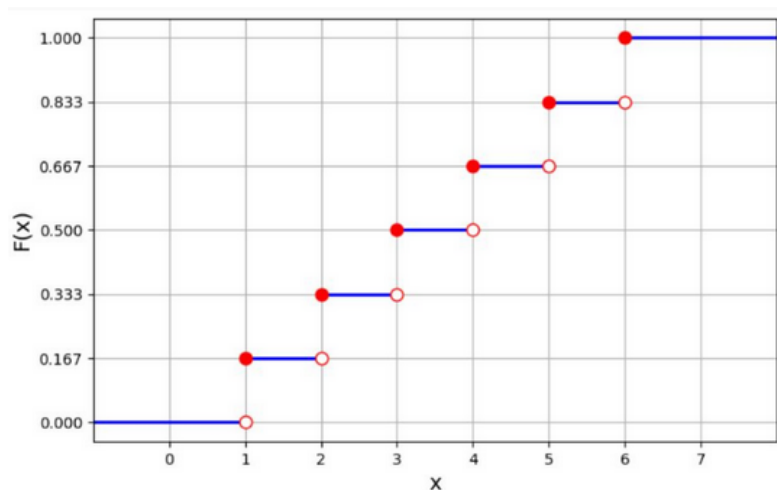


(04) Spojité náhodné premenné

1. Úvodné príklady

Zaoberali sme sa **diskrétnymi náhodnými premennými**, ktoré nadobúdali konečne alebo spočítateľne veľa hodnôt, napríklad počet bodov pri hádzaní kockou. Distribučná funkcia je v tom prípade skokovitá, veľkosť skoku je pravdepodobnosť príslušnej hodnoty:



Nie každá náhodná premenná má takýto charakter, napríklad sme mali úlohu:

Romeo a Júlia sa chcú stretnúť. Na miesto stretnutia prídu v náhodnom čase, ktorý si každý nezávisle zvolí *rovnomerne náhodne* medzi 18:00 a 20:00. Kto príde prvý, počká na druhého.

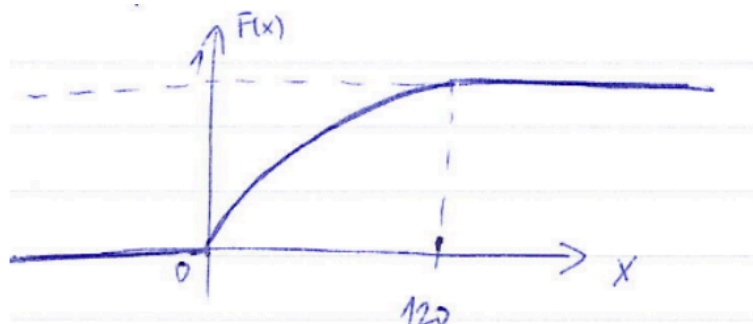
Diskrétnou náhodnou premennou nie je

- Príchod Romea (analogicky príklad Júlie) – možná je ľubovoľná hodnota medzi 0 a 120 (ak čas meriame v minútach od 18:00)
- Čas čakania – tiež ľubovoľná hodnota medzi 0 a 120

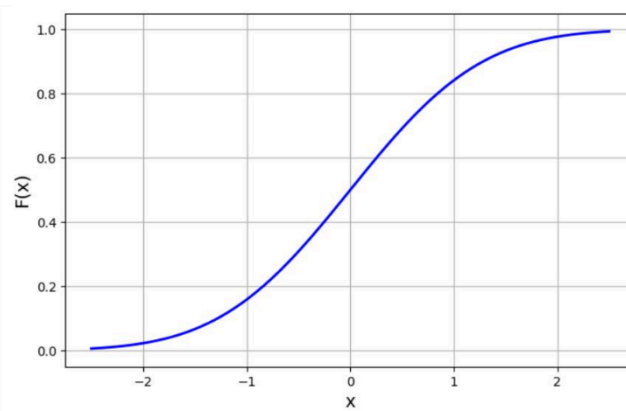
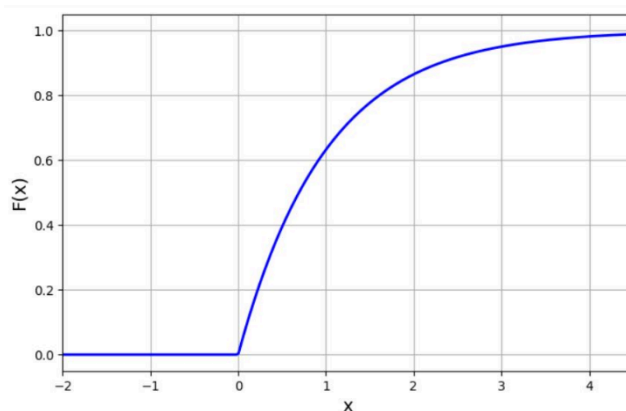
Predpoklad o “rovnomernom rozdelení” príchodu intuitívne rozumieme, takisto tomu, čo je “náhodné číslo rovnomerne zvolené medzi 0 a 1 v Pythone”:

```
n = 10**4
x = np.random.rand(n) * 120
y = np.random.rand(n) * 120
cakanie = np.abs(x - y)
```

Pre čas čakania sme odvodili distribučnú funkciu, ktorá naozaj nemá skok:



Ukazovali sme si aj iné príklady distribučných funkcií (len graficky, bez predpisu pre funkciu): vľavo náhodná premenná nadobúda ľubovoľné kladné hodnoty, vpravo ľubovoľné reálne.



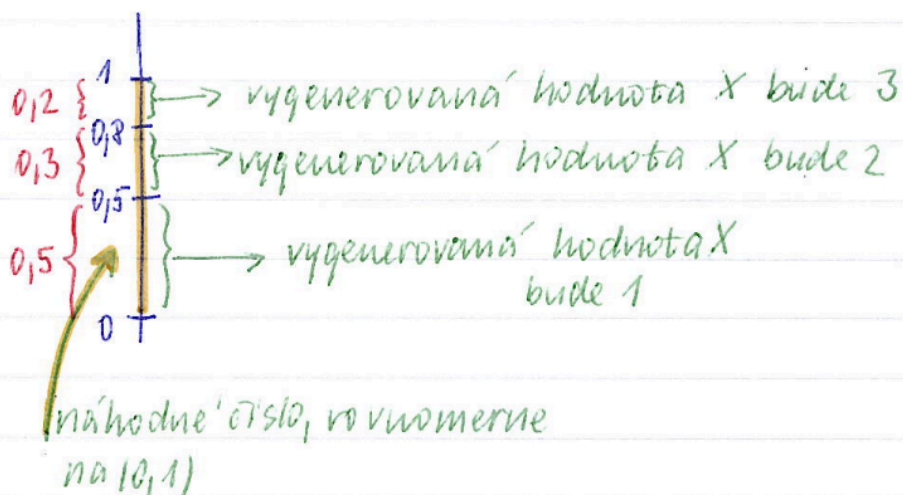
2. Generovanie náhodných čísel z rozdelenia daného distribučnou funkciou

Ak by sme mali iba generátor rovnomerne rozložených čísel z intervalu $(0, 1)$, mohli by sme ho použiť na generovanie hodnôt diskkrétnej náhodnej premennej.

Napríklad majme náhodnú premennú s rozdelením:

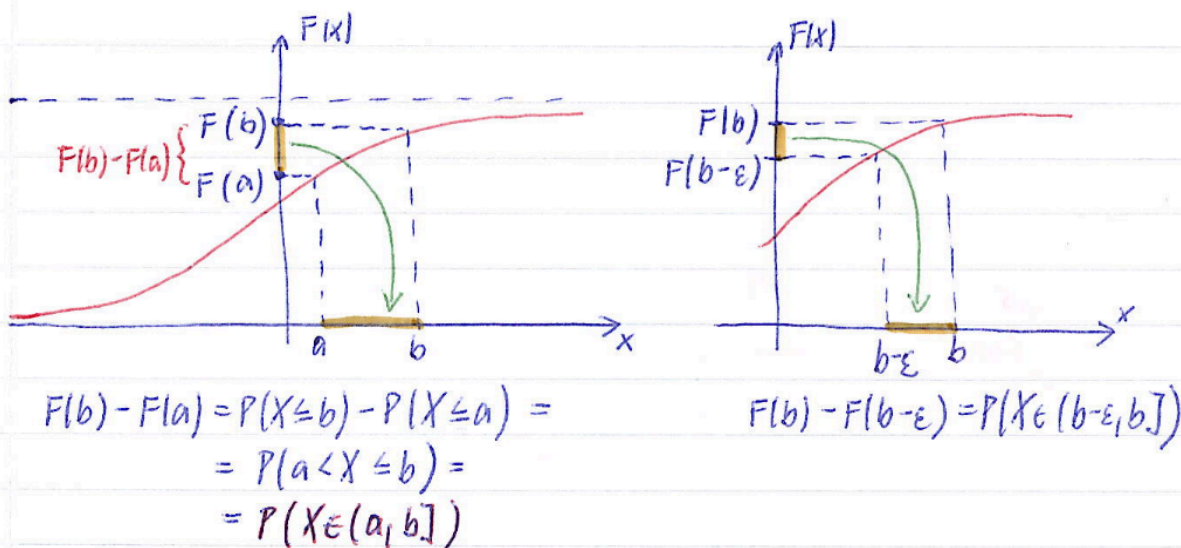
x_i	1	2	3
$P(X=x_i)$	0,5	0,3	0,2

Potom môžeme postupovať nasledovne:



