

STATISTIKA 2

Statistika a statistické zpracování dat

Blok 2-1

Jiří Fišer

22. listopadu 2024

1 Náhodná veličina

- ▶ Diskrétní náhodná veličina
- ▶ Spojitá náhodná veličina

2 Číselné charakteristiky náhodné veličiny

- ▶ polohy
- ▶ variability

Náhodná veličina

Definice (Náhodná veličina)

Náhodná veličina X je reálná funkce $X: \Omega \rightarrow \mathbb{R}$

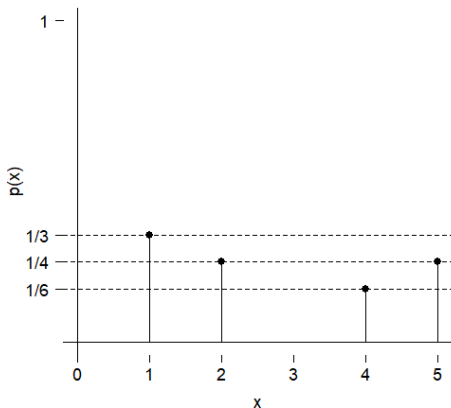
- $X, Y, Z \dots$ náhodná veličina
- $x, y, z \dots$ hodnota náhodné veličiny = **realizace náhodné veličiny**

Dělení náhodných veličin podle **oboru hodnot** $M \subset \mathbb{R}$:

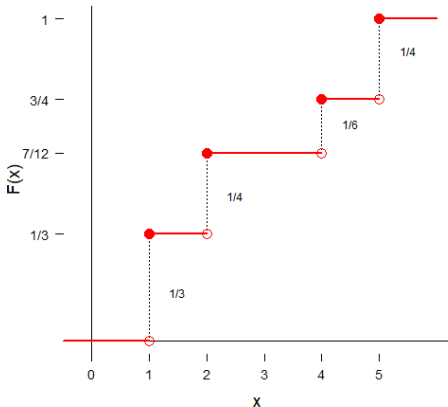
- **diskrétní:** M konečná nebo spočetná
- **spojitá:** M interval

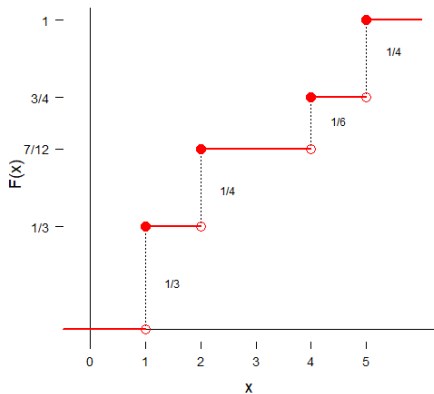
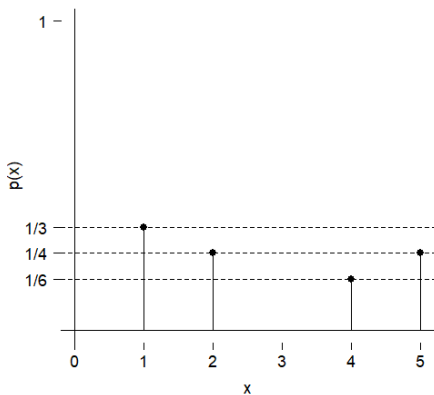
Diskrétní náhodná veličina

Pravděpodobnostní funkce



Distribuční funkce

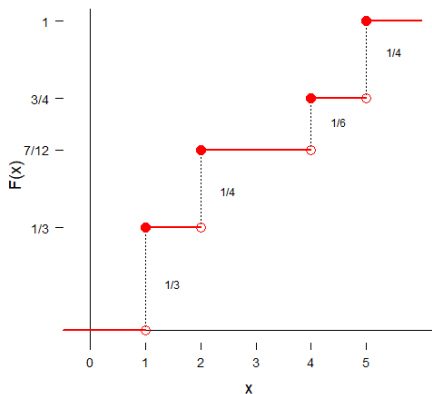
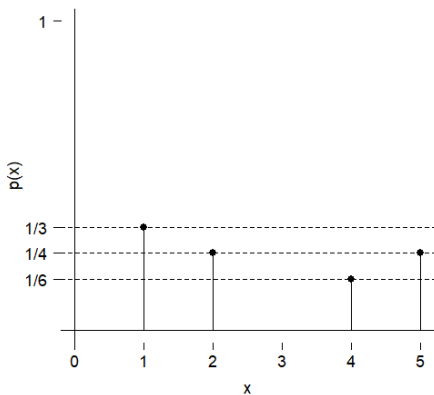




