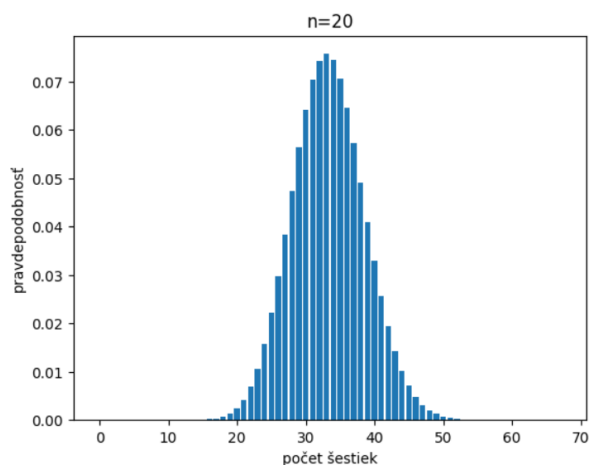
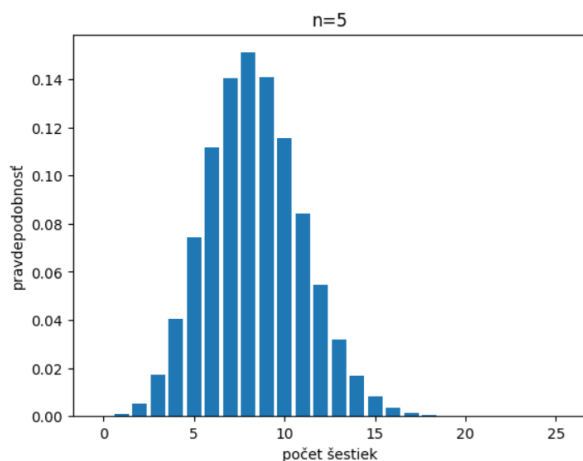
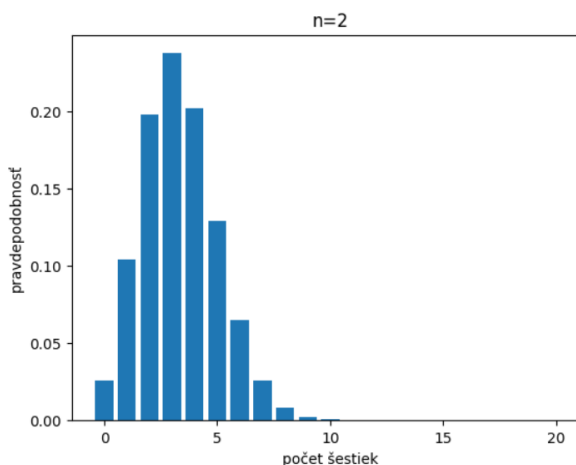
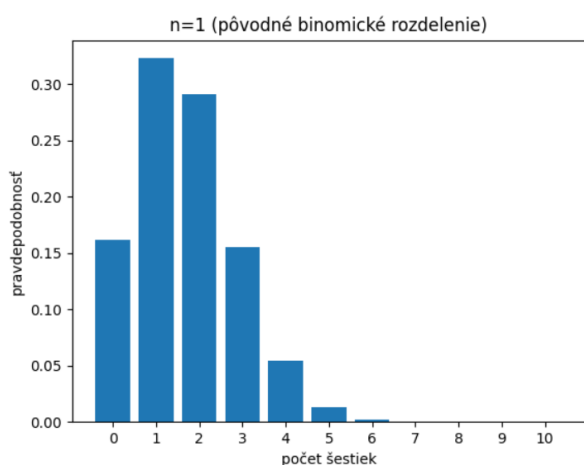


(05) Súčet nezávislých rovnako rozdelených náhodných premenných, centrálna limitná veta

1. Príklad: Binomické rozdelenie

Majme postupnosť N nezávislých pokusov s pravdepodobnosťou úspechu p , počet úspechov má binomické rozdelenie. Ak toto zopakujeme nezávisle n krát, dostaneme náhodné premenné $X_1, X_2, \dots, X_n$ označujúce počty úspechov. Na ich súčet sa však môžeme pozeráť aj ako na počet úspechov v postupnosti $n \cdot N$ nezávislých pokusoch, takže má znovu binomické rozdelenie, a to $\text{Bin}(n \cdot N, p)$.

Ilustrácia pre $N=10$, $p=\frac{1}{6}$ (hádzame desaťkrát kockou, chceme hodiť šestku), opakujeme to n krát.