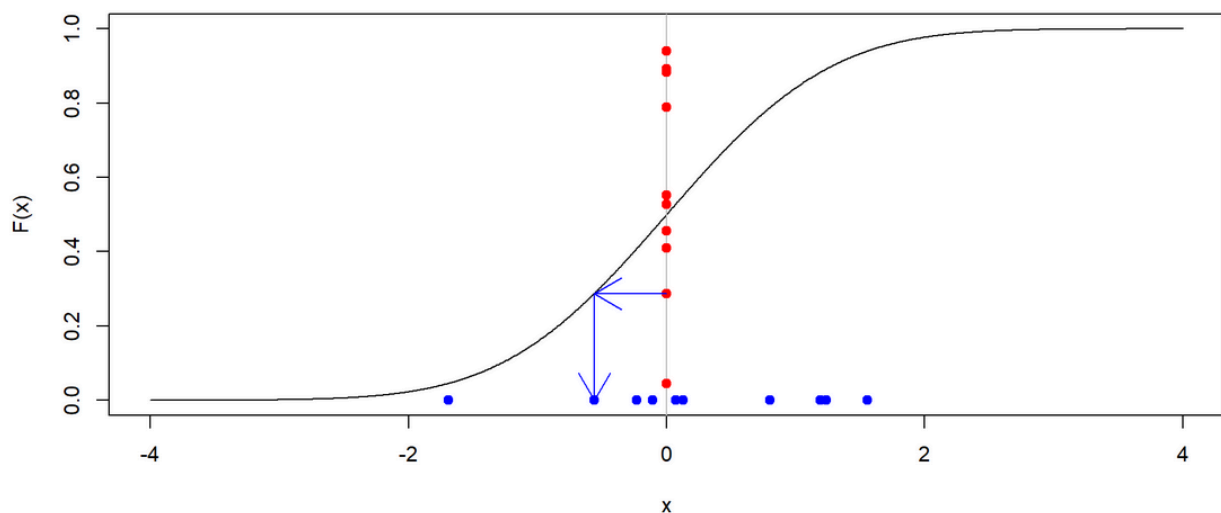
Teraz by sme chceli podobný postup aj pre náhodné premenné s horeuvedenými distribučnými funkciami. Budeme predpokladať, že distribučná funkcia F nášho rozdelenia (z ktorého chceme generovať náhodné čísla) je spojitá a rastúca (okrem hodnôt 0 a 1, ktoré sa môžu nadobúdať na intervaloch vľavo, resp. vpravo - príslušné hodnoty daná náhodná premenná nenadobúda).

Myšlienka:



Postup:

- Generujeme čísla s rovnomerným rozdelením na $(0, 1)$.
- Dosadzujeme ich do inverznej funkcie k danej distribučnej funkcii F
- Výsledkom je náhodné číslo z rozdelenia s distribučnou funkciou F .



Prečo to funguje:

Označenie: Y je číslo z rovnomerného rozdelenia na $(0, 1)$, F je distribučná funkcia rozdelenia, z ktorého chceme generovať náhodné číslo.

Nájdeme distribučnú funkciu náhodného čísla, ktoré vznikne navrhnutým spôsobom a ukážeme, že sa zhoduje s F . Teda náš postup dáva to, čo potrebujeme.

$$\begin{aligned}
 Y &= F(X), \text{ t.j. } X = F^{-1}(Y) \\
 F_X(x) &= P(X \leq x) = P(F^{-1}(Y) \leq x) \stackrel{\text{F je rastúca}}{=} P(F(F^{-1}(Y)) \leq F(x)) = \\
 &= P(Y \leq F(x)) \stackrel{\substack{Y \text{ má rovnomerné} \\ \text{rozd. na } (0, 1)}}{=} F(x)
 \end{aligned}$$

Príklad 1: náhodná premenná vyjadrujúca čas čakania pri stretnutí Romea a Júlie:

$$F(x) = P(\text{čakanie} \leq x) = \begin{cases} 0 & \text{pre } x < 0 \\ 1 - \frac{(120-x)^2}{120^2} & \text{pre } x \in [0, 120] \\ 1 & \text{pre } x > 120 \end{cases}$$

$y \in (0,1)$
 Máme $y = F(x)$ $\rightarrow$ chceme vyjadriť $x \in (0,120)$

$$y = 1 - \frac{(120-x)^2}{120^2}$$

$$(120-x)^2 = (1-y) \cdot 120^2$$

$$120-x = \pm \sqrt{(1-y) \cdot 120^2}$$

$$\frac{120 - \sqrt{120^2(1-y)}}{120(1 - \sqrt{1-y})} = x$$

$$\frac{120 - \sqrt{120^2(1-y)}}{120(1 - \sqrt{1-y})} = x$$

Python:

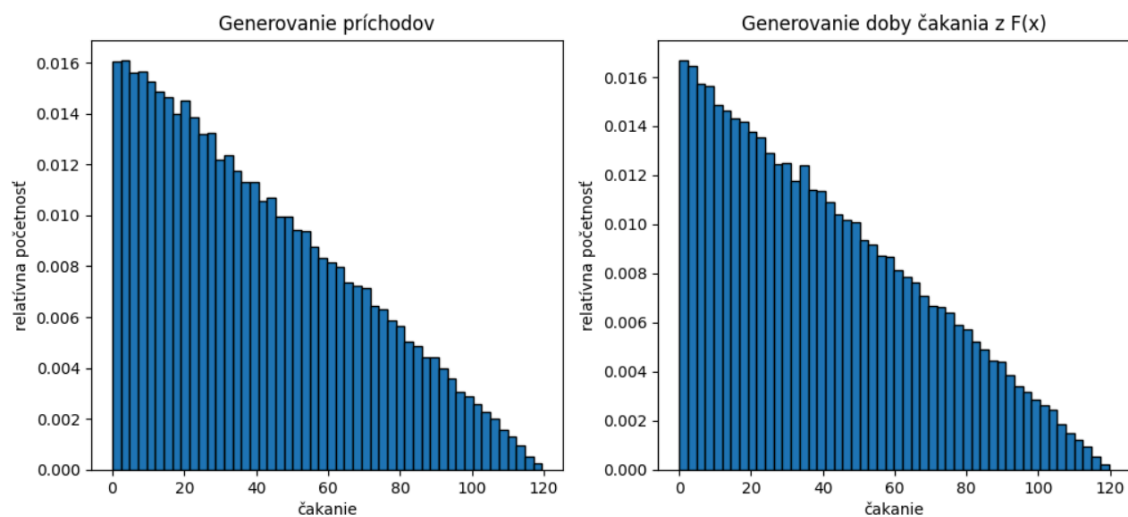
```
n = 5
rovnomerne = np.random.rand(n)
print(rovnomerne)
```

```
[0.70900422 0.63573366 0.29063671 0.44699994 0.32389354]
```

```
nase = 120*(1 - np.sqrt(1 - rovnomerne))
print(nase)
```

```
[55.26716991 47.57462258 18.9315511 30.76323132 21.32916801]
```

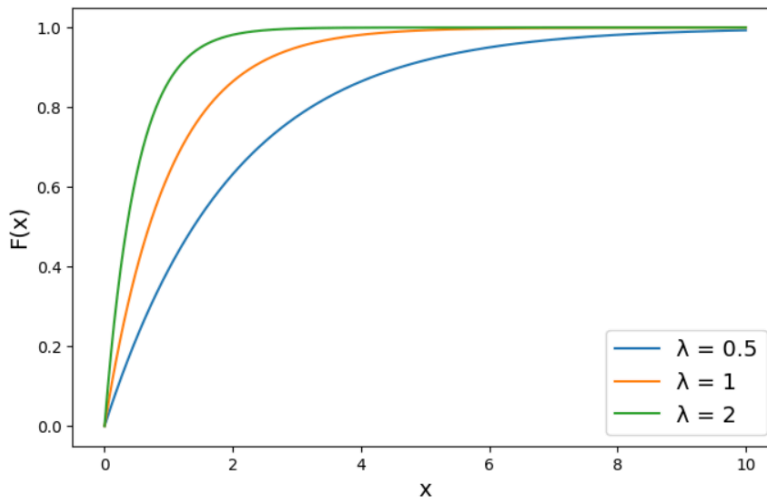
Výsledok + porovnanie s generovaním času čakania pomocou príchodov Romea a Júlie:



Príklad 2:

Majme náhodnú premennú s distribučnou funkciou (λ je kladný parameter)

$$F(x) = \begin{cases} 0, & x < 0 \\ 1 - e^{-\lambda x}, & x \geq 0 \end{cases}$$



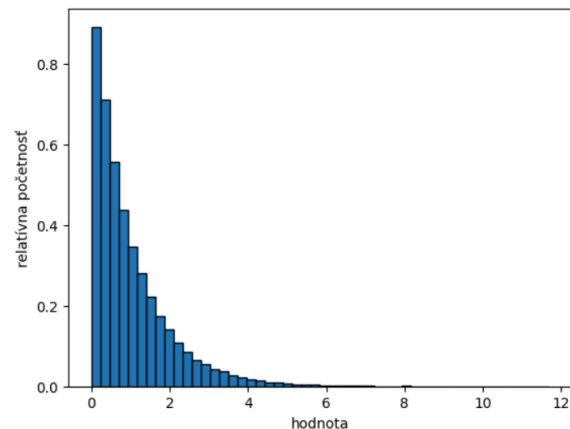
Je to spojitá funkcia, nulová pre x záporné a rastúca pre x nezáporné, nulová v mínus nekonečne a jednotka v limite pre plus nekonečno - je to distribučná funkcia nejakej náhodnej premennej. Chceme z nej generovať náhodné čísla.

Odvodenie postupu:

$$\begin{aligned} y &= 1 - e^{-\lambda x} \rightarrow \text{chceme vyjadriť } x > 0 \\ e^{-\lambda x} &= 1 - y \\ -\lambda x &= \ln(1 - y) \\ x &= \left\lceil -\frac{1}{\lambda} \ln(1 - y) \right\rceil \end{aligned}$$

Poznámka: Ak Y má rovnomerné rozdelenie na $(0, 1)$, tak aj $1 - Y$ má rovnomerné rozdelenie na $(0, 1)$. Takže môžeme spraviť napríklad pre **lambda=1**:

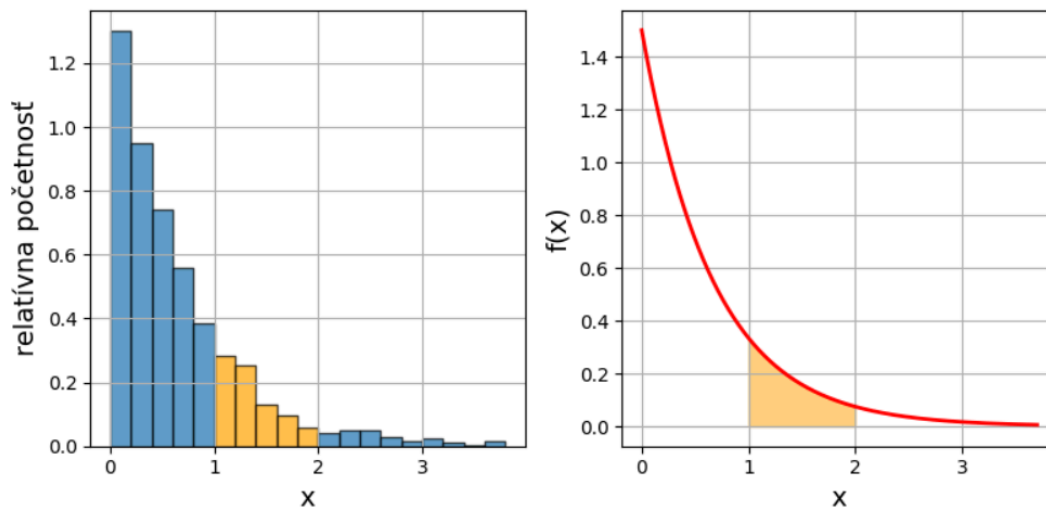
```
n = 10**5
rovnomerne = np.random.rand(n)
nase2 = -np.log(rovnomerne)
```



3. Čo sú spojité náhodné premenné + funkcia hustoty

Distribučnej funkcii náhodnej premennej zodpovedá empirická distribučná funkcia vypočítaná z dát. Ak máme diskretnú náhodnú premennú, analógiou grafu početností z dát je tabuľka s pravdepodobnosťami možných hodnôt náhodnej premennej. Aká je analógia histogramu hodnôt pre náhodnú premennú, ktorá nadobúda nespočítateľne veľa hodnôt (ľubovoľnú hodnotu z intervalu, ľubovoľné kladné číslo, ľubovoľné reálne číslo)?

Myšlienka:



Matematický zápis:

$$P(a \leq X \leq b) = \int_a^b f(x) dx$$

Špeciálne platí:

$$F(x) = P(X \leq x) = \int_{-\infty}^x f(t) dt$$

Definícia spojitej náhodnej premennej:

Náhodná premenná sa nazýva **spojitá**, ak existuje funkcia **f**, pre ktorú platí

$$F(x) = \int_{-\infty}^x f(t) dt$$

Funkcia **f** sa nazýva **hustota** náhodnej premennej.

Funkcia hustoty spĺňa

$$f(x) \geq 0, \quad \int_{-\infty}^{\infty} f(x) dx = 1$$

A naopak, každá funkcia s touto vlastnosťou definuje náhodnú premennú.

Príklad 3

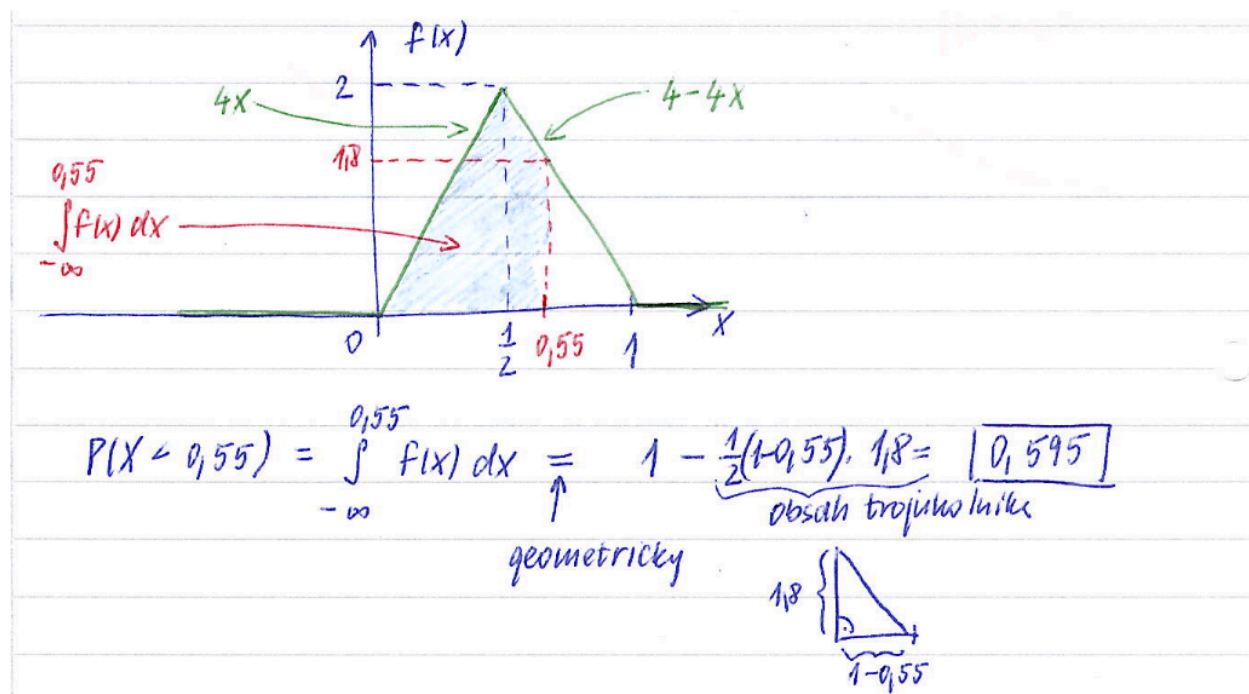
5.7 The score of a student on a certain exam is represented by a number between 0 and 1. Suppose that the student passes the exam if this number is at least 0.55. Suppose we model this experiment by a continuous random variable S , the score, whose probability density function is given by

$$f(x) = \begin{cases} 4x & \text{for } 0 \leq x \leq \frac{1}{2} \\ 4 - 4x & \text{for } \frac{1}{2} \leq x \leq 1 \\ 0 & \text{elsewhere.} \end{cases}$$

a. What is the probability that the student fails the exam?

