4 Diskrétní a spojité náhodné veličiny - OSNOVA

4.1 Základy pravděpodobnosti

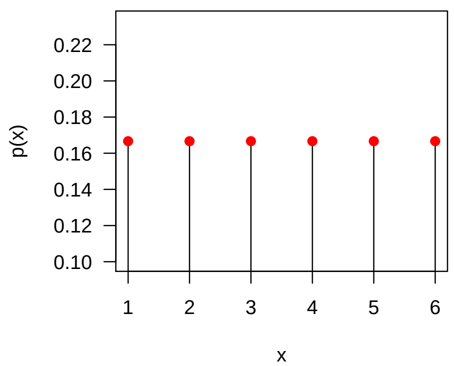
- Motivace:
  - reálná situace (data) $\rightarrow$ popíšeme ji nějakým známým rozdělením (binomické, poissonovo, normální) $\rightarrow$ z dat odhadneme parametry rozdělení $\rightarrow$ stanovíme nové závěry na základě vlastností rozdělení
- *Experiment* $\rightarrow$ založen na *náhodném pokusu*
  - porodní hmotnost: náhodný pokus = zvážení 1 novorozence;
  - vzdělání matky: náhodný pokus = dotaz na jednu matku;
  - číslo na kostce: náhodný pokus = hod kostkou;
- *Základní prostor* Ω = množina všech možných výsledků
  - porodní hmotnost: $0 - \infty$; $0 - 6\,000\,\mathrm{g}$;
  - počet starších sourozenců: 0, 1, 2, 3, 4, 5, 6, 7, 8, 9 a více;
  - kostka . . . 1–6;
- *Jev* = výsledek náhodného pokusu
  - Hodila jsem kostkou $\rightarrow$ nastal *jev*: (a) padla 5; (b) padlo liché číslo; (c) padlo číslo ≤ 2
  - Zvážila jsem novorozence $\rightarrow$ nastal *jev*: (a) vážil $2\,654\,\mathrm{g}$; (b) vážil více než $2\,500\,\mathrm{g}$, apod.

Pravděpodobnost

- vyjadřuje, jak velká je naděje, že nějaký jev nastane
- $\Pr(A) = \Pr(\text{nastal jev A})$
- $\Pr(A) \in \langle 0; 1\rangle$; resp. $\langle 0\,\% - 100\,\%\rangle$
- příklad: hodím kostkou:
  - $\Pr(\text{padne } 1) = 1/6 \ldots 16.7\,\%$
  - $\Pr(\text{padne liché číslo}) = 1/2 \ldots 50\,\%$
  - $\Pr(\text{padne } 3{,}4{,}5{,}6) = 2/3 \ldots 66.67\,\%$

