Centrální limitní věta

CLT.py

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

```
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=12
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

`CLT.py`

náhodná proměnná y

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

předpověď CLT:

- gaussián

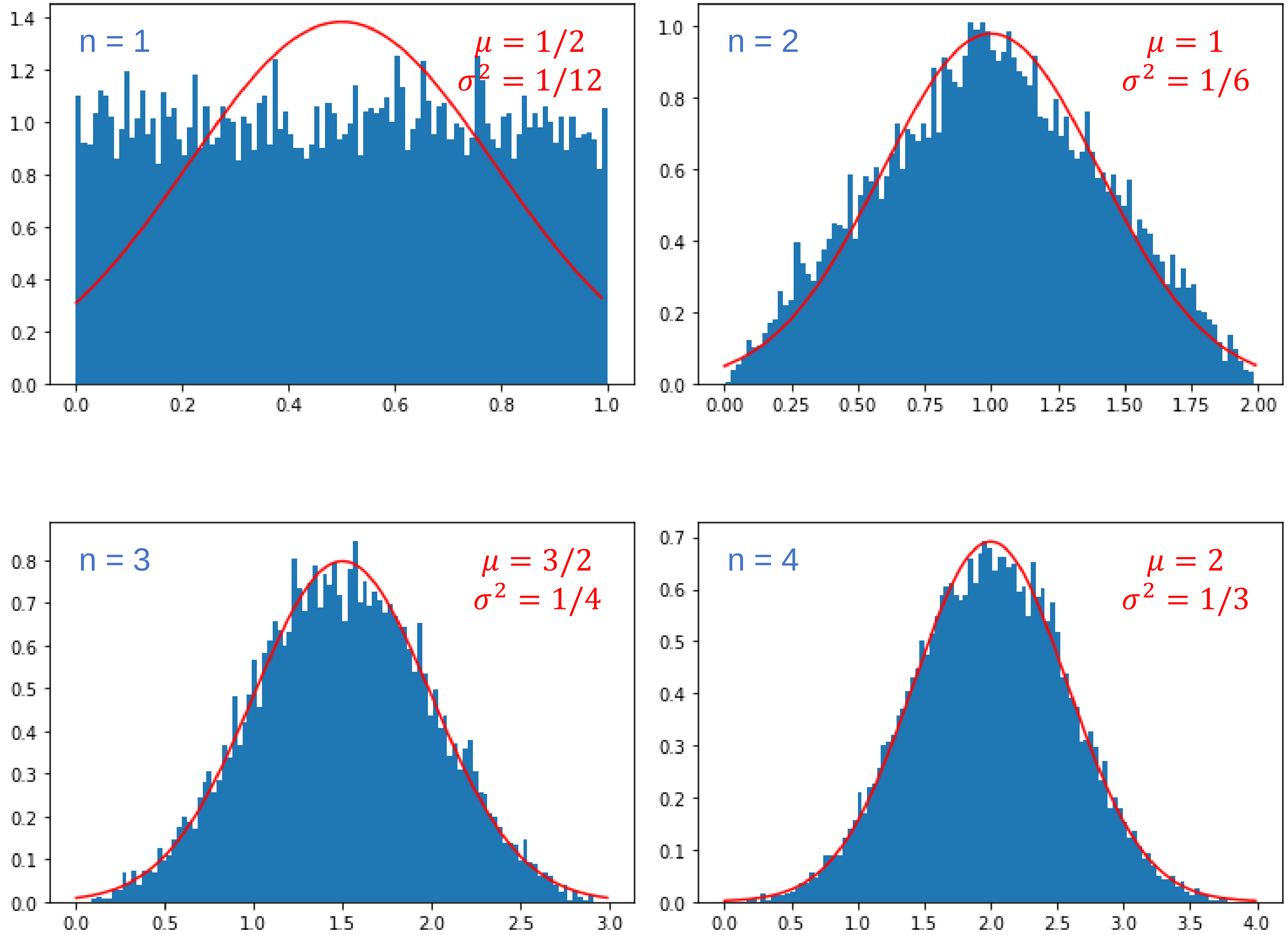
$$\frac{1}{\sqrt{2\pi\sigma^2}} \exp\left[-\frac{(y-\mu)^2}{2\sigma^2}\right]$$

- očekávaná hodnota

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