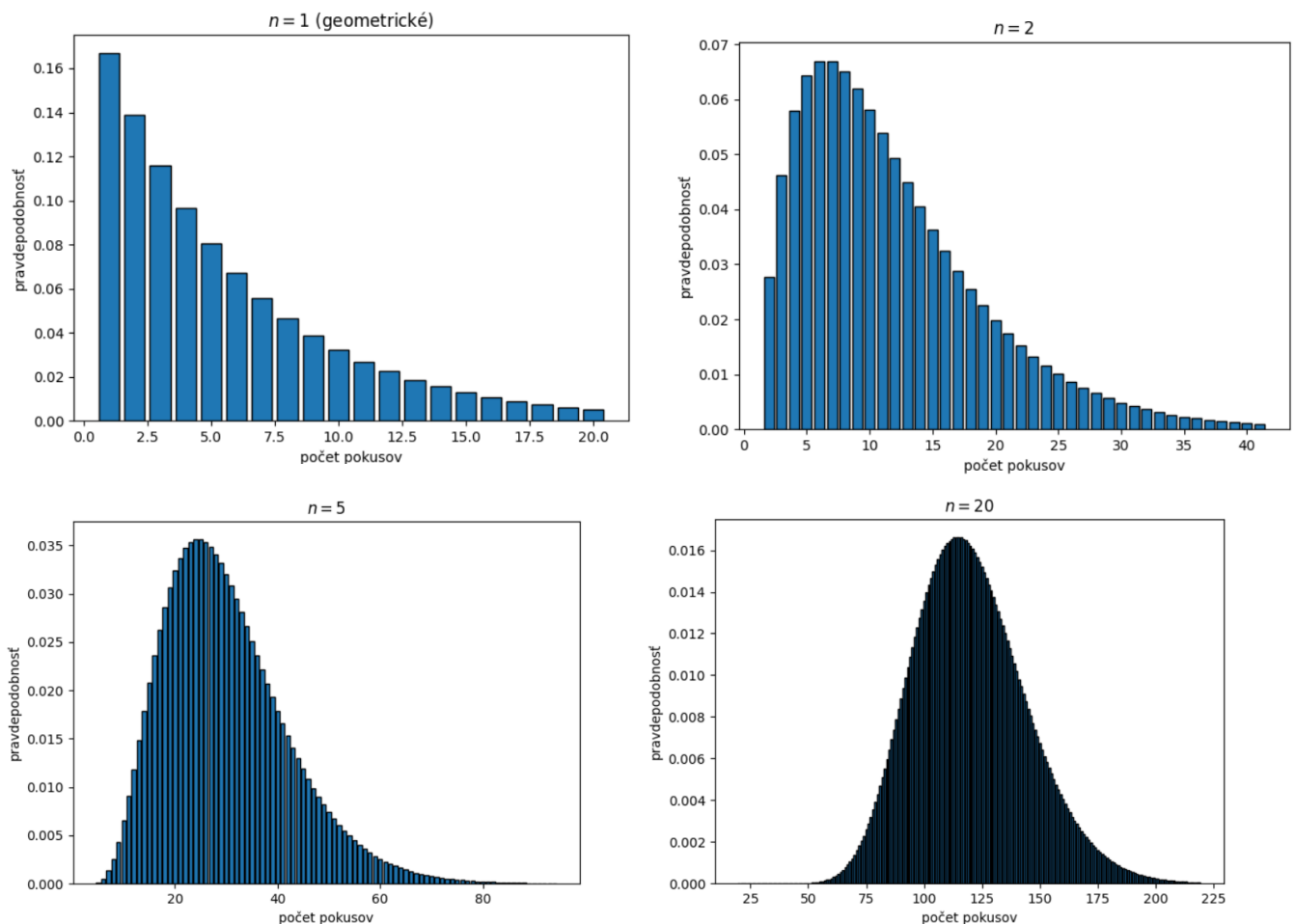
2. Príklad: Geometrické rozdelenie

Majme znovu postupnosť nezávislých pokusov v s pravdepodobnosťou úspechu p . Ak definujeme $X_1, X_2, \dots, X_n$ ako nezávislé náhodné premenné s geometrickým rozdelením, tak každá reprezentuje počet pokusov potrebných na dosiahnutie jedného úspechu. Ich súčet X potom vyjadruje počet pokusov potrebných na dosiahnutie n -tého úspechu. Pravdepodobnosti získame kombinatorickou úvahou ako

$$\mathbb{P}(X = k) = \binom{k-1}{n-1} p^n (1-p)^{k-n}$$

pričom možné hodnoty sú prirodzené čísla väčšie alebo rovné n .