Zdroj: A Modern Introduction to Probability and Statistics. Understanding Why and How
Z fakultnej siete je kniha dostupná na <https://link.springer.com/book/10.1007/1-84628-168-7>

Riešenie:



4. Vlastnosti spojitej náhodnej premennej

- **Distribučná funkcia spojitej náhodnej premennej je spojitá** (lebo integrál spojitosti závisí od svojej hornej hranice).
- Uvidíme, že hustota nemusí byť spojitá.
- Hustota sa dá vypočítať z distribučnej funkcie derivovaním:

$$F(x) = \int_{-\infty}^x f(t) dt \quad \Rightarrow \quad f(x) = \frac{d}{dx} F(x)$$

- **Pravdepodobnosť, že nastane jedna konkrétna hodnota, je nulová.**
 - Vyplýva to z toho, že integrál cez jeden bod (plocha pod krivkou v jednom bode) sa rovaná nule. Alebo aj z distribučnej funkcie, ktorá nikde nemá skok.
 - Z toho napríklad vyplýva, že intervaly **(a, b)**, **[a, b]**, **[a, b)**, **(a, b]** majú rovnakú pravdepodobnosť.
 - Pravdepodobnosť každého z nich je **F(b) - F(a)**.
 - Analogicky, pravdepodobnosti udalostí **X < a**, **X ≤ a** sú rovnaké. A pod.
- Ak náhodná premenná nenadobúda nejaké hodnoty (napríklad z nejakého intervalu), jej hustota je tam nulová. To môže byť užitočné pri výpočte hustoty náhodnej premennej – na základe jej možných hodnôt napíšeme, kde je hustota nulová a zaoberáme sa potom výpočtom jej nenulových hodnôt.
- Aj naopak, ak je na nejakom intervale hustota nulová, tak pravdepodobnosť, že sa nadobudne hodnota z toho intervalu, je nulová (pri jej výpočte integrujeme nulu).

5. Príklady: Hustota náhodnej premennej

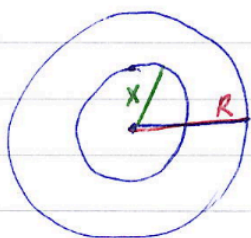


Príklad 4

Strieľame šípkami do kruhového terča s polomerom **R**. Predpokladajme, že ho vždy trafíme, a miesto, ktoré trafíme, je *rovnomerne náhodné* na ploche terča (teda pravdepodobnosť zásahu ľubovoľnej oblasti je úmerná jej ploche). Odvodte hustotu náhodnej premennej **X**, ktorá vyjadruje vzdialenosť zasiahnutého bodu od stredu terča.

Riešenie:

$$F(x) = P(X \leq x) = \frac{\pi x^2}{\pi R^2} = \frac{x^2}{R^2} \quad \text{pre } x \in (0, R)$$



$$f(x) = F'(x) = \frac{2x}{R^2} \quad \text{pre } x \in (0, R)$$

Mimo tohto intervalu je $f(x) = 0$ (tie hodnoty X nenadobúdajú)

Príklad 5

5.9 □ Suppose we choose arbitrarily a point from the square with corners at (2,1), (3,1), (2,2), and (3,2). The random variable A is the area of the triangle with its corners at (2,1), (3,1) and the chosen point (see Figure 5.9).

- a. What is the largest area A that can occur, and what is the set of points for which $A \leq 1/4$?

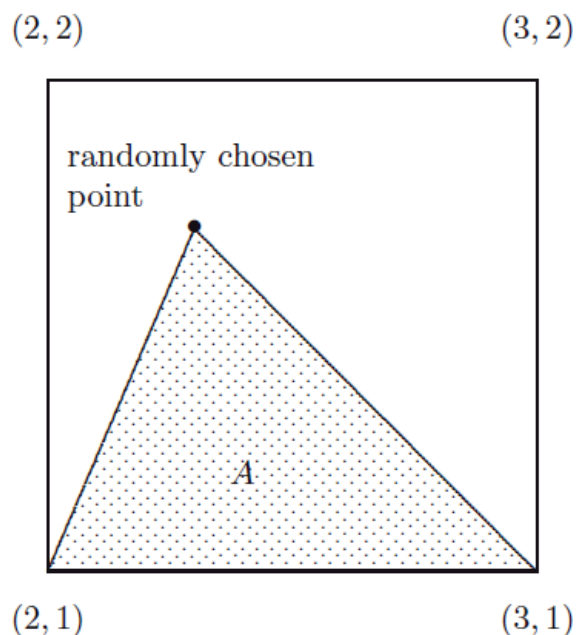
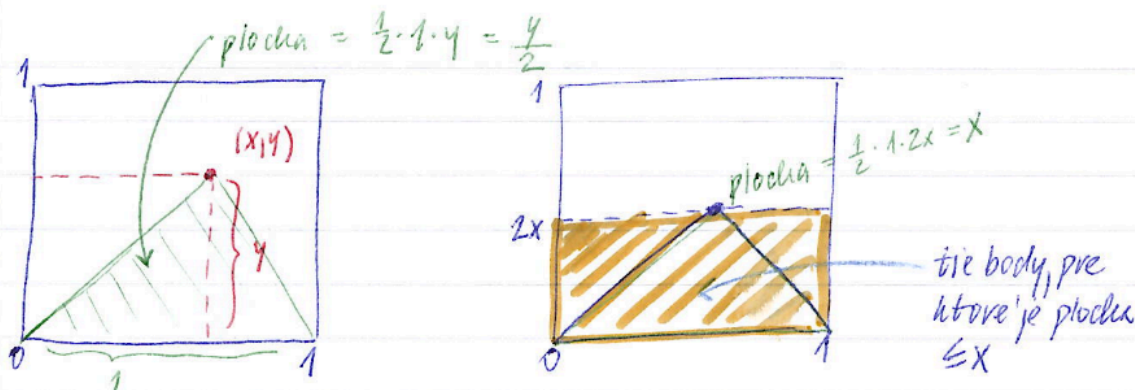


Fig. 5.9. A triangle in a square.

- b. Determine the distribution function F of A .
c. Determine the probability density function f of A .

Zdroj: A Modern Introduction to Probability and Statistics. Understanding Why and How

Riešenie:

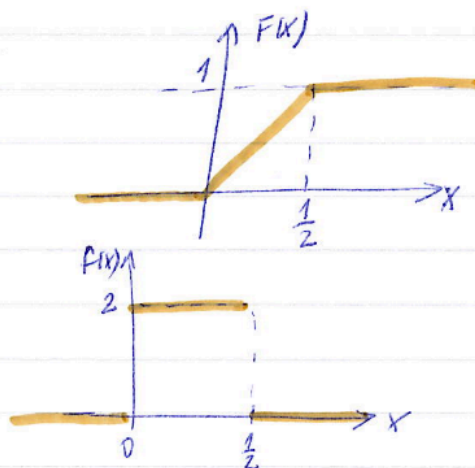


$$F(x) = P(A \leq x) = \frac{1 \cdot 2x}{1} = 2x \quad \text{pre } x \in (0, \frac{1}{2})$$

$$F(x) = \begin{cases} 0 & \text{pre } x \leq 0 \\ 2x & \text{pre } x \in (0, \frac{1}{2}) \\ 1 & \text{pre } x \geq \frac{1}{2} \end{cases}$$

$$f(x) = F'(x) = 2 \quad \text{pre } x \in (0, \frac{1}{2})$$

indek je $f(x) = 0$



6. Stredná hodnota spojitej náhodnej premennej

- **Definícia strednej hodnoty** (za podmienky, že integrál absolútne konverguje):

$$E(X) = \int_{-\infty}^{\infty} x f(x) dx$$

Príklad 6: Vypočítame strednú hodnotu čakania v príklade o stretnutí Romea a Júlie a porovnáme s priemerom zo simulácií.

```
priemerne_cakanie = np.mean(cakanie)
print(f"{priemerne_cakanie:.4f}")
```

39.9153

Riešenie:

Najskôr potrebujeme hustotu (všimnime si, že sa podobá na histogram zo simulácií):

$$F(x) = 1 - \frac{(120-x)^2}{120^2} \quad \text{pre } x \in [0, 120]$$

$$\text{pre tieto } x \text{ je: } f(x) = F'(x) = -\frac{2 \cdot (120-x)}{120^2} \cdot (-1) = \frac{120-x}{120 \cdot 60}$$

$$\text{inde je } f(x) = 0$$

Teraz môžeme vypočítať strednú hodnotu:

$$\begin{aligned} E(X) &= \int_{-\infty}^{\infty} x f(x) dx \quad \xrightarrow{\text{def. } E(X)} \int_0^{120} x f(x) dx \quad \xrightarrow{\text{inde je } f(x)=0} \int_0^{120} \frac{x(120-x)}{120 \cdot 60} dx = \frac{1}{120 \cdot 60} \int_0^{120} (120x - x^2) dx = \\ &= \frac{1}{120 \cdot 60} \left[60x^2 - \frac{x^3}{3} \right]_0^{120} = \boxed{40} \end{aligned}$$

Príklad 7 (pokračovanie príkladu 5):

7.14 Suppose we choose arbitrarily a point from the square with corners at (2,1), (3,1), (2,2), and (3,2). The random variable A is the area of the triangle with its corners at (2,1), (3,1), and the chosen point. (See also Exercise 5.9 and Figure 7.5.) Compute $E[A]$.

