Přednáška 3 – Spojitá náhodná veličina, rozdělení

$x \in (-\infty, \infty)$ nebo $x \in (a, b)$ – nespočetný počet možných realizací (intervaly)

Příklady:

rychlost $\in (0, 280)$ – spojitá náhodná veličina

realizace: 22.5, 56, 138.563, 88.16, 110.32223, ...

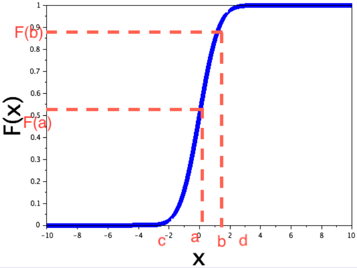
spotřeba paliva, teplota, tlak, váha, délka, výška, cena, ...

- pravděpodobnosti ~~jednotlivých realizací~~ intervalů realizací

Distribuční funkce (DF)

$F(x) = P(x \leq i)$

- neklesající
- Pro $x \in (c, d)$
  - $F(x) = 0$ pro $x < c$,
  - $F(x) = 1$ pro $x \geq d$



Vlastnost DF: $P(x \in (a, b)) = F(b) - F(a)$

Hustota pravděpodobnosti (hp) spojité náhodné veličiny

rozdělení Sp.N.V. – hp, rozdělení D.N.V. - pf

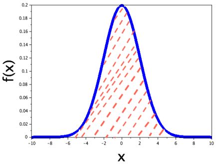
$$f(x) = \frac{\partial F(x)}{\partial x} \quad \text{nebo} \quad F(x) = \int_{-\infty}^{x} f(t)\, dt, \qquad t - \text{čas}$$

Vlastnosti hp:

- Nezáporná na celém definičním oboru $f(x) \geq 0$, $x \in (-\infty, \infty)$
- Integral hp přes celý definiční obor se rovná 1

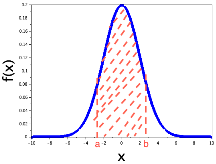
Příklad: Gaussovo (normální) rozdělení:

$$\underbrace{\int_{-\infty}^{\infty} f(x)\, dx = 1}_{\text{plocha pod křivkou hp } = 1}$$



- Pravděpodobnost, že $x \in (a, b)$

$$\int_{a}^{b} f(x)\, dx = P(x \in (a, b))$$



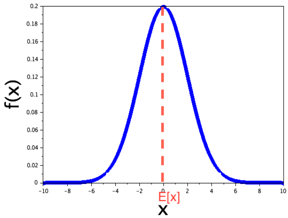
Střední hodnota, rozptyl, modus

Střední hodnota

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

Modus:

$\hat{x} = \arg\max(f(x))$

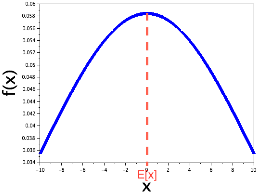


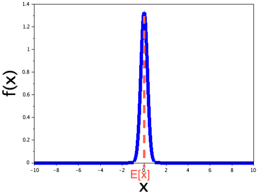
Rozptyl:

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

Odvození:

$$\begin{aligned} D[x] &= E\left[(x - E[x])^2\right] \\ &= E\left[x^2 - 2xE[x] + (E[x])^2\right] \\ &= E[x^2] - 2E[x]E[x] + (E[x])^2 \\ &= E[x^2] - (E[x])^2 \end{aligned}$$

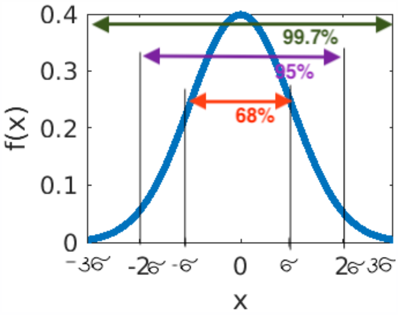




Směrodatná odchylka: $\sigma = \sqrt{D[x]}$

Pravidlo 3σ :

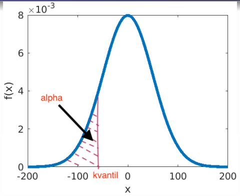
68% dat $\in$ (střední hodnota $\pm\ \sigma$)
95% dat $\in$ (střední hodnota $\pm\ 2\sigma$)
99.7% dat $\in$ (střední hodnota $\pm\ 3\sigma$)