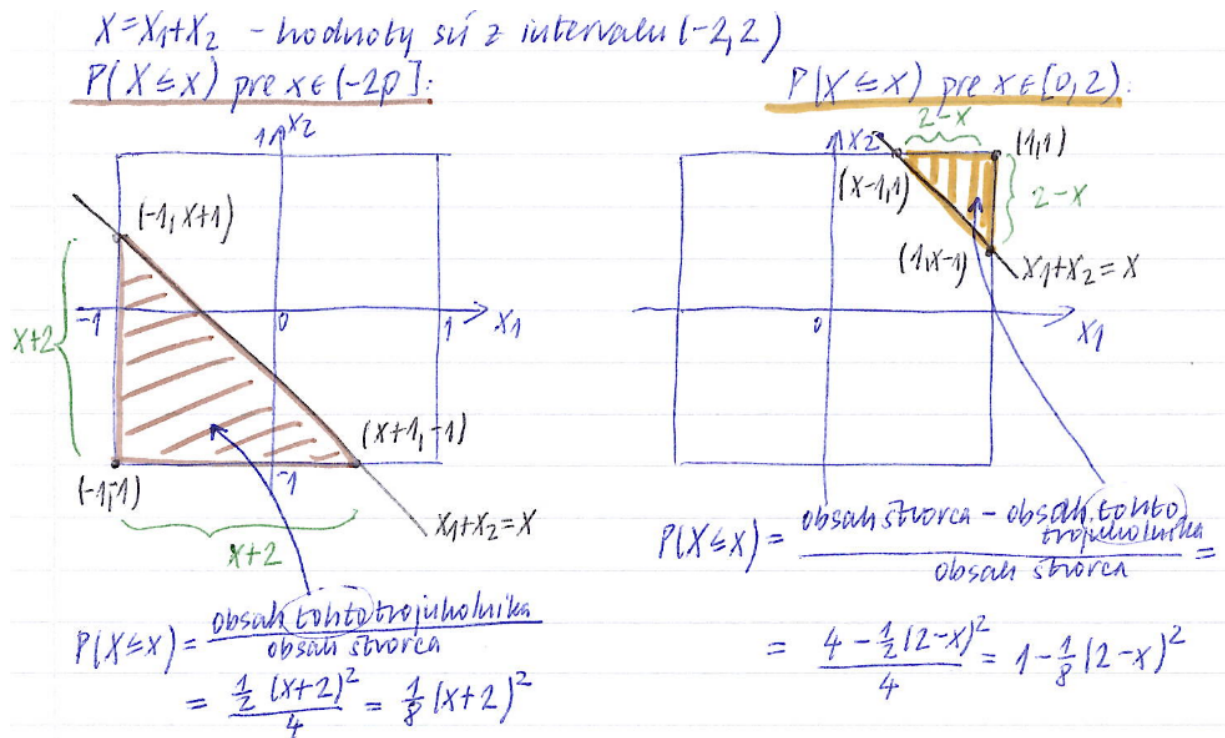
Ilustrácia pre $p = \frac{1}{6}$ (hádzame desaťkrát kockou, chceme hodiť šestku), opakujeme to n krát.



3. Príklad: rovnomerné rozdelenie

Uvažujme rovnomerné rozdelenie na $(-1, 1)$ a nezávislé náhodné premenné $X_1, X_2, \dots, X_n$ s týmto rozdelením.

Pre $n=2$ je odvozenie hustoty súčtu dobré cvičenie na geometrickú pravdepodobnosť, výpočet distribučnej funkcie a hustoty:

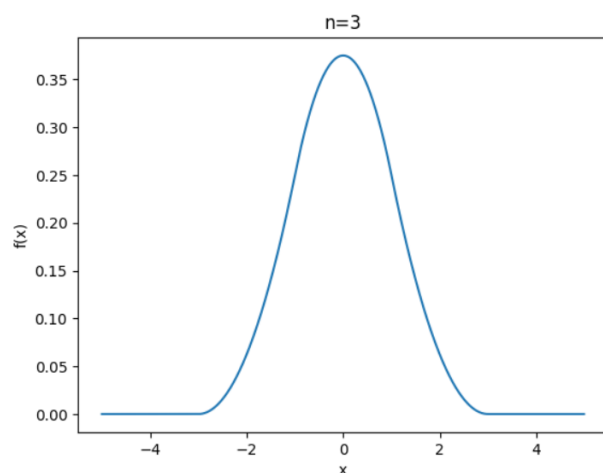
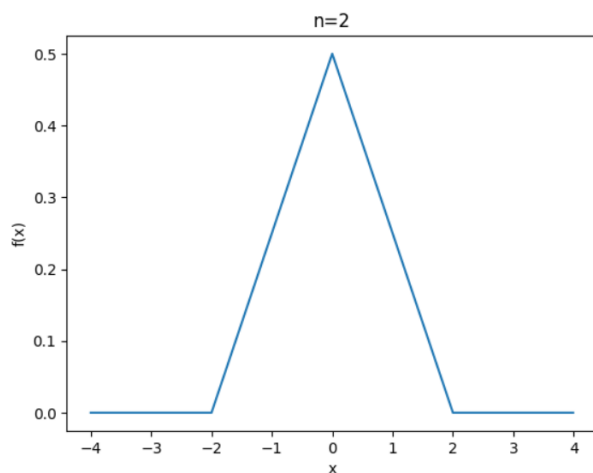