Zdroj: A Modern Introduction to Probability and Statistics. Understanding Why and How

Riešenie:

$$\begin{aligned} E(A) &= \int_{-\infty}^{\infty} x f(x) dx \quad \xrightarrow{\text{definícia } E(A)} \int_0^{\frac{1}{2}} x f(x) dx \quad \xrightarrow{\text{inde je } f(x)=0} \int_0^{\frac{1}{2}} x \cdot 2 dx = \left[x^2 \right]_0^{\frac{1}{2}} = \frac{1}{4} - 0 = \boxed{\frac{1}{4}} \end{aligned}$$

- **Vlastnosti** sú rovnaké ako pri diskretných náhodných premenných:

$$\mathbb{E}(aX) = a\mathbb{E}(X)$$

$$\mathbb{E}(X + Y) = \mathbb{E}(X) + \mathbb{E}(Y)$$

- **Stredná hodnota transformovanej náhodnej premennej:**

$$E[g(X)] = \int_{-\infty}^{\infty} g(x) f(x) dx$$

Nemusíme teda počítať hustotu transformovanej náhodnej premennej (ak to nie je pre výpočet jednoduchšie), ale spraviť napríklad:

$$E[X^2] = \int_{-\infty}^{\infty} x^2 f(x) dx$$

$$E[e^X] = \int_{-\infty}^{\infty} e^x f(x) dx$$

$$E\left[\frac{1}{X}\right] = \int_0^{\infty} \frac{1}{x} f(x) dx$$

Príklad 8: Vypočítame strednú hodnotu druhej čakania v príklade o stretnutí Romea a Júlie a porovnáme s priemerom zo simulácií.

```
priemerne_cakanie_2mocnina = np.mean(cakanie**2)
print(f"{priemerne_cakanie_2mocnina:.4f}")
```

2397.9991

$$E(X^2) = \int_{-\infty}^{\infty} x^2 f(x) dx = \int_0^{120} x^2 f(x) dx = \int_0^{120} \frac{x^2 (120-x)}{120 \cdot 60} dx =$$

$$= \frac{1}{120 \cdot 60} \int_0^{120} (120x^2 - x^3) dx = \frac{1}{120 \cdot 60} \left[40x^3 - \frac{x^4}{4} \right]_0^{120} = \boxed{2400}$$

7. Disperzia spojitej náhodnej premennej

- **Definícia disperzie** je rovnaká ako pre diskkrétne náhodné premenné (predpokladáme, že všetky potrebné stredné hodnoty existujú):

$$\mathbb{D}(X) = \mathbb{E}((X - \mathbb{E}(X))^2)$$

- Rovnako ako pre diskkrétne náhodné premenné, aj teraz platí (dôkazy využívali len horeuvedenú definíciu, nie to, že išlo o diskkrétne náhodné premenné):

$$\mathbb{D}(aX) = a^2 \mathbb{D}(X), \quad \mathbb{D}(a + X) = \mathbb{D}(X)$$

$$\mathbb{D}(X) = \mathbb{E}(X^2) - (\mathbb{E}(X))^2$$

- **Štandardná odchýlka** náhodnej premennej je odmocnina z jej disperzie

Príklad 9: Vypočítame disperziu čakania v príklade o stretnutí Romea a Júlie.

$$D(X) = E(X^2) - [E(X)]^2 = 2400 - 40^2 = 2400 - 1600 = \boxed{800}$$

- Pre nezávislé náhodné premenné platí to isté ako pre diskkrétne náhodné premenné.

Stredná hodnota a disperzia pre nezávislé náhodné premenné:

$$\mathbb{D}(X + Y) = \mathbb{D}(X) + \mathbb{D}(Y)$$

$$\mathbb{E}(XY) = \mathbb{E}(X) \mathbb{E}(Y)$$

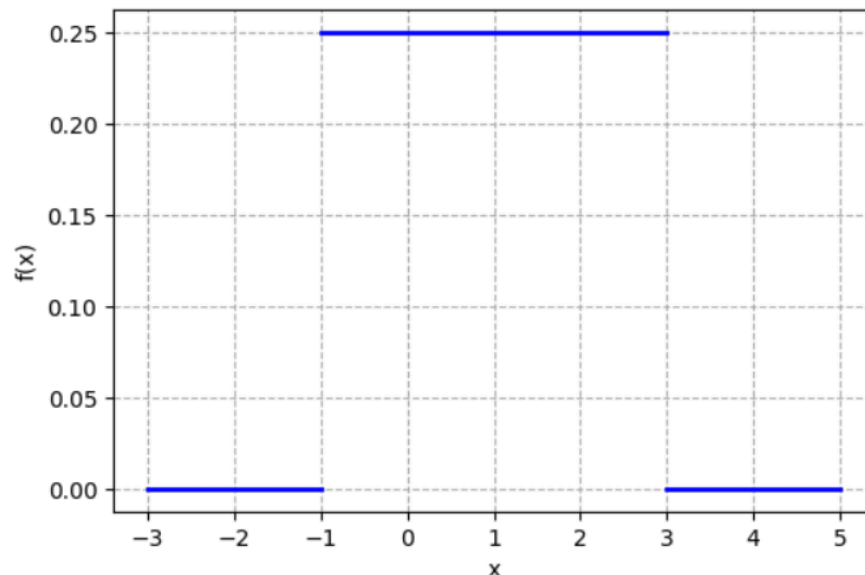
POZNÁMKA: Indukciou dostaneme, že tvrdenia platia aj pre n náhodných premenných.

8. Rovnomerné rozdelenie

- Parametre: $a < b$
- Hustota:

$$f(x) = \begin{cases} \frac{1}{b-a}, & a \leq x \leq b \\ 0, & \text{inak} \end{cases}$$

- Graf hustoty pre $a = -1, b = 3$:



- Stredná hodnota a disperzia:

$$E(X) = \frac{a+b}{2}, \quad D(X) = \frac{(b-a)^2}{12}$$

Odvodenie:

$$\begin{aligned} E(X) &= \int_{-\infty}^{\infty} x f(x) dx = \int_a^b x f(x) dx = \int_a^b x \cdot \frac{1}{b-a} dx = \frac{1}{b-a} \left[\frac{x^2}{2} \right]_a^b = \frac{b^2 - a^2}{2(b-a)} = \left[\frac{b+a}{2} \right] \\ E(X^2) &= \int_{-\infty}^{\infty} x^2 f(x) dx = \int_a^b x^2 \frac{1}{b-a} dx = \frac{1}{b-a} \left[\frac{x^3}{3} \right]_a^b = \frac{b^3 - a^3}{3(b-a)} = \frac{b^2 + ab + a^2}{3} \\ D(X) &= E(X^2) - [E(X)]^2 = \frac{b^2 + ab + a^2}{3} - \frac{(a+b)^2}{4} = \frac{1}{12} (4a^2 + 4ab + 4b^2 - 3a^2 - 6ab - 3b^2) = \\ &= \frac{1}{12} (a^2 - 2ab + b^2) = \left[\frac{(b-a)^2}{12} \right] \end{aligned}$$

- Vypočítame distribučnú funkciu:

$$\begin{aligned}
 & - \text{pre } x \leq a \text{ je } F(x) = 0 \\
 & - \text{pre } x \geq b \text{ je } F(x) = 1 \\
 & - \text{pre } x \in (a, b) \text{ je } F(x) = \int_{-\infty}^x f(t) dt = \int_a^x \frac{1}{b-a} dt = \frac{1}{b-a} [t]_a^x = \\
 & \quad = \frac{x-a}{b-a}
 \end{aligned}$$

- V Pythone: `scipy.stats.uniform`, parametre `loc`, `scale` dajú rovnomerné rozdelenie na intervale `[loc, loc + scale]`

<code>rvs(loc=0, scale=1, size=1, random_state=None)</code>	náhodné čísla (<i>random variates</i>)
<code>pdf(x, loc=0, scale=1)</code>	hustota (<i>probability density function</i>)
<code>cdf(x, loc=0, scale=1)</code>	distribučná funkcia (<i>cumulative distribution function</i>)
<code>ppf(q, loc=0, scale=1)</code>	inverzná funkcia k distribučnej (<i>percent point function</i>), percentily

Funkcie **rvs**, **pdf**, **cdf**, **ppf** (aj iné) sú univerzálne pre všetky rozdelenia.

scipy.stats - <https://docs.scipy.org/doc/scipy/reference/stats.html>

Príklad 10

The warranty on a machine specifies that it will be replaced at failure or age 4, whichever occurs first. The machine's age at failure, X , has density function

$$f(x) = \begin{cases} 1/5 & \text{for } 0 < x < 5, \\ 0 & \text{otherwise.} \end{cases}$$

Let Y be the age of the machine at the time of replacement. Determine the variance of Y .