Pravděpodobnost, že X nabývá hodnoty **alespoň 3**

$$P(X \geq 3) = P(X = 4) + P(X = 5) = \frac{1}{6} + \frac{1}{4} = \frac{5}{12}$$

$$P(X \geq 3) = 1 - P(X < 3) = 1 - P(X \leq 2) = 1 - F(2) = 1 - \frac{7}{12} = \frac{5}{12}$$



Pravděpodobnost, že X nabývá hodnoty **nejvýše 4**

$$P(X \leq 4) = P(X = 1) + P(X = 2) + P(X = 4) = \frac{1}{3} + \frac{1}{4} + \frac{1}{6} = \frac{3}{4}$$

$$P(X \leq 4) = F(4) = \frac{3}{4}$$

Diskrétní vs. spojitá náhodná veličina

Diskrétní náhodná veličina

- značí „počet“
- počet autohavárií v daném časovém intervalu (Poissonovo rozdělení)
- počet 6, které padly v 10 hodech pravidelnou kostkou (Binomické rozdělení)

$$P(X = x_0) = p(x_0) \in \langle 0, 1 \rangle \quad \forall x_0 \in \mathbb{R}$$

Spojité náhodná veličina

- značí „měření“

$$P(X = x_0) = 0 \quad \forall x_0 \in \mathbb{R}$$

- jsme schopni najít pouze interval, kde se X realizuje

$$P(a \leq X \leq b) \in \langle 0, 1 \rangle \quad \forall a \leq b, a, b \in \mathbb{R}$$

Spojité náhodná veličina

Definice

Náhodná veličina X se nazývá (absolutně) **spojitá**, jestliže existuje nezáporná funkce $f : \mathbb{R} \rightarrow \mathbb{R}$ taková, že

$$F(x) = \int_{-\infty}^x f(t) dt, \quad \forall x \in \mathbb{R}.$$

Funkce $f(x)$ se nazývá **hustota** (rozdělení pravděpodobností) náhodné veličiny X .

Vlastnosti hustoty

$$f(x) \geq 0$$

$$\int_{-\infty}^{\infty} f(t) dt = 1$$

Vlastnosti hustoty a distribuční funkce spojité náhodné veličiny

$f(x) = F'(x)$ v každém bodě x , kde je F diferencovatelná

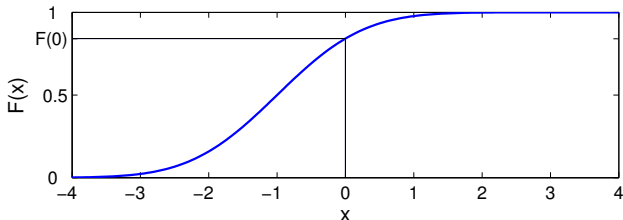
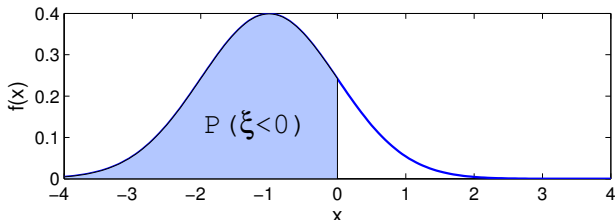
$$P(a \leq X \leq b) = F(b) - F(a) = \int_a^b f(t) dt$$

$$P(a \leq X \leq b) = P(a < X \leq b) = P(a \leq X < b) = P(a < X < b)$$

$$P(X \in B) = \int_B f(t) dt$$

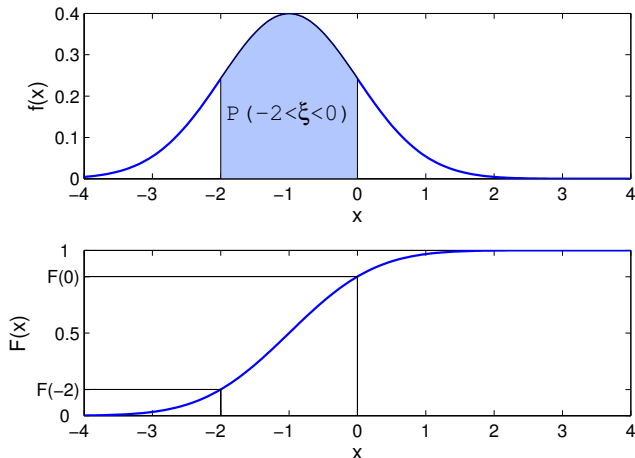
Výpočet pravděpodobností pomocí $F(x)$ a $f(x)$