$$F(x) = P(X \leq x) = \begin{cases} 0 & \text{pre } x \leq -2 \\ \frac{1}{8}(x+2)^2 & \text{pre } x \in (-2, 0] \\ 1 - \frac{1}{8}(2-x)^2 & \text{pre } x \in [0, 2] \\ 1 & \text{pre } x \geq 2 \end{cases}$$

$$f(x) = F'(x) = \begin{cases} \frac{1}{4}(x+2) & \text{pre } x \in (-2, 0] \\ \frac{1}{4}(2-x) & \text{pre } x \in [0, 2] \\ 0 & \text{inak} \end{cases}$$

Pre $n=3$ zoberieme výsledok z knihy:

$$h(x) = \begin{cases} 0 & \text{pro } |x| \geq 3, \\ \frac{(3 - |x|)^2}{16} & \text{pro } 1 \leq |x| \leq 3, \\ \frac{3 - x^2}{8} & \text{pro } 0 \leq |x| \leq 1. \end{cases}$$



Je známa aj všeobecná hustota pre n sčítancov:

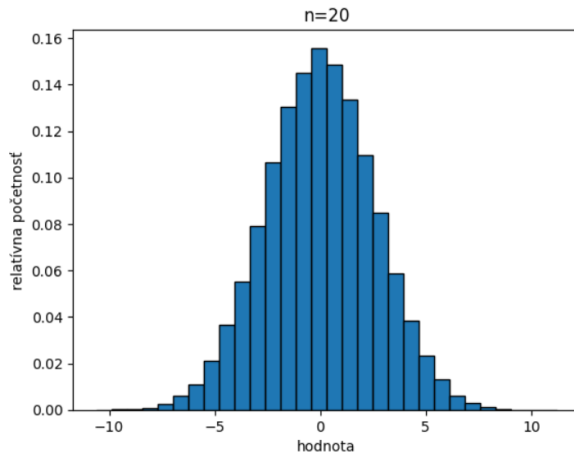
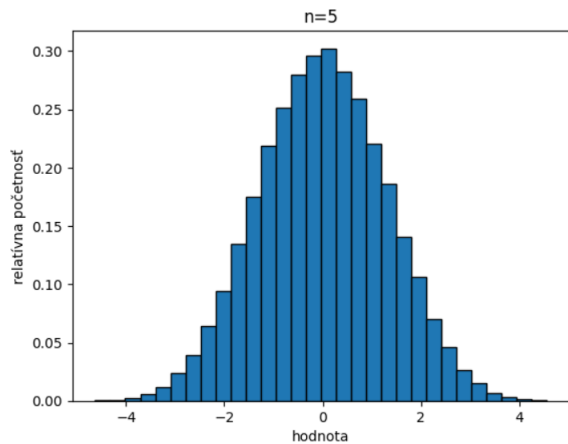
$$f_n(x) = \begin{cases} \frac{1}{2^n(n-1)!} \sum_{k=0}^{\lfloor (n+x)/2 \rfloor} (-1)^k \binom{n}{k} (n+x-2k)^{n-1} & \text{pro } |x| < n, \\ 0 & \text{pro ostatní } x. \end{cases}$$