Kvantily

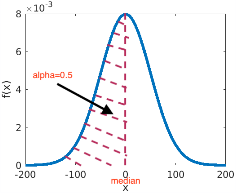
Kvantil $$\int_{-\infty}^{\xi_\alpha} f(x)\,dx = \alpha$$

ξ_α – kvantil, α – plocha pod křivkou



Medián $$\int_{-\infty}^{\tilde{x}_{0.5}} f(x)\,dx = 0.5$$

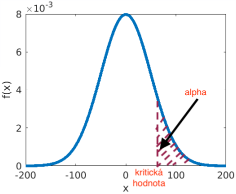
$$\alpha = 0.5$$

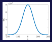


Kritická hodnota:

$$\int_{z_\alpha}^{\infty} f(x)\,dx = \alpha$$

z_α – kritická hodnota,
α – plocha pod křivkou



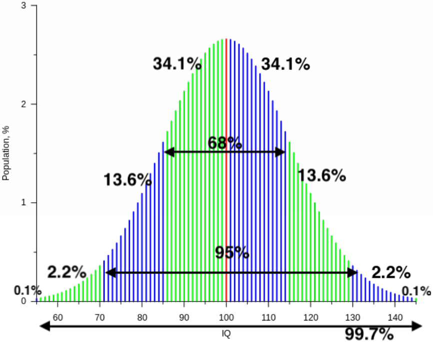
Gaussovo (normální) rozdělení – 

$$f(x) = \frac{1}{\sqrt{2\pi\sigma^2}} \exp\{-\frac{1}{2\underbrace{\sigma^2}_{D[x]}}(x-\underbrace{\mu}_{E[x]})^2\}$$

μ – střední hodnota
σ^2 – rozptyl
$x \sim N(\mu, \sigma^2)$

Příklad: Inteligenční kvocient

$$\mathrm{IQ} \sim N(100, 15^2)$$



Zdroj: Wikipedia, under the Creative Commons Attribution-Share Alike 3.0 Unported license

Normované Gaussovo rozdělení

$$z = \frac{x-\mu}{\sigma}, \qquad z \sim N(0,1)$$

$$f(z) = \frac{1}{\sqrt{2\pi}} e^{-\frac{1}{2}z^2}$$

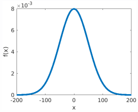
Příklady:

- jedno z nejvýznamnějších rozdělení N.V.
- při velkém počtu dat aproximuje jiná rozdělení

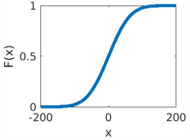
Gaussovo (normální) rozdělení – hp a DF

Normální rozdělení:

Hustota pravděpodobnosti

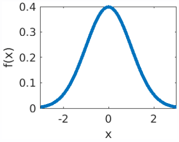


Distribuční funkce

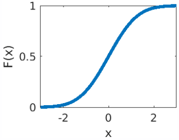


Normované normální rozdělení:

Hustota pravděpodobnosti



Distribuční funkce



Poznámka

- DF normálního rozdělení se počítá numericky
  $F(x) = \int_{-\infty}^{x} \frac{1}{\sqrt{2\pi\sigma^2}} \exp\{-\frac{1}{2\sigma^2}(t-\mu)^2\}\, dt$
- používá se pro výpočet kvantilů pro testy hypotéz
- tabulky DF pro normovanou N.V. + statistický software

Rovnoměrné rozdělení – stejná pravděpodobnost na (a,b)

$$f(x) = \begin{cases} \frac{1}{b-a} & pro \quad x \in (a,b) \\ 0 & pro \quad x \notin (a,b) \end{cases}$$

Příklad: doba čekání na autobus s intervalem 2 minuty

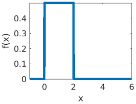
$f(x) = \frac{1}{2} \quad pro \quad x \in (0,2), \quad$ jinde $\; 0$

$$F(x) = \begin{cases} 0 & x < 0 \\ \frac{1}{2}x & 0 \leq x < 2 \\ 1 & x \geq 2 \end{cases} \quad - \quad \text{protože} \quad F(x) = \int_0^x \frac{1}{2}\,dt = \left[\frac{1}{2}t\right]_0^x = \frac{1}{2}x$$

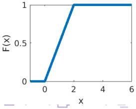
S jakou pravděpodobností budeme čekat déle, než 1 minutu?