- standardní odchylka

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



Centrální limitní věta

CLT.py

náhodná proměnná y

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

předpověď CLT:

- gaussián

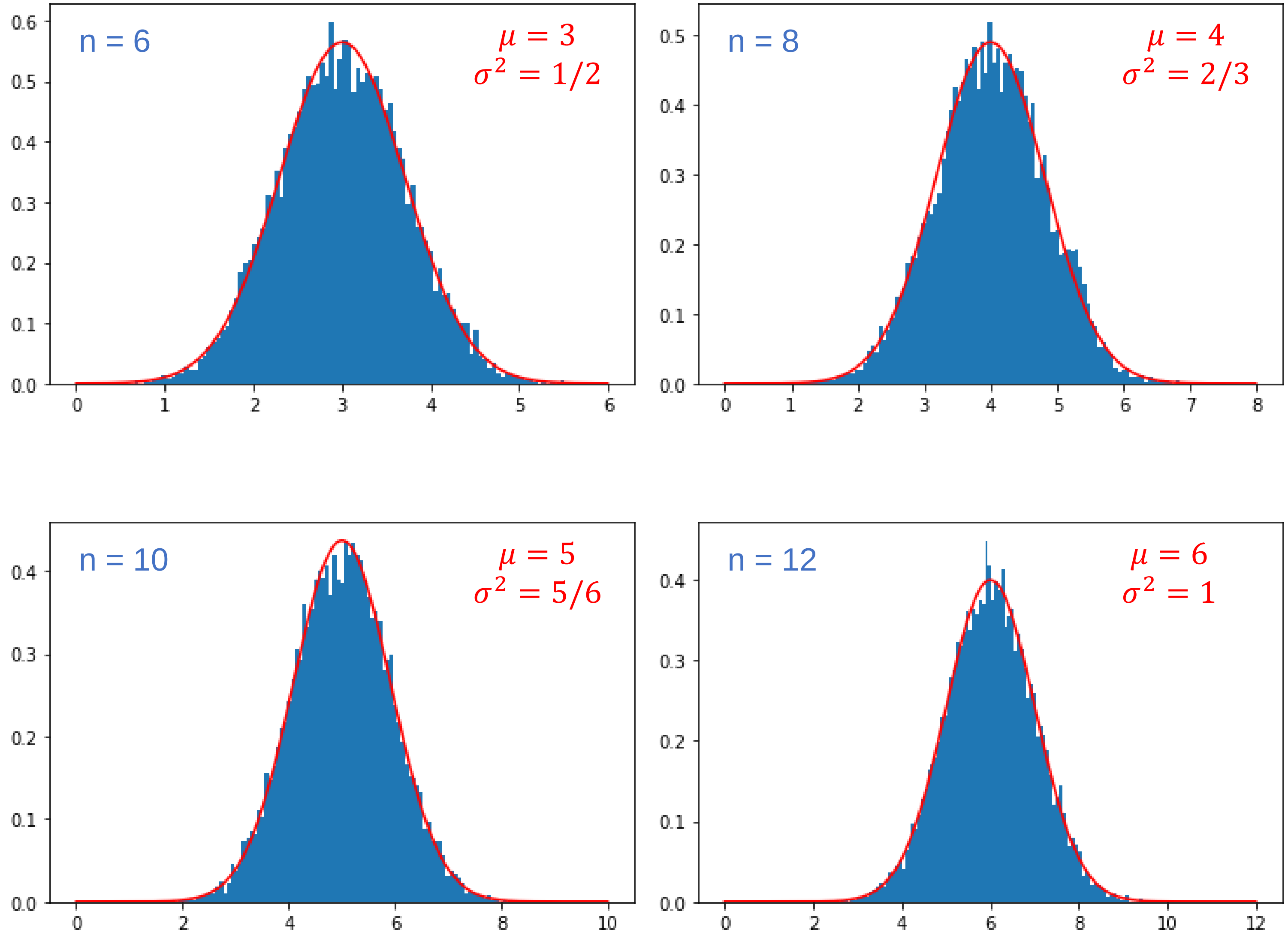
$$\frac{1}{\sqrt{2\pi\sigma^2}} \exp\left[-\frac{(y-\mu)^2}{2\sigma^2}\right]$$

- očekávaná hodnota

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

- standardní odchylka

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



Centrální limitní věta

`CLT-kyvadlo.py`

