpověď CLT:

gaussián:

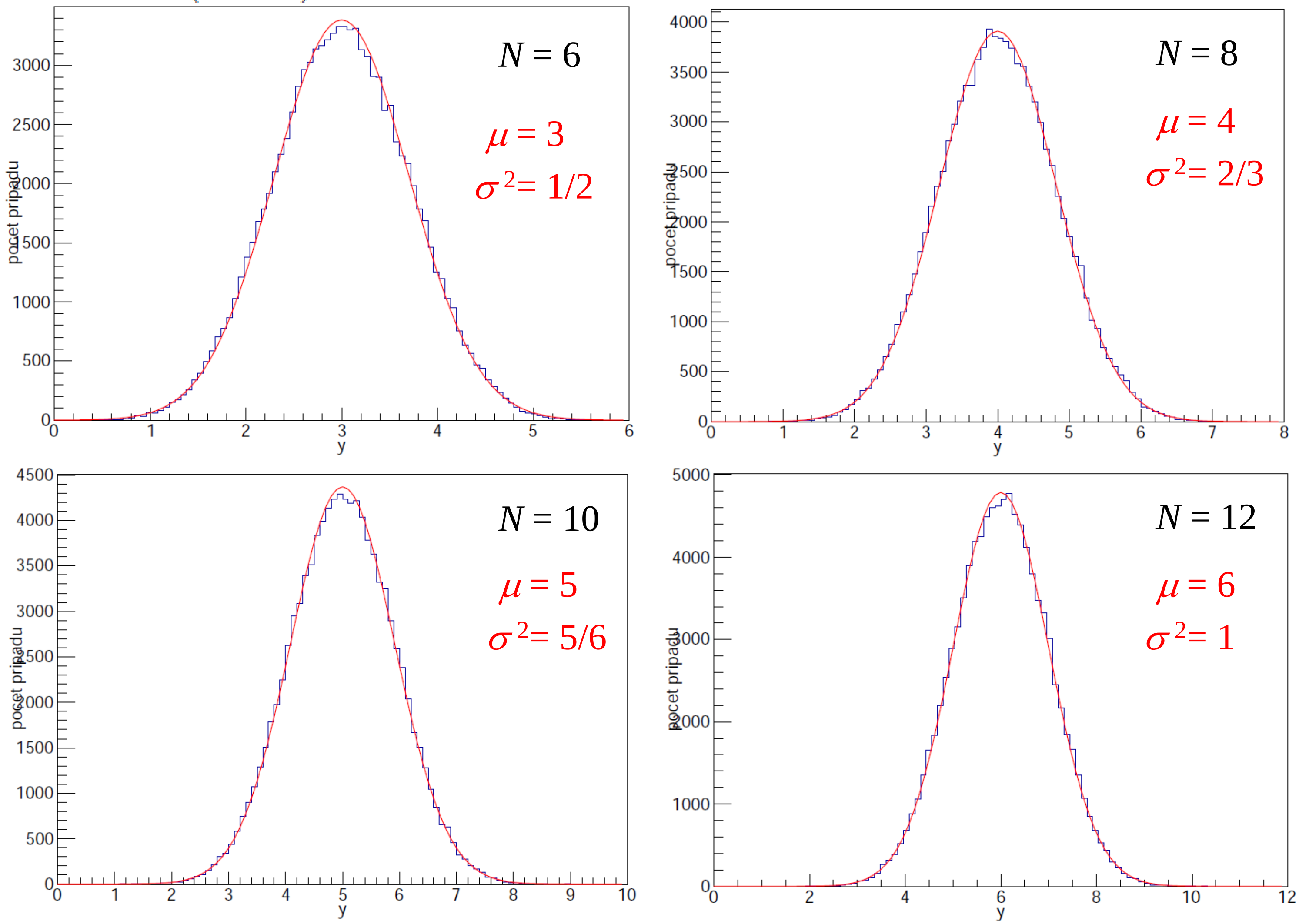
$$\frac{1}{\sqrt{2\pi}\sigma} \exp\left(-\frac{(y-\mu)^2}{2\sigma^2}\right)$$

očekávaná hodnota:

$$\mu = \frac{N}{2}$$

standardní odchylka:

$$\sigma = \sqrt{\frac{N}{12}}$$



Centrální limitní věta

Python program `CLT-kyvadlo.py`

hustota pravděpodobnosti: $f(x) = \frac{1}{\pi} \frac{1}{\sqrt{x_0^2 - x^2}}$ (kyvadlo)

distribuční funkce: $F(x) = \frac{1}{\pi} \int_{-x_0}^{x} \frac{\frac{1}{x_0} 1}{\sqrt{1 - \frac{t^2}{x_0^2}}} \, dt = \frac{1}{\pi} \left[\frac{\pi}{2} + \arcsin\left(\frac{x}{x_0}\right) \right]$

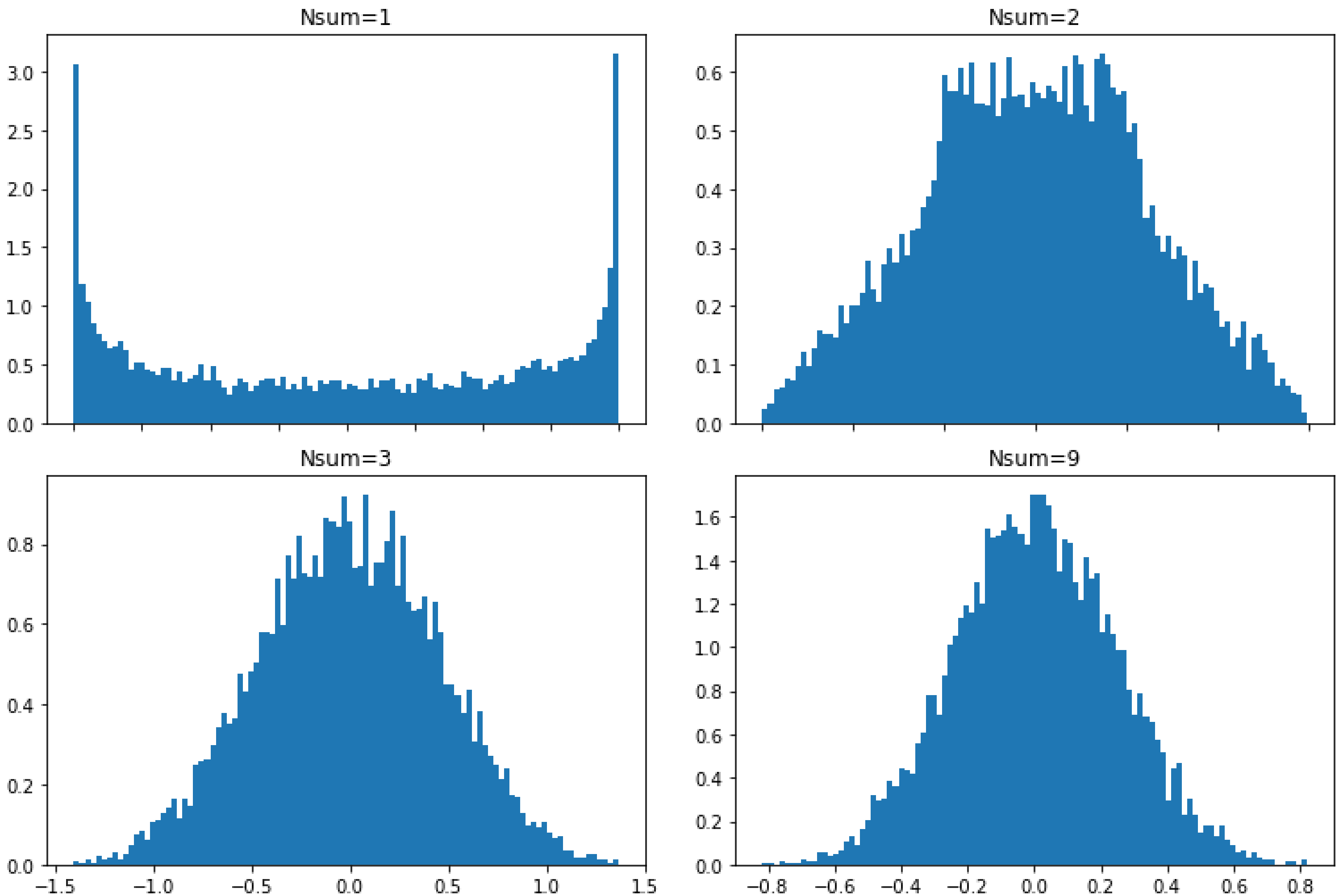
metoda inverzní funkce: $\frac{1}{\pi} \left[\frac{\pi}{2} + \arcsin\left(\frac{x}{x_0}\right) \right] = r$