Ale jednoduchšie bude spraviť si simuláciu a súčty $X_1 + \dots + X_n$ si vygenerovať:

```
import numpy as np
from scipy.stats import uniform

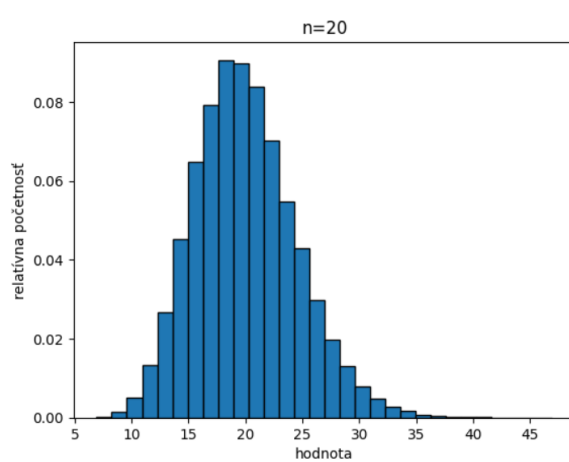
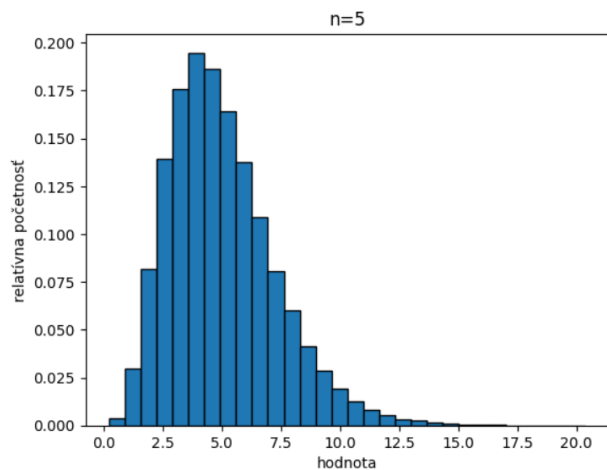
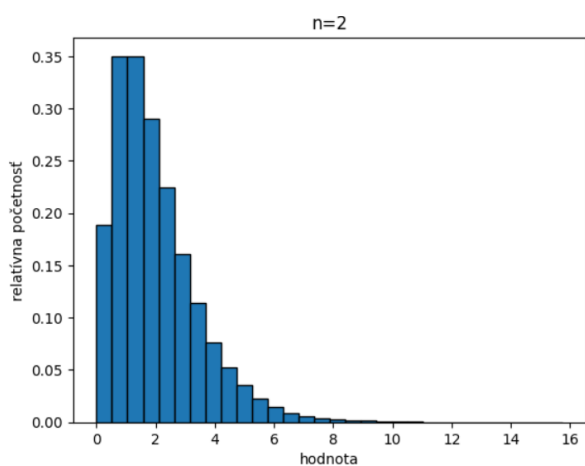
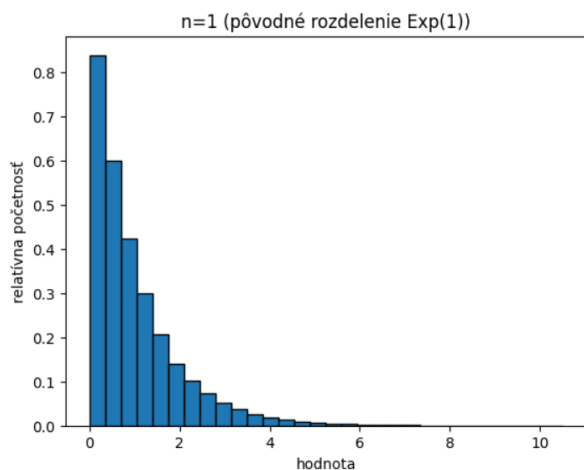
def sucet_rovnomernych(loc, scale, n):
    X_i = uniform.rvs(loc=loc, scale=scale, size=n)
    return np.sum(X_i)

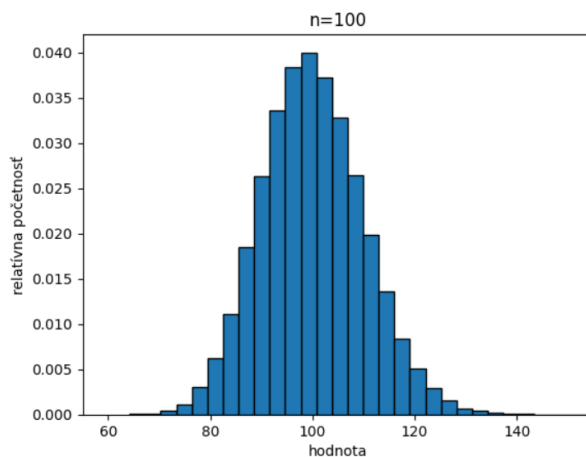
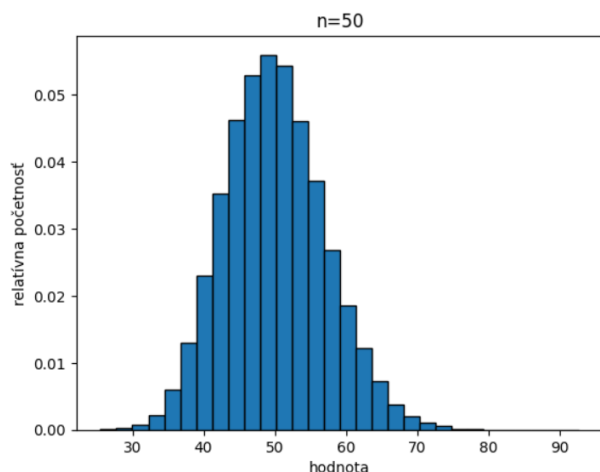
n = 20
sucty = [sucet_rovnomernych(-1, 2, n) for _ in range(10**5)]
```