$$P(\xi < 0) = F(0) = \int_{-\infty}^0 f(t) dt$$

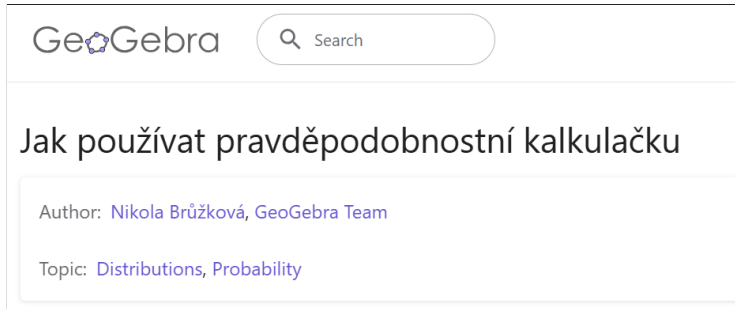


Výpočet pravděpodobností pomocí $F(x)$ a $f(x)$

$$P(-2 < \xi < 0) = F(0) - F(-2) = \int_{-2}^0 f(t) dt$$



Interaktivní práce s hustotou a distribuční funkcí (GeoGebra)



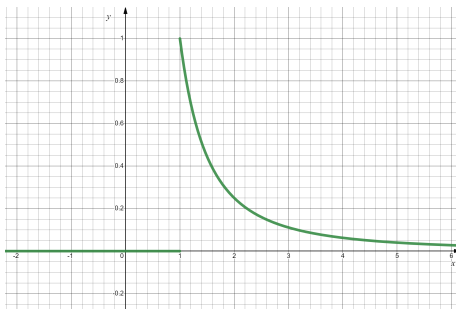
<https://www.geogebra.org/m/zyd8gta9>

Příklad

Spojité náhodná veličina X má rozdělení dané hustotou

$$f(x) = \begin{cases} 0 & x \leq 1 \\ \frac{1}{x^2} & x > 1 \end{cases}$$

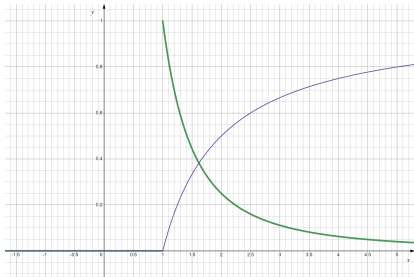
- 1 Určete její distribuční funkci.
- 2 Spočítejte pravděpodobnost $P(1 \leq X \leq 5)$.



$$f(x) = \begin{cases} 0 & x \leq 1 \\ \frac{1}{x^2} & x > 1 \end{cases}$$

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

$$F(x) = \begin{cases} 0 & x \leq 1 \\ 1 - \frac{1}{x} & x > 1 \end{cases}$$



$$F(x) = \int_{-\infty}^x 0 dt = 0 \quad \text{pro } x \leq 1$$

$$F(x) = \int_{-\infty}^x f(t) dt = \int_{-\infty}^1 0 dt + \int_1^x \frac{1}{t^2} dt = \left[-\frac{1}{t} \right]_1^x = -\frac{1}{x} + 1 \quad \text{pro } x > 1$$

$$P(1 \leq X \leq 5) = \int_1^5 \frac{1}{x^2} dx = \left[-\frac{1}{x} \right]_1^5 = -\frac{1}{5} + 1 = \frac{4}{5}$$

$$P(1 \leq X \leq 5) = F(5) - F(1) = \left(1 - \frac{1}{5} \right) - 0 = \frac{4}{5}$$

2.3 Číselné charakteristiky

Náhodnou veličinu X jednoznačně určují a plně popisují

- X diskrétní
 - ▶ pravděpodobnostní funkce
 - ▶ distribuční funkce
- X spojitá
 - ▶ hustota
 - ▶ distribuční funkce