$$P(x \in (1,2)) = F(2) - F(1) = 1 - \frac{1}{2} = \frac{1}{2}$$

Hustota pravděpodobnosti



Distribuční funkce



Střední hodnota

$$E[x] = \int_0^2 x f(x) dx = \int_0^2 \frac{1}{2} x \, dx = \frac{1}{2} \left[\frac{x^2}{2} \right]_0^2 = \frac{1}{2} \cdot \frac{4}{2} = 1$$

Rozptyl

$$D[x] = \int_0^2 x^2 f(x) dx - 1^2 = \int_0^2 \frac{1}{2} x^2 \, dx - 1 = \frac{1}{2} \left[\frac{x^3}{3} \right]_0^2 - 1 = \frac{1}{2} \cdot \frac{8}{3} - 1 = 0.33$$

5%ní kvantil

$$\int_0^{\xi_{0.05}} \frac{1}{2} \, dx = 0.05,$$

$$\frac{1}{2} [x]_0^{\xi_{0.05}} = 0.05$$

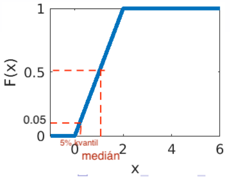
$$\frac{1}{2} \xi_{0.05} = 0.05$$

$$\xi_{0.05} = 0.1$$

Medián

$$\int_0^{\tilde{x}_{0.5}} \frac{1}{2} \, dx = 0.5,$$

$$\frac{1}{2} \tilde{x}_{0.5} = 0.5, \quad \tilde{x}_{0.5} = 1$$



Exponenciální rozdělení

vyjadřuje dobu mezi náhodně se vyskytujícími událostmi

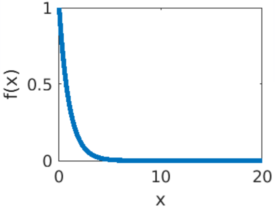
$$f(x) = \lambda e^{-\lambda x}, \quad \lambda > 0, x \geq 0$$

pouze nezáporná data

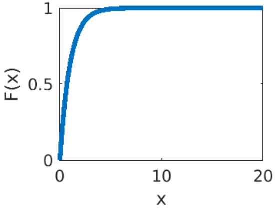
$$F(x) = \begin{cases} 0 & x < 0 \\ 1 - e^{-\lambda x} & x \geq 0, \end{cases}$$

Příklad: doba do vzniku zemětřesení,
doba, po kterou vydrží baterie v autě,
doba fungování přístroje, než se porouchá,
doba do výskytu nehody, onemocnění atd.

Hustota pravděpodobnosti



Distribuční funkce

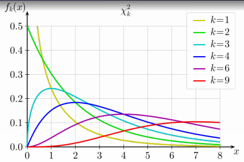


chí-kvadrát, Studentovo a Fisherovo rozdělení

chí-kvadrát nebo χ^2-rozdělení

$$\chi^2(n) = \sum_{i=1}^{n} (N_i(0,1))^2$$

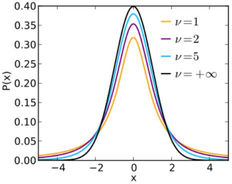
popis rozptylu (nezáporná data)



Zdroj: Wikipedia, under the Creative Commons Attribution 3.0 Unported license

Studentovo rozdělení

$$t(n) = \frac{N(0,1)}{\chi^2(n)/n}$$

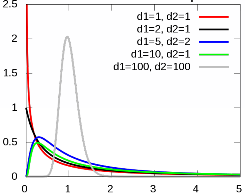


Zdroj: Wikipedia, under the Creative Commons Attribution 3.0 Unported license

Fisherovo rozdělení

$$F(n_1, n_2) = \frac{\chi^2(n_1)}{n_1} \Big/ \frac{\chi^2(n_2)}{n_2}$$

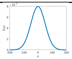
podíl rozptylů (nezáporná data)



Zdroj: Wikipedia, under the Creative Commons Attribution 3.0 Unported license

Poznámky: • základ – N(0,1) • používají se v testech hypotéz

Jak poznám rozdělení spojité náhodné veličiny?

Gaussovo (normální) rozdělení	křivka – symetrický kopeček 
Rovnoměrné rozdělení	stejná pravděpodobnost na intervalu
Normované normální rozdělení:	gaussovka $N(0, 1)$
Exponenciální rozdělení	doba mezi náhodně se vyskytujícími událostmi (nezáporná data)
chí-kvadrát, Studentovo a Fisherovo rozdělení	používají se v testech hypotéz