4. Príklad: exponenciálne rozdelenie (simulačne)

Súčet $X_1 + \dots + X_n$ nezávislých exponenciálnych náhodných premenných s rovnakým parametrom má tiež známe rozdelenie (tzv. gama rozdelenie), ale tiež si ho pre názornosť vygenerujeme:

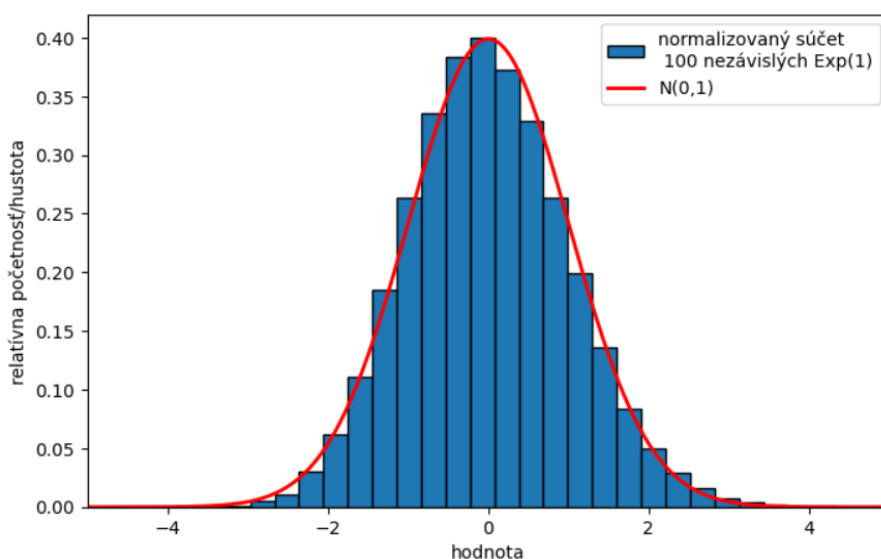




5. Centrálna limitná veta

Tvrdenie:

- Technický predpoklad: uvažované rozdelenie má strednú hodnotu a disperziu.
- Majme súčet $X_1 + \dots + X_n$ nezávislých náhodných premenných s rovnakým rozdelením.
- Normalizujme ho, teda odpočítajme strednú hodnotu a vydeľme štandardnou odchýlkou. Získaná náhodná premenná má nulovú strednú hodnotu a jednotkovú disperziu.
- Tieto normalizované náhodné premenné konvergujú pre n idúce do nekonečna k normálnemu rozdeleniu $N(0,1)$.



Prakticky:

- Súčet $X_1 + \dots + X_n$ nezávislých náhodných premenných s rovnakým rozdelením má pre veľké n približne normálne rozdelenie. Za strednú hodnotu a disperziu tohto aproximatívneho rozdelenia zoberieme strednú hodnotu a disperziu súčtu.

Ako dôsledok:

- Priemer nezávislých náhodných premenných s rovnakým rozdelením má pre veľké n približne normálne rozdelenie. Za strednú hodnotu a disperziu tohto aproximatívneho rozdelenia zoberieme strednú hodnotu a disperziu priemeru.

Stredná hodnota a disperzia súčtu a priemeru:

$$E(X_i) = \mu, D(X_i) = \sigma^2, X_i \text{ nezávislé}$$

$$S = X_1 + \dots + X_n$$

$$E(S) = E(X_1) + \dots + E(X_n) = n\mu, D(S) \overset{\text{nez.}}{=} D(X_1) + \dots + D(X_n) = n\sigma^2$$

$$\bar{X} = \frac{1}{n} S$$

$$E(\bar{X}) = \frac{1}{n} E(S) = \frac{\mu}{n}, D(\bar{X}) = D\left(\frac{1}{n} S\right) = \frac{1}{n^2} D(S) = \frac{\sigma^2}{n}$$

6. Príklad 1

Poistovňa vydala 1250 poistiek, týkajúcich sa starostlivosti o zrak. Počet nárokov, ktoré poistenec uplatní počas roka, má Poissonovo rozdelenie so strednou hodnotou 2. Predpokladajte, že nároky jednotlivých poistencov sú nezávislé. Aká je približná pravdepodobnosť, že celkový počet nárokov počas roka bude medzi 2450 a 2600?