2. Otestujte centrální limitní větu na hustotě pravděpodobnosti polohy matematického kyvadla.

hustota pravděpodobnosti
$$f(x) = \frac{1}{\pi}\frac{1}{\sqrt{x_m^2 - x^2}} \qquad \mu_x = 0 \qquad \sigma_x = \frac{x_m}{\sqrt{2}}$$

distribuční funkce
$$F(x) = \frac{1}{\pi}\int_{-x_m}^{x} \frac{\frac{1}{x_m}}{\sqrt{1-\left(\frac{y}{x_m}\right)^2}}\,\mathrm{d}y = \frac{1}{\pi}\left[\frac{\pi}{2} + \arcsin\left(\frac{x}{x_m}\right)\right]$$

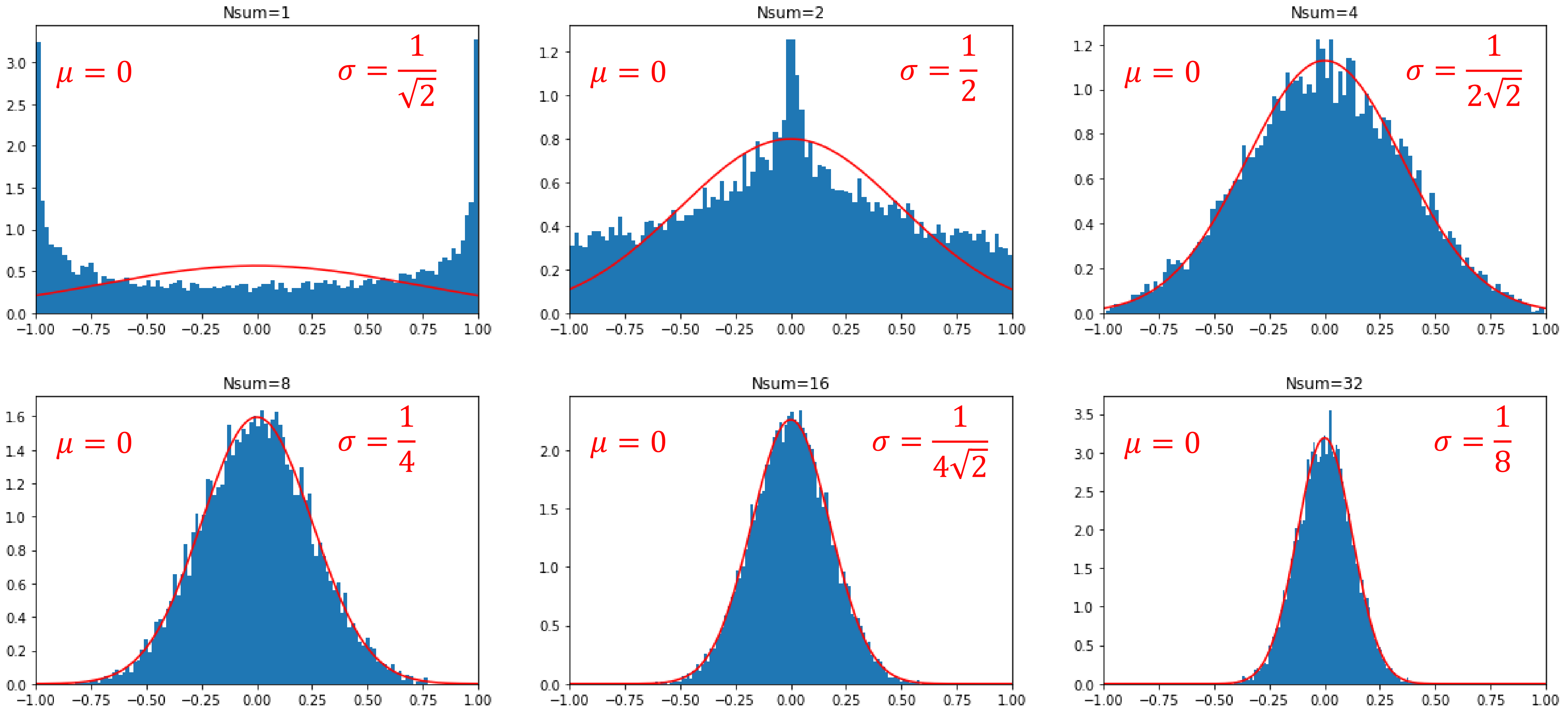
metoda inverzní funkce
$$r = \frac{1}{\pi}\left[\frac{\pi}{2} + \arcsin\left(\frac{x}{x_m}\right)\right] \qquad r \in U(0,1)$$

$$x = x_m \sin\left[\pi\left(r - \frac{1}{2}\right)\right]$$

Centrální limitní věta

`CLT-kyvadlo.py`