číselné charakteristiky = shrnutí informací o X do několika čísel, které ji dostatečně charakterizují

Typy číselných charakteristik

Členění dle popisované vlastnosti

- **polohy:** „střed“, kolem kterého jsou koncentrovány hodnoty X
 - ▶ střední hodnota
 - ▶ modus
 - ▶ kvantily
- **variability:** rozptýlenost hodnot X kolem charakteristiky polohy
 - ▶ rozptyl
 - ▶ směrodatná odchylka
- **šikmost:** tvar rozdělení pravděpodobností (symetrické nebo asymetrické)
- **špičatost:** tvar rozdělení pravděpodobností (špičaté nebo zploštělé)

Charakteristiky polohy

- „střed“ skupiny údajů, kolem kterého všechny hodnoty kolísají
- chceme-li charakterizovat sledovanou veličinu jediným číslem, pak to bude nějaká charakteristika polohy
 - ▶ dospělá liška obecná je obvykle velká přibližně 70 cm
 - ▶ studenti prvního ročníku VŠ mají kolem 20 let věku
- jaké by měly mít **vlastnosti**
 - ▶ př. mzda ve firmě
 - ★ všichni zaměstnanci dostanou přidáno 1000 Kč $\Rightarrow$ „střed“ se zvětší o 1000 Kč

$$f(x_1 + c, x_2 + c, \dots, x_n + c) = f(x_1, x_2, \dots, x_n) + c$$

- ★ přepočteme-li jejich mzdu na euro $\Rightarrow$ „střed“ v eurech bude „střed“ v Kč krát 0,042

$$f(x_1 \cdot b, x_2 \cdot b, \dots, x_n \cdot b) = f(x_1, x_2, \dots, x_n) \cdot b$$

Střední hodnota $E(X)$

- základní charakteristika polohy
- „střed“ (těžiště), kolem kterého jsou koncentrovány hodnoty X

X diskrétní s oborem hodnot M

$$E(X) = \sum_{x_j \in M} x_j P(X = x_j)$$

X spojitá s hustotou $f(x)$

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

Jestliže řada nebo integrál absolutně nekonverguje, střední hodnota neexistuje.

Vlastnosti střední hodnoty

- $E(c) = c$ pro libovolnou konstantu $c \in \mathbb{R}$
- zvětšíme-li všechny hodnoty o konstantu c , zvětší se střední hodnota též o c

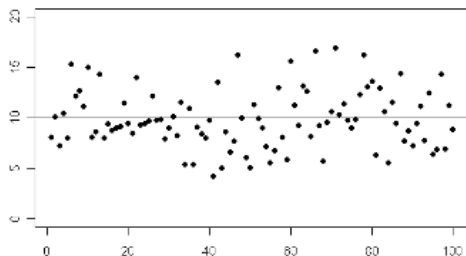
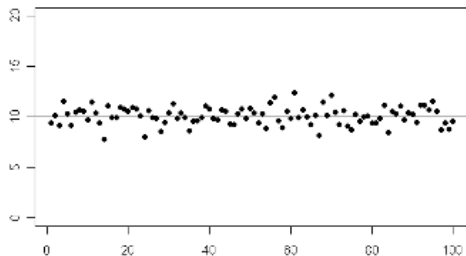
$$E(c + X) = c + E(X)$$

- násobíme-li všechny hodnoty nějakou konstantou b , pak nová střední hodnota bude rovna původní střední hodnotě krát konstanta b

$$E(bX) = bE(X)$$

Charakteristiky variability

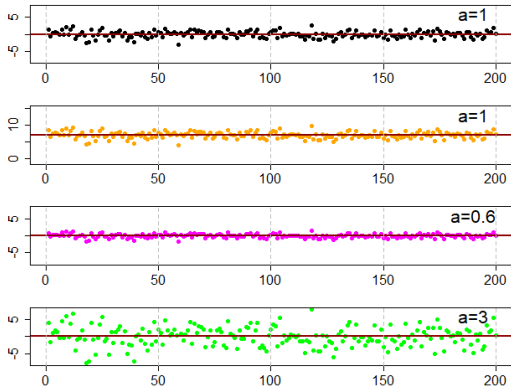
- číslo udávající **koncentraci (rozptýlení)** hodnot okolo středu