(A) 0,68 (B) 0,82 (C) 0,87 (D) 0,95 (E) 1,00

Zdroj: Society of Actuaries

Riešenie:

$$X_i \sim Po(\lambda), \lambda = E(X_i) = 2 \Rightarrow D(X_i) = 2$$

$$S = X_1 + \dots + X_n, n = 1250$$

$$P(2450 \leq S \leq 2600) = ?$$

$$CLV \Rightarrow S \overset{\text{aprox.}}{\sim} N(n \cdot E(X_i), n \cdot D(X_i))$$

$$N(1250 \cdot 2, 1250 \cdot 2)$$

$$N(2500, 2500)$$

$$P(2450 \leq S \leq 2600) \doteq P(2450 \leq N(2500, 2500) \leq 2600) = F(2600) - F(2450)$$

F je tu distribučná funkcia normálneho rozdelenia **N(2500, 2500)**. Python chce štandardnú odchýlku, nie disperziu, preto:

```
from scipy.stats import norm
```

```
pr1 = norm.cdf(2600, loc=2500, scale=50) - norm.cdf(2450, loc=2500, scale=50)
print(f"{pr1:.4f}")
```

0.8186

Ak potrebujeme výsledok vyjadriť pomocou distribučnej funkcie normalizovaného normálneho rozdelenia **N(0, 1)**:

$$\begin{aligned} P(2450 \leq S \leq 2600) &\doteq P(2450 \leq N(2500, 2500) \leq 2600) = \\ &= P\left(\frac{2450 - 2500}{\sqrt{2500}} \leq \frac{N(2500, 2500) - 2500}{\sqrt{2500}} \leq \frac{2600 - 2500}{\sqrt{2500}}\right) \\ &\quad \underbrace{\hspace{10em}}_{N(0,1)} \\ &= P(-1 \leq N(0,1) \leq 2) = \Phi(2) - \Phi(-1) \end{aligned}$$

kde Φ je distribučná funkcia $N(0,1)$

```
pr1_inak = norm.cdf(2) - norm.cdf(-1)
print(f"{pr1_inak:.4f}")
```

0.8186

7. Príklad 2

Pri analýze údajov o zdravotnej starostlivosti bol vek pacientov zaokrúhlený na najbližší násobok 5 rokov. Rozdiel medzi skutočným a zaokrúhleným vekom má rovnomerne rozdelenie od -2.5 do 2.5 rokov. Náhodne sme vybrali 48 pacientov. Aká je približná pravdepodobnosť, že priemer zaokrúhlených vekov je od priemeru skutočných vekov vzdialený nie viac ako štvrt' roka?

- (A) 0,14 (B) 0,38 (C) 0,57 (D) 0,77 (E) 0,88

Zdroj: Society of Actuaries

Výsledok vyjadríme pomocou distribučnej funkcie normalizovaného normálneho rozdelenia $N(0, 1)$.

Riešenie:

$$X_i \sim \text{Ro}(-2,5; 2,5) \Rightarrow E(X_i) = 0, D(X_i) = \frac{25}{12}$$

$$X_1, \dots, X_n \text{ nezávislé, } n = 48$$

$$P\left(\left|\frac{1}{n} \sum \text{zaokrúhlené}_i - \frac{1}{n} \sum \text{presné}_i\right| \leq \frac{1}{4}\right) = ?$$

$$\frac{1}{n} \sum (\text{zaokrúhlené} - \text{presné})_i = \frac{1}{n} \sum X_i = \bar{X}$$

$$E(\bar{X}) = 0$$

$$D(\bar{X}) = \frac{1}{n} \cdot D(X_i) = \frac{1}{48} \cdot \frac{25}{12} = \frac{25}{24^2}$$

$$P\left(|\bar{X}| \leq \frac{1}{4}\right) = P\left(-\frac{1}{4} \leq \bar{X} \leq \frac{1}{4}\right) \doteq P\left(-\frac{1}{4} \leq N\left(0, \frac{25}{24^2}\right) \leq \frac{1}{4}\right) =$$

$$= P\left(\frac{-\frac{1}{4} - 0}{\sqrt{\frac{25}{24^2}}} \leq \frac{N\left(0, \frac{25}{24^2}\right) - 0}{\sqrt{\frac{25}{24^2}}} \leq \frac{\frac{1}{4} - 0}{\sqrt{\frac{25}{24^2}}}\right) = P\left(-\frac{24}{4 \cdot 5} \leq N(0, 1) \leq \frac{24}{4 \cdot 5}\right)$$