$$x = x_0 \sin\left(\pi\left(r - \frac{1}{2}\right)\right) \qquad r \in U(0,1)$$

Centrální limitní věta

`CLT-kyvadlo.py`

hustota pravděpodobnosti: $$f(x) = \frac{1}{\pi} \frac{1}{\sqrt{x_0^2 - x^2}}$$ (kyvadlo)



Korelace

2. Proveďte analýzu korelace mezi výškou, váhou a měsícem narození studentů Matfyzu.

   Excel soubor `korelace-dotaznik.xlsx`

`COVAR(A1:A50,B1:B50)` – odhad kovariance

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

`PEARSON(A1:A50,B1:B50)` – odhad korelace
nebo `CORREL (A1:A50,B1:B50)`

$$\hat{\varrho}_E(x,y) = \frac{\langle xy \rangle - \langle x \rangle \langle y \rangle}{\hat{\sigma}_x \hat{\sigma}_y}$$

$$\hat{\varrho}(x,y) = \frac{N-1}{N} \hat{\varrho}_E(x,y)$$

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

nepředpojatý odhad standardní odchylky

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

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

předpojatý odhad standardní odchylky

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

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

Korelace

2. Proveďte analýzu korelace mezi výškou, váhou a měsícem narození studentů Matfyzu.

Excel soubor `korelace-dotaznik.xlsx`

Test statistické signifikance korelace pomocí Studentova rozdělení

$$t = \varrho\sqrt{\frac{N-2}{1-\varrho^2}}$$

$$t \in T(N-2)$$

`TDIST(x,`ν`,chvosty)` – 1 - hodnota distribuční funkce Studentova rozdělení s počtem stupňů volnosti ν, tj. $P(t > x \mid \nu)$

P - hodnota pro korelaci: `P = 2*(TDIST(x,`ν`,1)) = TDIST(x,`ν`,2)`

`TINV(`P_α`,`ν`)` – spočítá takovou hodnotu t_P, pro kterou platí, že pravděpodobnost, že náhodná proměnná ze Studentova rozdělení s počtem stupňů volnosti ν se bude od nuly lišit o víc než t_P je P, tj. $P(|t| > t_P \mid \nu)$

konfidenční interval pro hladinu signifikance P_α : `(-TINV(`P_α`,`ν`), TINV(`P_α`,`ν`))`

Korelace

2. Proveďte analýzu korelace mezi výškou, váhou a měsícem narození studentů Matfyzu.

   Python soubor `korelace-dotaznik.py`

Test statistické signifikance Pearsonova korelačního koeficientu pomocí Beta rozdělení

Za předpokladu, že korelace náhodných proměnných *x*,*y* je nula, má Pearsonův korelační koeficient rozdělení s hustotou pravděpodobnosti

$$f(\varrho|N) = \frac{(1-\varrho^2)^{\frac{N-4}{2}}}{B\left(\frac{1}{2}, \frac{N-2}{2}\right)}$$

kde *B*(*x*,*y*) je Beta funkce $B(x,y) = \int_0^1 t^{x-1}(1-t)^{y-1}\,dt$ $\qquad B(x,y) = \frac{\Gamma(x)\Gamma(y)}{\Gamma(x+y)}$

P-hodnota: $P = 2F(-|\varrho|, N) = 2\int_{-1}^{-|\varrho|} f(r|N)\,dr$

```
import scipy.stats as stat
Pearson,P=stat.pearsonr(vyska,vaha)
print("korelace =", Pearson, "P-hodnota = ", P)
```