2. Otestujte centrální limitní větu na hustotě pravděpodobnosti polohy matematického kyvadla.



Centrální limitní věta

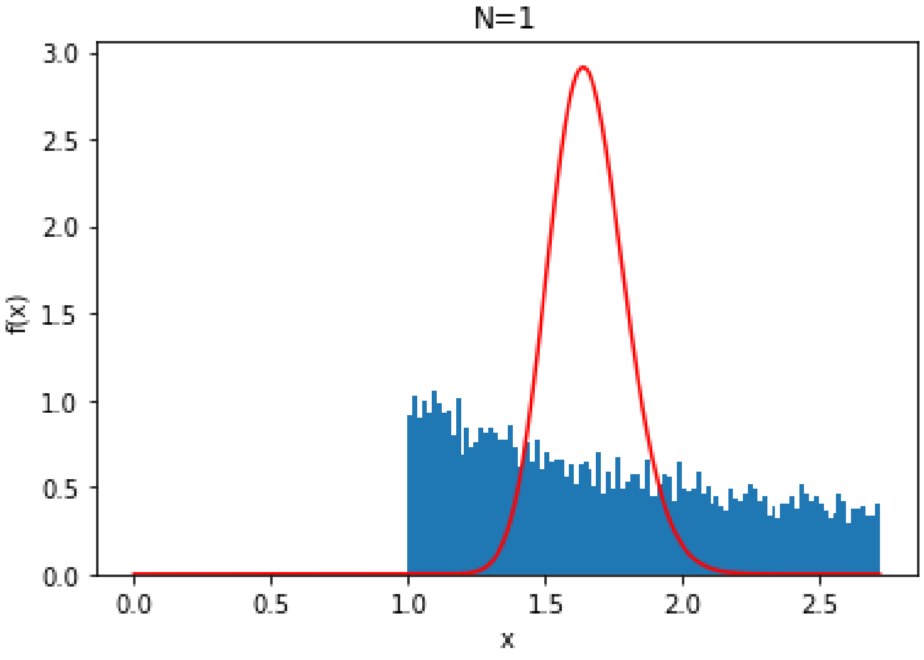
`MLT.py`