$\sqrt{\frac{25}{24^2}} = \frac{5}{24} \quad N(0, 1)$

$$= P(-1,2 \leq N(0, 1) \leq 1,2) = \Phi(1,2) - \Phi(-1,2)$$

```
pr2 = norm.cdf(1.2) - norm.cdf(-1.2)
```

```
print(f"{pr2:.4f}")
```

0.7699

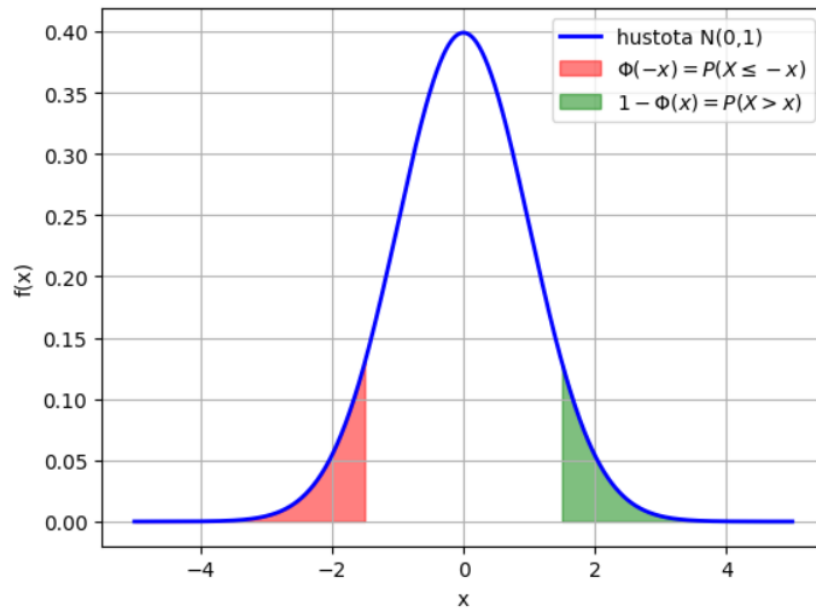
8. Symetria distribučnej funkcie $N(0, 1)$

Pre distribučnú funkciu normalizovaného normálneho rozdelenia $N(0, 1)$ platí:

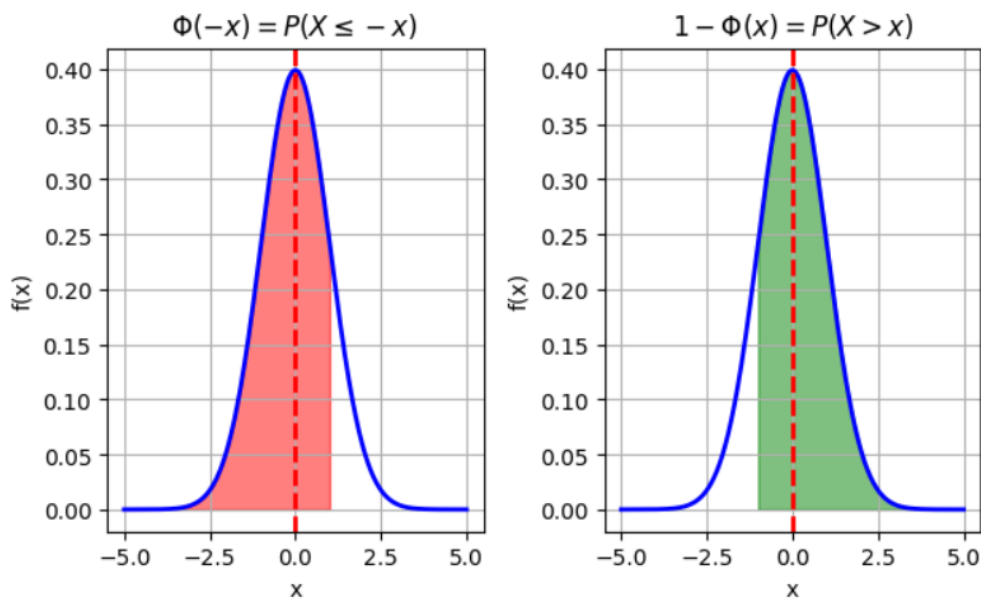
$$\Phi(-x) = 1 - \Phi(x)$$

Prečo:

- ak je x kladné:



- ak je x záporné:



- V bode $x=0$ je hodnota distribučnej funkcie rovná $\frac{1}{2}$ a tvrdenie tiež platí.

Toto môžeme využiť na taký zápis, že vo výsledku bude distribučná funkcia $N(0, 1)$ len s kladnými argumentami.

Takže v predchádzajúcom príklade:

```
pr2_inak = 2*norm.cdf(1.2) - 1  
print(f"{pr2_inak:.4f}")
```

0.7699