- (A) 1.3 (B) 1.4 (C) 1.7 (D) 2.1 (E) 7.5

Riešenie:

$$Y = g(X) = \begin{cases} X & \text{pre } X \in [0, 4) \\ 4 & \text{pre } X \geq 4 \end{cases} \quad X \sim R(0, 5)$$

$$E(Y) = \int_{-\infty}^{\infty} g(x) f(x) dx = \int_0^5 g(x) f(x) dx = \int_0^5 g(x) \cdot \frac{1}{5} dx =$$

$$= \int_0^4 x \cdot \frac{1}{5} dx + \int_4^5 4 \cdot \frac{1}{5} dx = \frac{1}{5} \int_0^4 x dx + \frac{4}{5} \int_4^5 1 dx =$$

$$= \frac{1}{5} \left[\frac{x^2}{2} \right]_0^4 + \frac{4}{5} \underbrace{[x]_4^5}_{=1} = \frac{1}{5} \cdot 8 + \frac{4}{5} \cdot 1 = \frac{12}{5}$$

$$\text{Analogicky: } E(Y^2) = \int_0^4 x^2 \cdot \frac{1}{5} dx + \int_4^5 4^2 \cdot \frac{1}{5} dx = \frac{1}{5} \left[\frac{x^3}{3} \right]_0^4 + \frac{16}{5} [x]_4^5 =$$
$$= \frac{64}{15} + \frac{16}{5} = \frac{112}{15}$$

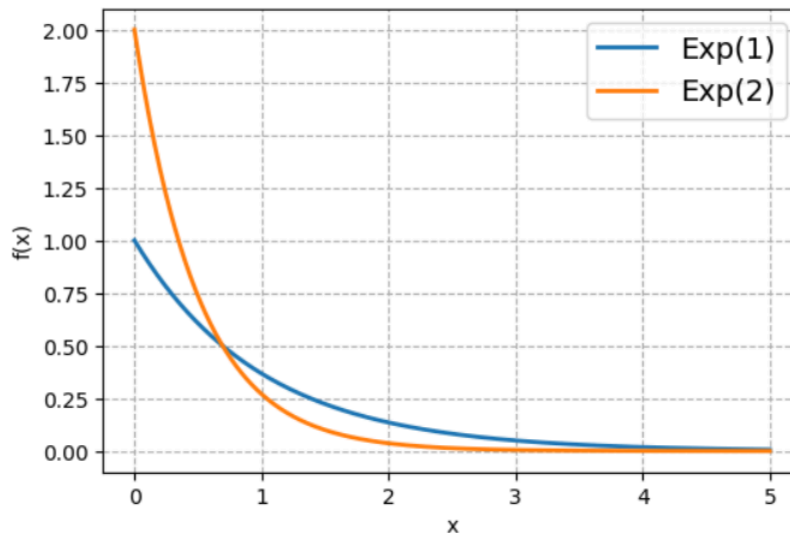
$$\Rightarrow D(Y) = E(Y^2) - [E(Y)]^2 = \frac{112}{15} - \left(\frac{12}{5} \right)^2 = \frac{112}{15} - \frac{144}{25} = \frac{128}{75}$$

9. Exponenciálne rozdelenie

- Parameter: $\lambda > 0$
- Hustota:

$$f(x) = \begin{cases} \lambda e^{-\lambda x}, & x \geq 0 \\ 0, & \text{inak} \end{cases}$$

- Graf hustoty:



- Stredná hodnota a disperzia:

$$E(X) = \frac{1}{\lambda}, \quad D(X) = \frac{1}{\lambda^2}$$

- Vypočítame distribučnú funkciu:

$$\begin{aligned} \text{pre } x > 0: \quad F(x) &= \int_0^x f(t) dt = \int_0^x \lambda e^{-\lambda t} dt = \lambda \int_0^x \frac{e^{-\lambda t}}{-\lambda} dt = \\ &= -[e^{-\lambda t}]_0^x = -(e^{-\lambda x} - 1) = 1 - e^{-\lambda x} \end{aligned}$$

Pre záporné a nulové x je distribučná funkcia nulová.

Všimnime si, že ide o distribučnú funkciu z príkladu 2. Máme teda napríklad jednoduchý postup, ako generovať náhodné čísla z exponenciálneho rozdelenia, ak vieme generovať náhodné čísla z rovnomerného rozdelenia.

- V Pythone: `scipy.stats.expon`, parameter `scale` dá exponenciálne rozdelenie s parametrom `scale = 1/lambda`

- Pri tejto parametrizácii sa teda parameter **scale** zhoduje so strednou hodnotou rozdelenia.

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

par_scale = 5
pocet_simulacii = 10**5

nahodne_cisla = expon.rvs(scale=par_scale, size=pocet_simulacii)
priemer = np.mean(nahodne_cisla)
print(f"Parameter scale: {par_scale}")
print(f"Priemerná hodnota: {priemer:.5f}")
```

Parameter scale: 5
Priemerná hodnota: 5.00235

- Momenty rozdelenia v Pythone, medián:

```
stredna_hodnota = expon.mean(scale=5)
disperzia = expon.var(scale=5)
standardna_odchylka = expon.std(scale=5) # dispersion = variance
median = expon.median(scale=5) # P(X <= median) = P(X >= median) = 1/2