Charakteristiky variability

Vlastnosti:

- po **přičtení konstanty** ke každé hodnotě se variabilita **nezmění**
- při **násobení číslem** (v absolutní hodnotě) **mezi 0 a 1** se variabilita **zmenší**
- při **násobení číslem** (v absolutní hodnotě) **větším než 1** se variabilita **zvětší**



Rozptyl $D(X)$

- základní charakteristika variability
- charakterizuje měřítko (šířku) rozdělení
- popisuje rozptýlenost hodnot kolem střední hodnoty

Rozptyl

$$D(X) = E[X - E(X)]^2, \text{ existuje-li } E(X)$$

Směrodatná odchylka

$$\sqrt{D(X)}$$

Vlastnosti rozptylu

1 $D(X) \geq 0$

2 $D(X) = E(X^2) - [E(X)]^2$

$$D(X) = \begin{cases} \sum_{x_i \in M} x_i^2 P(X = x_i) - [E(X)]^2 & X \text{ diskrétní} \\ \int_{-\infty}^{\infty} x^2 f(x) dx - [E(X)]^2 & X \text{ spojitá} \end{cases}$$

3 $D(a + cX) = c^2 D(X)$ pro libovolné konstanty $a, c \in \mathbb{R}$

4 $D(X) = 0 \Leftrightarrow P(X = c) = 1$

Modus $\hat{x}$

- charakteristika polohy
- X diskrétní: nejpravděpodobnější hodnota
- X spojitá: bod x , ve kterém hustota pravděpodobností $f(x)$ nabývá lokálního maxima

Příklad: diskrétní náhodná veličina

Příklad

Diskrétní náhodná veličina X

- *nabývá hodnot $M = \{1, 2, 4, 5\}$
s pravděpodobnostmi $p(k) = P[X = k]$, kde*
- *$p(1) = \frac{1}{3}$, $p(2) = \frac{1}{4}$, $p(4) = \frac{1}{6}$, $p(5) = \frac{1}{4}$ a $p(x) = 0$ jinak.*

Spočítejte střední hodnotu, rozptyl, směrodatnou odchylku a modus náhodné veličiny X .

Nějaký nápad?

Rozdělení pravděpodobností náhodné veličiny X

k	1	2	4	5
$P(X = k)$	$\frac{1}{3}$	$\frac{1}{4}$	$\frac{1}{6}$	$\frac{1}{4}$

Střední hodnota $E(X) = \sum_n x_n p(x_n)$

- $E(X) = 1 \cdot \frac{1}{3} + 2 \cdot \frac{1}{4} + 4 \cdot \frac{1}{6} + 5 \cdot \frac{1}{4} = \frac{11}{4} = 2,75$

Rozptyl a směrodatná odchylka

$$D(X) = E[X - E(X)]^2 = E(X^2) - [E(X)]^2$$

$$E(X^2) = \sum_n x_n^2 p(x_n) = 1^2 \cdot \frac{1}{3} + 2^2 \cdot \frac{1}{4} + 4^2 \cdot \frac{1}{6} + 5^2 \cdot \frac{1}{4} = \frac{41}{4} = 10,25$$

- $D(X) = E(X^2) - [E(X)]^2 = \frac{41}{4} - \left(\frac{11}{4}\right)^2 = \frac{43}{16} \doteq 2,69$

- $\sqrt{D(X)} = \sqrt{\frac{43}{16}} \doteq 1,64$

Modus: nejpravděpodobnější hodnota: $\hat{x} = 1$

Příklad: spojitá náhodná veličina

Příklad

Hustota pravděpodobnosti náhodné veličiny X má tvar

$$f(x) = \begin{cases} 6x(1-x) & 0 < x \leq 1 \\ 0 & \text{jinak} \end{cases}$$

Určete střední hodnotu, rozptyl a modus náhodné veličiny X .