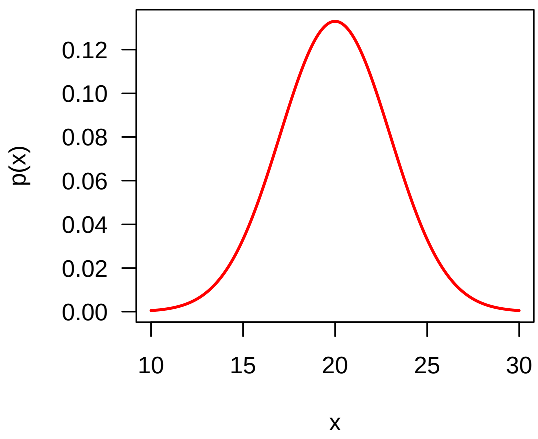
4.2 Náhodné veličiny

- Víc než výsledek nás často zajímají jeho číselné interpretace
- *Náhodná veličina* X = pravidlo, které zobrazuje základní prostor možných výsledků do množiny reálních čísel
- i-tá realizace náh. veličiny X se značí x_i
  (a) X …počet puntíků na vrchní straně kostky: $x_1 = 4$, $x_2 = 1$ …
  (b) Y …dokončené vzdělání; $y_1 = 1$ (ZŠ), $y_2 = 3$ (SŠm) …
  (c) Y …počet starších sourozenců $y_1 = 0$, $y_2 = 2$ …
  (d) X …porodní hmotnost v g; $x_1 = 3470$, $x_2 = 3240$ …
  (e) Y …největší šířka mozkovny v mm; $y_1 = 145$, $y_2 = 139$ …
- Dva typy náhodných veličin:
  - **Diskrétní** – z podstaty nabývají převážně celých hodnot (a), (b), (c)
    - počet sourozenců: 0, 3, 2, …; novorozenec nemůže mít 2.4 sourozence
    - hod kostkou: padne 1, 2, 3, 4, 5, 6; nemůže padnout 3.5
      - $\Pr(X = 4) = \ldots$
      - $\Pr(X \leq 4) = \ldots$
      - $\Pr(X > 4) = \Pr(X \geq 5) = \ldots$
      - $\Pr(3 < X \leq 5) = \ldots$
  - **Spojité** – z podstaty mohou nabývat libovolných i neceločíselných hodnot (d), (e)
    - porodní hmotnost:
      - základní prostor rozdělíme na intervaly: I1: 0–1500; I2: 1500–2500; I3: 2500–3500; I4: 3500–4500; I5:>4500
      - $\Pr(X \in \langle 3500\,;\,4500\rangle) = \ldots$
      - $\Pr(X \leq 3500) = \ldots$
      - $\Pr(X > 3500) = \ldots$
- Pravděpodobnostní funkce $p(x)$ (X je diskrétní):
  - $p(x) = \Pr(X = x)$
  - pstní fce pro případ hodu kostkou:

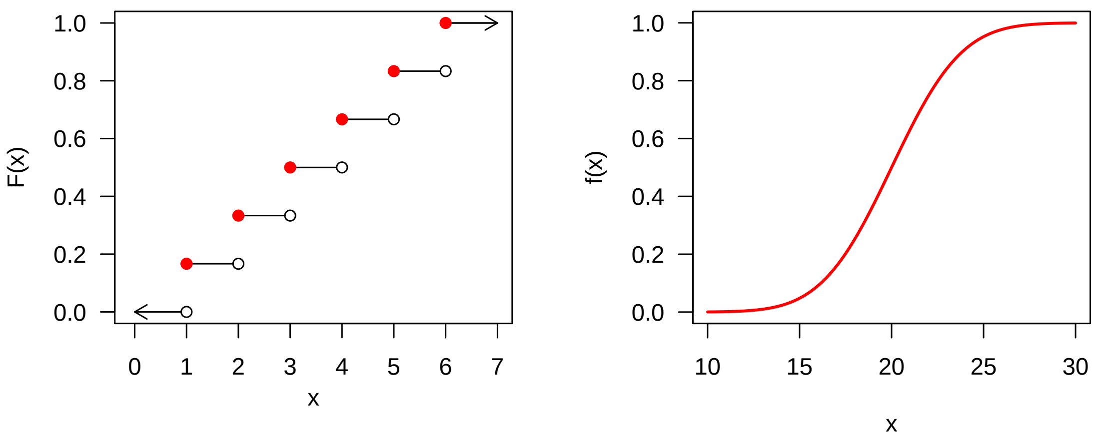


  - nezáporná: $\Pr(x) \geq 0$; normovaná: $\sum_{i=1}^{\infty} \Pr(X = x_i) = 1$

- hustota $f(x)$ (X je spojitá)
  - Pst realizace X v libovolném intervalu I se dá vyjádřit jako plocha pod křivkou pomocí integrálního tvaru:
  $$\Pr(X \in I) = \int_{x \in I} f(x)dx,$$
  kde $f(x)$ je *hustota* pravděpodobnosti spojité náhodné veličiny
  - nezáporná: $f(x) \geq 0$; normovaná (plocha pod křivkou hustoty = 1)
  - Hustota normálního rozdělení (Gaussova křivka):



- distribuční funkce $F(x)$ (X je diskrétní / spojitá)
  - $F(x) = \Pr(X \leq x)$
  - příklad: distribuční fce hodu kostkou:



  - **Komplementarita**: $\Pr(X > x) = 1 - \Pr(X \leq x) = 1 - F(x)$
  - platí v diskrétním případě:
    - $\Pr(X = x) = p(x)$
  - platí ve spojitém případě:
    - $\Pr(X = x) = 0$

Přehled rozdělení

- máme náhodný výběr a rádi bychom věděli z jakého pochází rozdělení.
  - **Diskrétní**
    - Binomické $\mathrm{Bin}(N, p)$;
    - Poissonovo $\mathrm{Po}(\lambda)$
  - **Spojité**
    - Normální $N(\mu, \sigma^2)$
    - Dvourozměrné normální $N_2(\boldsymbol{\mu}, \boldsymbol{\Sigma})$
- Nejdříve odhadneme typ rozdělení, potom parametry rozdělení