print(f"Stredná hodnota: {stredna_hodnota}")
print(f"Disperzia: {disperzia}")
print(f"Štandardná odchýlka: {standardna_odchylka}")
print(f"Medián: {median:.5f}")
```

Stredná hodnota: 5.0
Disperzia: 25.0
Štandardná odchýlka: 5.0
Medián: 3.46574

- Presnú hodnotu mediánu nájdeme z distribučnej funkcie:

$$F(x) = \frac{1}{2}$$

$$1 - e^{-\lambda x} = \frac{1}{2}$$

$$e^{-\lambda x} = \frac{1}{2} \quad / \ln$$

$$-\lambda x = \ln\left(\frac{1}{2}\right)$$

$$-\lambda x = -\ln 2$$

$$x = \frac{1}{\lambda} \cdot \ln 2$$

Príklad 11:

An actuary is reviewing a study she performed on the size of the claims made ten years ago under homeowners insurance policies. In her study, she concluded that the size of claims followed an exponential distribution and that the probability that a claim would be less than \$1,000 was 0.250. The actuary feels that the conclusions she reached in her study are still valid today with one exception: every claim made today would be twice the size of a similar claim made ten years ago as a result of inflation. Calculate the probability that the size of a claim made today is less than \$1,000.

Riešenie:

$$X = \text{staré hodnoty} \sim \text{Exp}(\lambda), \quad E(X) =$$

$$X = \text{staré hodnoty} \sim \text{Exp}(\lambda)$$

$$P(X < 1000) = 0,250$$

$$F(1000) = \frac{1}{4}$$

$$1 - e^{-\lambda \cdot 1000} = \frac{1}{4}$$

$$e^{-1000\lambda} = \frac{3}{4}$$

$$-1000\lambda = \ln \frac{3}{4}$$

$$\lambda = -\frac{1}{1000} \ln \frac{3}{4} \quad (\text{>0, lebo } \ln \frac{3}{4} < 0)$$

$$Y = \text{dnešné hodnoty}$$

$$Y = 2X$$

$$P(Y < 1000) = P(2X < 1000) = P(X < 500) = F(500) = 1 - e^{-\lambda X} \quad \left| \begin{array}{l} X=500 \\ \lambda = -\frac{1}{1000} \ln \frac{3}{4} \end{array} \right.$$
$$= 1 - e^{\frac{1}{1000} (\ln \frac{3}{4}) \cdot 500} = 0,13397$$

10. Normálne rozdelenie $N(\mu, \sigma^2)$

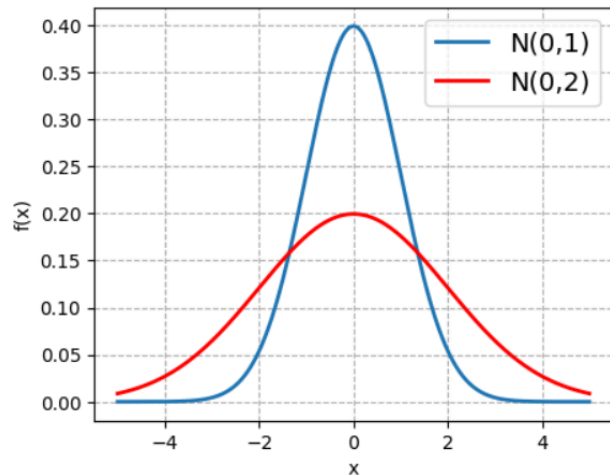
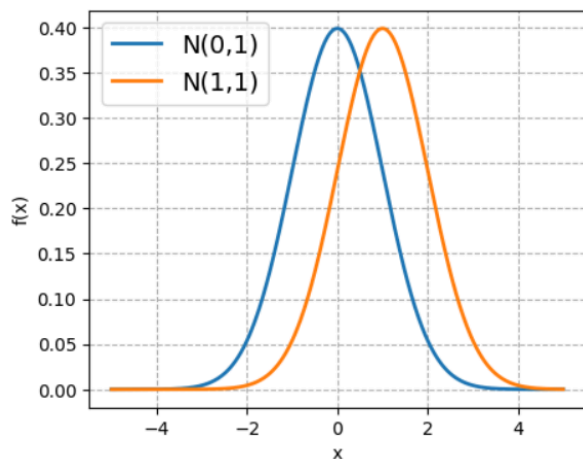
- Parametre: μ (ľubovoľné reálne), $\sigma^2 > 0$ zodpovedajú strednej hodnote a disperzii
- Hustota:

$$f(x) = \frac{1}{\sqrt{2\pi\sigma^2}} \exp\left(-\frac{(x-\mu)^2}{2\sigma^2}\right), \quad x \in \mathbb{R}$$

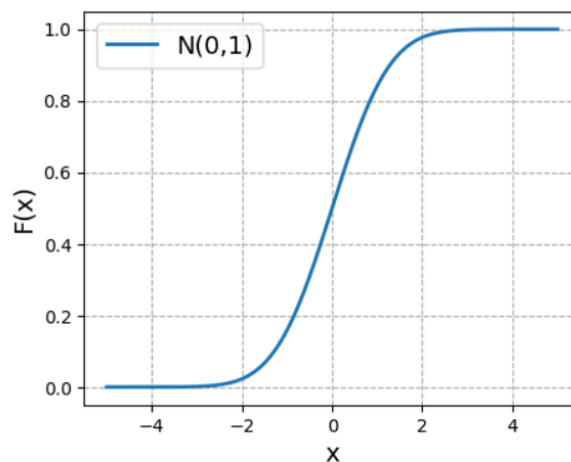
- Stredná hodnota a disperzia:

$$E(X) = \mu, \quad D(X) = \sigma^2$$

- Ak je stredná hodnota nulová a disperzia rovná jednej, hovoríme o **štandardizovanom normálnom rozdelení** alebo **normalizovanom normálnom rozdelení**
- Distribučná funkcia nemá jednoduchý predpis (integrál z hustoty sa nedá "pekné" vyjadriť), ale distribučná funkcia normalizovaného normálneho rozdelenia sa zvykne označovať ako Φ .
- Graf hustoty:



- Graf distribučnej funkcie normalizovaného normálneho rozdelenia:



- **Násobok normálneho rozdelenia má tiež normálne rozdelenie.** Jeho stredná hodnota a disperzia je daná nasledovne (takúto strednú hodnotu a disperziu má súčet pri ľubovoľnom rozdelení, normálne rozdelenie je zaujímavé tým, že sa zachová typ rozdelenia):

$$Y = aX, \quad X \sim N(\mu, \sigma^2)$$

$$E(Y) = E(aX) = aE(X) = a\mu$$

$$D(Y) = D(aX) = a^2 D(X) = a^2 \sigma^2 \quad Y \sim N(a\mu, a^2 \sigma^2)$$

- **Ak k normálnemu rozdeleniu pripočítame konštantu a , výsledná náhodná premenná má tiež normálne rozdelenie.** Jej stredná hodnota sa zväčší o a , disperzia sa nezmení (takúto strednú hodnotu a disperziu má súčet náhodnej premennej a konštanty pri ľubovoľnom rozdelení, normálne rozdelenie je zaujímavé tým, že sa zachová typ rozdelenia):

$$Y = a + X, \quad X \sim N(\mu, \sigma^2)$$

$$E(Y) = E(a + X) = E(a) + E(X) = a + E(X) = a + \mu$$

$$D(Y) = D(a + X) = D(X) = \sigma^2 \quad Y \sim N(a + \mu, \sigma^2)$$

- Dôsledok: **Normalizácia náhodnej premennej**

Ak $X \sim N(\mu, \sigma^2)$, tak $\frac{X - \mu}{\sigma}$ má normálne rozdelenie, pričom

$$E\left(\frac{X - \mu}{\sigma}\right) = \frac{1}{\sigma} E(X - \mu) = \frac{1}{\sigma} (E(X) - \mu) = 0$$

$$D\left(\frac{X - \mu}{\sigma}\right) = \frac{1}{\sigma^2} D(X - \mu) = \frac{1}{\sigma^2} D(X) = \frac{1}{\sigma^2} \cdot \sigma^2 = 1$$

$$\boxed{\frac{X - \mu}{\sigma} \sim N(0, 1)}$$

- **Súčet nezávislých normálnych náhodných premenných má tiež normálne rozdelenie.** Jeho stredná hodnota a disperzia je daná nasledovne (takúto strednú hodnotu a disperziu má súčet pri ľubovoľnom rozdelení, normálne rozdelenie je zaujímavé tým, že sa zachová typ rozdelenia):

$$X \sim N(\mu_1, \sigma_1^2), Y \sim N(\mu_2, \sigma_2^2) \text{ nezávislé}$$

$$E(X+Y) = E(X) + E(Y) = \mu_1 + \mu_2$$

$$D(X+Y) \stackrel{\text{nezávislost}}{=} D(X) + D(Y) = \sigma_1^2 + \sigma_2^2 \quad X+Y \sim N(\mu_1 + \mu_2, \sigma_1^2 + \sigma_2^2)$$

- Dôsledok: **Priemer nezávislých normálnych náhodných premenných má tiež normálne rozdelenie, a to s parametrami:**

$$X_1, \dots, X_n \text{ nezávislé}, \quad X_i \sim N(\mu_i, \sigma_i^2)$$

$$E(\bar{X}) = E\left(\frac{1}{n} \sum_{i=1}^n X_i\right) = \frac{1}{n} E\left(\sum X_i\right) = \frac{1}{n} \underbrace{\sum E(X_i)}_{=\mu_1 + \dots + \mu_n} = \frac{1}{n} (\mu_1 + \dots + \mu_n)$$

$$D(\bar{X}) = D\left(\frac{1}{n} \sum_{i=1}^n X_i\right) = \frac{1}{n^2} D\left(\sum X_i\right) \stackrel{\text{nezávislost}}{=} \frac{1}{n^2} \underbrace{\sum D(X_i)}_{\sigma_1^2 + \dots + \sigma_n^2} = \frac{1}{n^2} (\sigma_1^2 + \dots + \sigma_n^2)$$

- Dôležitý špeciálny príklad: **Priemer nezávislých náhodných premenných s rovnakým normálnym rozdelením má normálne rozdelenie s parametrami:**

V špeciálnom prípade, ak všetky X_i majú rozd. $N(\mu, \sigma^2)$, tak

$$\boxed{E(\bar{X})} = \frac{1}{n} (\underbrace{\mu + \dots + \mu}_{n \text{ členov}}) = \frac{1}{n} \cdot n\mu = \boxed{\mu}, \quad \boxed{D(\bar{X})} = \frac{1}{n^2} (\underbrace{\sigma^2 + \dots + \sigma^2}_{n \text{ členov}}) = \frac{1}{n^2} \cdot n\sigma^2 = \boxed{\frac{\sigma^2}{n}}$$

- V Pythone: `scipy.stats.norm`, parametre: `loc` je stredná hodnota a `scale` je štandardná odchýlka.

Príklad 12:

Claims filed under auto insurance policies follow a normal distribution with mean 19,400 and standard deviation 5,000. What is the probability that the average of 25 randomly selected claims exceeds 20,000 ?

- (A) 0.01 (B) 0.15 (C) 0.27 (D) 0.33 (E) 0.45

Riešenie:

- Treba doplniť predpoklad nezávislosti.
- Stredná hodnota priemeru je rovnaká priemer jednotlivých náhodných premenných.
- Pri disperzii je ľahšie si potrebnú hodnotu odvodiť, ako ju pamätať:

$$X_1, \dots, X_{25} \sim N(19400, 5000^2) \text{ nezávislé}$$
$$\Rightarrow \bar{X} \sim N(19400, \frac{5000^2}{25})$$

$$D(\bar{X}) = D\left(\frac{1}{n} \sum X_i\right) = \frac{1}{n^2} D\left(\sum X_i\right) \stackrel{\text{nez.}}{=} \frac{1}{n^2} \sum D(X_i) = \frac{1}{n^2} \cdot n \sigma^2 = \frac{\sigma^2}{n}$$

$$P(\bar{X} > 20000) = 1 - F(20000) \text{ kde } F \text{ je distribučná funkcia}$$

$$N(19400, \frac{5000^2}{25})$$

$$\sigma = \sqrt{\frac{5000^2}{25}} = \frac{5000}{5} = 1000$$

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