Nějaký nápad?

$$f(x) = \begin{cases} 6x(1-x) & 0 < x \leq 1 \\ 0 & \text{jinak} \end{cases}$$

Střední hodnota:

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

$$\begin{aligned} E(X) &= \int_{-\infty}^{\infty} xf(x) dx = \int_0^1 x \cdot 6x(1-x) dx = \\ &= \int_0^1 (6x^2 - 6x^3) dx = \left[\frac{6x^3}{3} - \frac{6x^4}{4} \right]_0^1 = \frac{6}{3} - \frac{6}{4} = \frac{24-18}{12} \\ &= \frac{6}{12} = \frac{1}{2} \end{aligned}$$

$$f(x) = \begin{cases} 6x(1-x) & 0 < x \leq 1 \\ 0 & \text{jinak} \end{cases}, \quad E(X) = \frac{1}{2}$$

Rozptyl:

$$D(X) = E[X - E(X)]^2 = E(X^2) - [E(X)]^2$$

$$\begin{aligned} E(X^2) &= \int_{-\infty}^{\infty} x^2 \cdot f(x) dx = \int_0^1 x^2 \cdot 6x(1-x) dx = \\ &= \int_0^1 6x^3 - 6x^4 dx = \left[\frac{6x^4}{4} - \frac{6x^5}{5} \right]_0^1 = \frac{6}{4} - \frac{6}{5} = \frac{30 - 24}{20} = \frac{6}{20} \end{aligned}$$

$$D(X) = \frac{6}{20} - \left(\frac{1}{2}\right)^2 = \frac{6}{20} - \frac{1}{4} = \frac{6 - 5}{20} = \frac{1}{20}$$

$$f(x) = \begin{cases} 6x(1-x) & 0 < x \leq 1 \\ 0 & \text{jinak} \end{cases}$$

Modus: bod, v němž je lokální maximum hustoty

$$f'(x) = 6 - 12x$$

$$6 - 12x = 0 \quad \Leftrightarrow \quad x = 0,5$$

$$f''(x) = -12 < 0 \quad \Rightarrow \quad \text{maximum} \quad \Rightarrow \quad \hat{x} = 0,5$$

Kvantily spojité náhodné veličiny

Definice

Nechť $\alpha \in (0, 1)$. α -kvantilem spojité náhodné veličiny X rozumíme kterékoli reálné číslo x_α , které splňuje

$$P(X \leq x_\alpha) = \alpha.$$

- X spojitá: α -kvantil určen jednoznačně vztahem

$$F(x_\alpha) = \alpha$$

- X diskrétní: kvantily nejsou určeny jednoznačně, nebudeme uvažovat

Speciální názvy kvantilů

- $x_{0,5}$ – medián
- $x_{0,25}$ – dolní kvartil
- $x_{0,75}$ – horní kvartil
- $x_{k/10}$, $k = 1, \dots, 9$ – k -tý decil
- $x_{k/100}$, $k = 1, \dots, 99$ – k -tý percentil

Příklad výpočtu kvantilů

Příklad

Vypočtěte první decil a horní kvartil náhodné veličiny X určené hustotou

$$f(x) = \begin{cases} \frac{1}{2} & \text{pro } x \in \langle 0, 2 \rangle \\ 0 & \text{jinak.} \end{cases}$$

- $F(x) = \begin{cases} 0 & \text{pro } x \leq 0 \\ \frac{1}{2}x & \text{pro } x \in (0; 2) \\ 1 & \text{pro } x \geq 2 \end{cases}$
- $F(x) = 0,1, \quad \frac{1}{2}x = 0,1, \quad x = 0,2, \quad \text{tedy } x_{0,1} = 0,2$
- $F(x) = 0,75, \quad \frac{1}{2}x = 0,75, \quad x = 1,5, \quad \text{tedy } x_{0,75} = 1,5$

Příklad výpočtu kvantilů

Příklad

Vypočtěte první decil a horní kvartil náhodné veličiny X určené hustotou

$$f(x) = \begin{cases} \frac{1}{2} & \text{pro } x \in \langle 0, 2 \rangle \\ 0 & \text{jinak.} \end{cases}$$

- $F(x) = \begin{cases} 0 & \text{pro } x \leq 0 \\ \frac{1}{2}x & \text{pro } x \in (0; 2) \\ 1 & \text{pro } x \geq 2 \end{cases}$
- $F(x) = 0,1, \quad \frac{1}{2}x = 0,1, \quad x = 0,2, \quad \text{tedy } x_{0,1} = 0,2$
- $F(x) = 0,75, \quad \frac{1}{2}x = 0,75, \quad x = 1,5, \quad \text{tedy } x_{0,75} = 1,5$