4.3 Diskrétní náhodné veličiny

4.3.1 Binomické rozdělení $\mathrm{Bin}(N, p)$

- Bernoulliho (opakované) pokusy $X_1, \ldots, X_N$:
  - $X_i = 1$ ... událost nastala; $X_i = 0$... událost nenastala; $i = 1, \ldots, N$.
  - $\Pr(X_i = 1) = p$
  - $\Pr(X_i = 0) = 1 - p = q$
- Binomické rozdělení:
  - X... počet událostí v posloupnosti N nezávislých Bernoulliho pokusů, přičemž pravděpodobnost nastání události v každém pokusu je vyjádřena parametrem p.
  - $\sum_{i=1}^{N} X_i = X \sim \mathrm{Bin}(N, p)$.
  - $\boldsymbol{\theta} = (N, p)$
  - pravděpodobnostní funkce:

$$p(x) = \binom{N}{x} p^x (1-p)^{N-x} \qquad x = 0, 1, \ldots, N;$$

  - vlastnosti: $\mathrm{E}[X] = Np$; $\mathrm{Var}[X] = Np(1-p)$
  - dbinom(x, N, p), pbinom(x, N, p), rbinom(M, N, p)

4.3.2 Poissonovo rozdělení $\mathrm{Po}(\lambda)$

- X ... počet událostí, které nastanou v jednotkovém časovém intervalu, přičemž k událostem dochází náhodně, jednotlivě a vzájemně nezávisle. Střední počet těchto událostí je vyjádřen parametrem $\lambda > 0$.
- $X \sim \mathrm{Po}(\lambda)$
- $\theta = \lambda$
- pravděpodobnostní funkce:

$$p(x) = \frac{\lambda^x}{x!} e^{-\lambda} \; x = 0, 1, \ldots;$$

- vlastnosti: $E[X] = \lambda$; $\mathrm{Var}[X] = \lambda$
- dpois(x, lambda), ppois(x, lambda)

4.4 Spojité náhodné veličiny

4.4.1 Normální rozdělení $N(\mu, \sigma^2)$

- $X_1, \dots, X_n$ ... nezávislé náhodné veličiny
- Normální rozdělení
  - $X \sim N(\mu, \sigma^2)$
  - $\boldsymbol{\theta} = (\mu, \sigma^2)^T$
  - hustota
  $$f(x) = \frac{1}{\sqrt{2\pi\sigma^2}} \exp\left\{\frac{-(x-\mu)^2}{2\sigma^2}\right\}, \qquad x \in \mathbb{R}.$$
  - vlastnosti $\mathrm{E}[X] = \mu$; $\mathrm{Var}[X] = \sigma^2$
  - dnorm(x, mu, sigma), pnorm(x, mu, sigma)
- Standardizované normální rozdělení
  - $X \sim N(0, 1)$
  - $\boldsymbol{\theta} = (0, 1)^T$
  - hustota
  $$f(x) = \phi(x) = \frac{1}{\sqrt{2\pi}} \exp\left\{-\frac{x^2}{2}\right\}, \qquad x \in \mathbb{R}.$$
  - vlastnosti $\mathrm{E}[X] = 0$; $\mathrm{Var}[X] = 1$
  - dnorm(x), pnorm(x)
- Vlastnosti normální rozdělení
  - **Věta 1:** Nechť $X_1, \dots, X_n$ jsou nezávislé náhodné veličiny z normálního rozdělení $N(\mu, \sigma^2)$. Potom náhodná veličina $\bar{X}_n = \frac{1}{n}\sum_{i=1}^{n} X_i \sim N\left(\mu, \frac{\sigma^2}{n}\right)$.

4.4.2 Aproximace binomického rozdělení normálním rozdělením

- Normální rozdělení je limitním rozdělením binomického rozdělení $\mathrm{Bin}(N, p)$, tedy
$$X \sim \mathrm{Bin}(N, p) \to X \sim N(\mu, \sigma^2),$$
kde $\mu = Np$ a $\sigma^2 = Np(1-p)$.
- Podmínky: $N \to \infty$, $p \to 0.5$
- Haldova podmínka: Nechť $X \sim \mathrm{Bin}(N, p)$ a platí, že $Np > 5$ a $N(1-p) > 5$. Potom rozdělení náhodné proměnné X můžeme aproximovat normálním rozdělením $X \sim N(Np, Np(1-p))$.