```
p_prekrozenia = 1 - norm.cdf(20000, loc=19400, scale=1000)
print(f"p_prekrozenia = {p_prekrozenia:.5f}")
```

```
p_prekrozenia = 0.27425
```

Ak požadujeme, aby bol výsledok zapísaný pomocou distribučnej funkcie normalizovaného normálneho rozdelenia Φ :

$$P(\bar{X} > 20000) = P\left(\frac{\bar{X} - 19400}{1000} > \frac{20000 - 19400}{1000}\right) =$$

stredná hodnota

~ N(0, 1)

$$= P(N(0, 1) > \frac{600}{1000}) = P(N(0, 1) > 0.6) =$$

$$= 1 - P(N(0, 1) \leq 0.6) = 1 - \Phi(0.6)$$

kde Φ je distribučná funkcia $N(0, 1)$

```
p_prekroenia = 1 - norm.cdf(0.6)
print(f"p_prekroenia = {p_prekroenia:.5f}")
```

```
p_prekroenia = 0.27425
```

Príklad 13

A machine dispenses cups of tea. The volumes of tea are Normally distributed with a standard deviation of 1.4 ml. What proportion of cups of tea contain a volume of tea within 2 ml of the mean?

Introduction to Probability, Statistics & R. Foundations for Data-Based Sciences

Z fakultnej siete je kniha dostupná na <https://link.springer.com/book/10.1007/978-3-031-37865-2>

Riešenie:

$$\begin{aligned}
 X &\sim N(\mu, 1.4^2) \\
 P(\mu - 2 \leq X \leq \mu + 2) &= P\left(\frac{(\mu - 2) - \mu}{1.4} \leq \frac{X - \mu}{1.4} \leq \frac{(\mu + 2) - \mu}{1.4}\right) = \\
 &= P\left(\frac{-2}{1.4} \leq N(0,1) \leq \frac{2}{1.4}\right) = \Phi\left(\frac{2}{1.4}\right) - \Phi\left(\frac{-2}{1.4}\right)
 \end{aligned}$$

Vidíme, že vieme odpovedať na otázku aj bez toho, aby sme poznali strednú hodnotu.

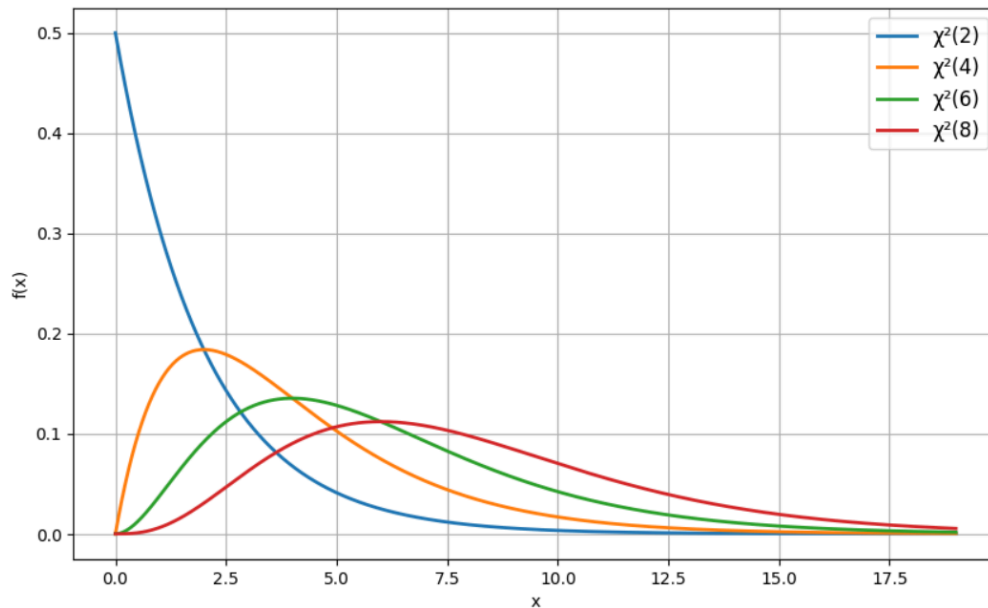
11. Chí kvadrát rozdelenie $\chi^2(k)$

- Parameter: prirodzené číslo k
- Definícia **chí-kvadrát rozdelenia s k stupňami voľnosti**:

$$X = Z_1^2 + Z_2^2 + \dots + Z_k^2$$

kde jednotlivé Z_i sú nezávislé náhodné premenné s normálnym rozdelením $N(0,1)$.

- Hustota má vo všeobecnosti zložitejší tvar, ktorý nebudeme potrebovať.
- Grafy hustoty:



- V Pythone: `scipy.stats.chi2`, parameter: `df` je počet stupňov voľnosti (degrees of freedom)

```
from scipy.stats import chi2
```

```
stredna_hodnota = chi2.mean(df = 5)
```

```
disperzia = chi2.var(df = 5)
```

```
# ekvivalentne:
```

```
# stredna_hodnota, disperzia = chi2.stats(df = 5, moments="mv")
```

```
print(f"Stredná hodnota: {stredna_hodnota}")
```

```
print(f"Disperzia: {disperzia}")
```

```
Stredná hodnota: 5.0
```

```
Disperzia: 10.0
```

- Vo všeobecnosti:

$$E(X) = k, \quad D(X) = 2k$$

Odvođenje strednej hodnoty a disperzie:

$$X = Z_1^2 + \dots + Z_k^2$$

$$E(X) = \underbrace{E(Z_1^2)}_{=1} + \dots + \underbrace{E(Z_k^2)}_{=1} = k$$

$$Z \sim N(0,1)$$

$$E(Z) = 0$$

$$D(Z) = 1$$

$$D(Z) = E(Z^2) - \underbrace{[E(Z)]^2}_{=0} = E(Z^2)$$

$$E(Z^2) = 1$$

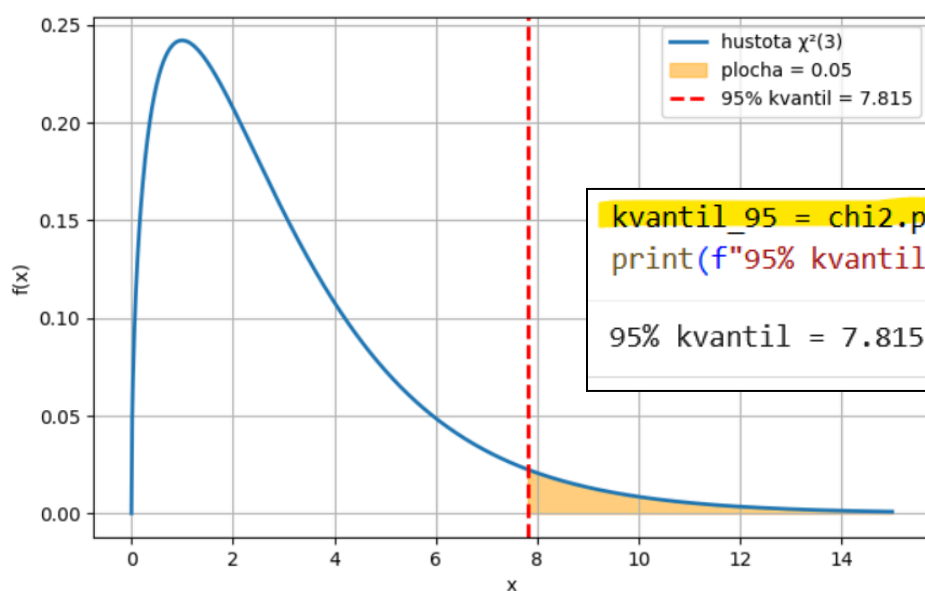
$$D(X) = D(Z_1^2 + \dots + Z_k^2) \stackrel{\text{nezávislost}}{=} \underbrace{D(Z_1^2)}_2 + \dots + \underbrace{D(Z_k^2)}_2 = 2k$$

Pre $Y \sim N(0, \sigma^2)$ platí $E(Y^4) = 3\sigma^4$ (nebudeme dokazovať, ale ešte sa na ňu to zide)

$$\Rightarrow D(Y^2) = E(Y^4) - [E(Y^2)]^2 = 3\sigma^4 - (\sigma^2)^2 = 2\sigma^4$$

↑
rovnako ako $E(Z^2)$ pre $Z \sim N(0,1)$

- Pri štatistických metódach budeme potrebovať kvantily, napríklad:



```
kvantil_95 = chi2.ppf(0.95, df=3)
print(f"95% kvantil = {kvantil_95:.3f}")
```

95% kvantil = 7.815