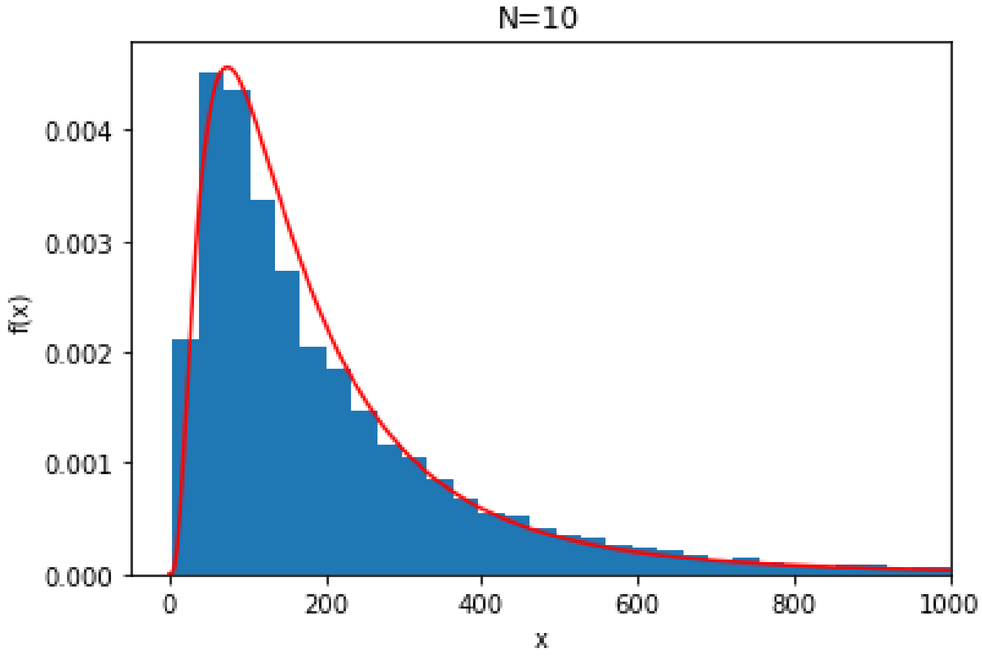
3. Jaké rozdělení bude mít součin N náhodných proměnných x_i pro $n \to \infty$?

$$y = x_1 \cdot x_2 \cdot \ldots \cdot x_n$$

$$z = \ln y = \ln x_1 + \ln x_2 + \cdots + \ln x_n$$ pro $n \to \infty$ je z výběrem z normálního rozdělení

$$z = \ln y \qquad \to y = e^z$$ pro $n \to \infty$ je y výběrem z log-normálního rozdělení





Centrální limitní věta

`MLT.py`

3. Jaké rozdělení bude mít součin N náhodných proměnných x_i pro $N \to \infty$?

log-normální rozdělení

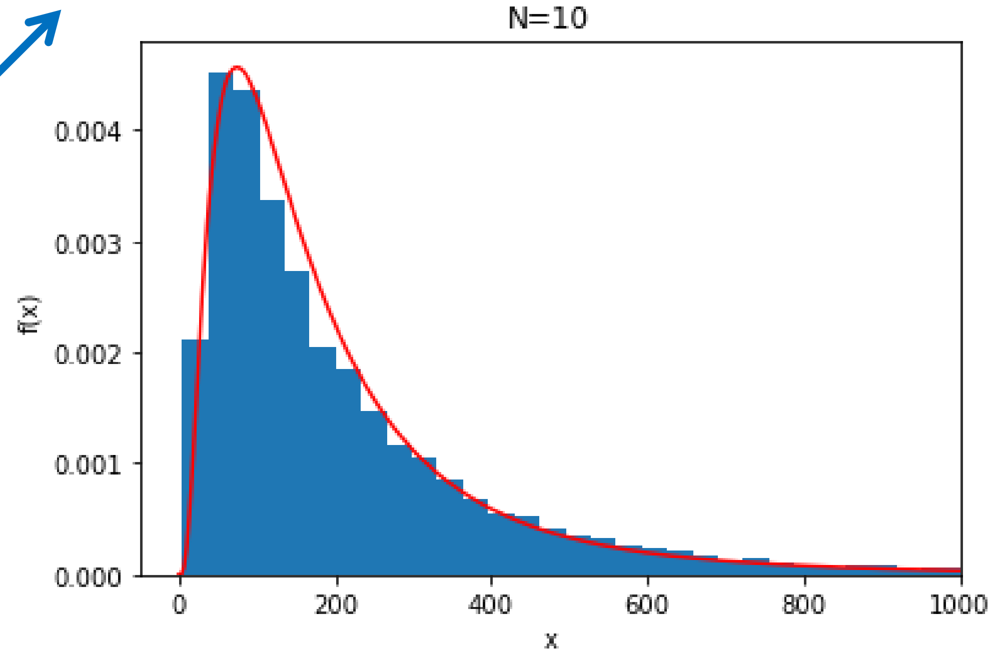
z má hustotu pravděpodobnosti
$$f(z) = \frac{1}{\sqrt{2\pi\sigma^2}} \exp\left[-\frac{(z-\mu)^2}{2\sigma^2}\right]$$

y má hustotu pravděpodobnosti
$$g(y) = \frac{1}{y}\frac{1}{\sqrt{2\pi\sigma^2}} \exp\left[-\frac{(\ln y-\mu)^2}{2\sigma^2}\right]$$

$$z = \ln y \to y = e^z$$

$$g(y)\mathrm{d}y = f(z)\mathrm{d}z$$

$$g(y) = \left|\frac{\mathrm{d}z}{\mathrm{d}y}\right| f\big(z(y)\big)$$



Korelace

`korelace-PH.xlsx`

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

kovariance	Excel:	`COVAR(A1:A50,B1:B50)`
	odhad kovariance:	$\widehat{\mathrm{cov}}(x,y) = \langle xy \rangle - \langle x \rangle \langle y \rangle$
korelace	Excel:	`PEARSON(A1:A50,B1:B50)` `CORREL(A1:A50,B1:B50)`
	odhad korelace:	$\hat{\rho}(x,y) = \dfrac{\langle xy \rangle - \langle x \rangle \langle y \rangle}{\hat{\sigma}_x \hat{\sigma}_y}$
	oprava:	$\hat{\rho}(x,y) = \dfrac{n-1}{n} \hat{\rho}_{\mathrm{Excel}}(x,y)$

Korelace

`korelace-PH.xlsx`

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

předpojatý odhad (Excel)

standardní odchylka:

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

korelace:

$$\hat{\rho}_{\mathrm{Excel}}(x, y) = \frac{\langle xy \rangle - \langle x \rangle \langle y \rangle}{\hat{s}_x \hat{s}_y}$$

nepředpojatý odhad

standardní odchylka:

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

korelace:

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