1. Pravděpodobnost a statistika *(MP leden 2010)*

Pravděpodobnost – pojmy

1. Diskrétní pravděpodobnostní prostor (definice, vlastnosti, příklad).
   * *Diskrétní pravděpodobnostní prostor* je trojice $(\Omega, \mathbf{A}, \mathbf{P})$, kde
     · Ω je množina všech *elementárních jevů*
     · $\mathbf{A}$ je množina *možných jevů* (platí $\mathbf{A} \subset 2^\Omega$)
     · $\mathbf{P}$ je *pravděpodobnost*, neboli funkce $\mathbf{P} : \mathbf{A} \to [0, 1]$

   splňující následující axiomy (první 3 se nazývají σ-algebra):
   (1) $\emptyset \in \mathbf{A}$
   (2) $A \in \mathbf{A} \Rightarrow \Omega \setminus A \in \mathbf{A}$
   (3) $A_1, A_2, \ldots, A_n \in \mathbf{A} \Rightarrow \bigcup_{i=1}^{\infty} A_i \in \mathbf{A}$
   (4) $\mathbf{P}(\Omega) = 1$
   (5) $A_1, A_2, \ldots, A_n$ disjunktní $\Rightarrow$

$$\mathbf{P}(\bigcup_{i=1}^{\infty} A_i) = \sum_{i=1}^{\infty} \mathbf{P}(A_i)$$

2. Elementární jev.
   * Prvek množiny Ω. Každý *elementární jev* má stejnou pravděpodobnost, že nastane a každé dva elementární jevy jsou disjunktní.
3. Nezávislost jevů.
   * Jevy A a B jsou *nezávislé*, jestliže $\mathbf{P}(A \cap B) = \mathbf{P}(A) \cdot \mathbf{P}(B)$.
   - Sdružená nezávislost dvou jevů.
     * Jevy $A_1, A_2, \ldots, A_n$ jsou *sdruženě (vesměs) nezávislé*, jestliže

$$\forall J \subset \{1, \ldots, n\} : \mathbf{P}(\bigcap_{j \in J} A_j) = \prod_{j \in J} \mathbf{P}(A_j).$$

   - Neslučitelnost (dvou) jevů.
     * Dva jevy jsou *neslučitelné*, jestliže nemohou nastat zároveň. Neboli pravděpodobnost jejich průniku je nulová.
   - Nezávislost po dvou.
     * Jevy $A_1, A_2, \ldots, A_n$ jsou *po dvou nezávislé*, jestliže jsou každé dva z nich nezávislé.

4. Podmíněná pravděpodobnost + příklad použití.

   * *Podmíněná pravděpodobnost* jevu A za předpokladu B je pravděpodobnost, že A nastal, za podmínky, že nastal B a definuje se
   $$\mathbf{P}(A \mid B) = \frac{\mathbf{P}(A \cap B)}{\mathbf{P}(B)}.$$

   * Příklad – házení kostkou. Jaká je pravděpodobnost, že padlo číslo vyšší než 3, za podmínky, že hozené číslo je sudé. Jev A je „padlo více než 3“ a jev B je „padlo sudé číslo“.
   $$\mathbf{P}(A \mid B) = \frac{\mathbf{P}(A \cap B)}{\mathbf{P}(B)} = \frac{\mathbf{P}(\text{„padlo 4 nebo 6“})}{\mathbf{P}(\text{„padlo 2, 4 nebo 6“})} = \frac{2}{3}.$$

5. Věta o úplné pravděpodobnosti + příklad použití.

   * Mějme pravděpodobnostní prostor $(\Omega, \mathbf{A}, \mathbf{P})$, jev $A \in \mathbf{A}$ a disjunktní jevy H_i pokrývající celé $\mathbf{A}$, neboli $\forall i \in I : H_i \in \mathbf{A}$, $\bigcup_{i \in I} H_i = \Omega$, $\forall i \neq j : H_i \cap H_j = \emptyset$. Pak platí:
   $$\mathbf{P}(A) = \sum_{i \in I} \mathbf{P}(A \mid H_i) \cdot \mathbf{P}(H_i).$$

   Důkaz: Učiňme pozorování, že jelikož $A \in \mathbf{A}$, tak $A \cap \Omega = A$. Proto platí první rovnost:
   $$\mathbf{P}(A) = \mathbf{P}(A \cap \Omega) = \mathbf{P}(A \cap (\bigcup_{i \in I} H_i)) = \mathbf{P}(\bigcup_{i \in I}(A \cap H_i)) =$$
   $$= \sum_{i \in I} \mathbf{P}(A \cap H_i) = \sum_{i \in I} \mathbf{P}(A \mid H_i) \cdot \mathbf{P}(H_i).$$
   Druhá rovnost je z definice jevů H_i (pokrývají celou množinu elementárních jevů). Třetí rovnost jsou de Morganova pravidla. Čtvrtá rovnost je pátý axiom, neboť pro různá i jsou jevy $A \cap H_i$ disjunktní, neboli $\forall i \neq j : (A \cap H_i) \cap (A \cap H_j) = \emptyset$. Poslendí rovnost je definice podmíněné pravděpodobnosti. ♡

   * Příklad – cesta do školy. Pravděpodobnost, že mi ujede tramvaj, je 0.3. Pravděpodobnost, že stihnu hodinu, pokud mi ujede tramvaj, je 0.1 (budu muset běžet). Pravděpodobnost, že stihnu hodinu, pokud tramvaj stihnu, je 0.8 (může se ještě něco stát na cestě od tramvaje). Jaká je pravděpodobnost, že stihnu hodinu? Označme $A =$ „stihnu hodinu“. $H_1 =$ „tramvaj mi ujede“. $H_1 =$ „tramvaj mi neujede“.
   $$\mathbf{P}(A) = \mathbf{P}(A \mid H_1) \cdot \mathbf{P}(H_1) + \mathbf{P}(A \mid H_2) \cdot \mathbf{P}(H_2)$$
   $$= 0.1 \cdot 0.3 + 0.8 \cdot 0.7 = 0.59$$

6. Bayesova věta.

   * Mějme pravděpodobnostní prostor $(\Omega, \mathbf{A}, \mathbf{P})$, jev $A \in \mathbf{A}$ a disjunktní jevy H_i pokrývající celé $\mathbf{A}$, neboli $\forall i \in I : H_i \in \mathbf{A}$, $\bigcup_{i \in I} H_i = \Omega$, $\forall i \neq j : H_i \cap H_j = \emptyset$. Pak platí:

   $$\mathbf{P}(H_i \mid A) = \frac{\mathbf{P}(A \mid H_i) \cdot \mathbf{P}(H_i)}{\sum_{j \in I} \mathbf{P}(A \mid H_j) \cdot \mathbf{P}(H_j)}.$$

   Důkaz:

   $$\mathbf{P}(H_i \mid A) = \frac{\mathbf{P}(H_i \cap A)}{\mathbf{P}(A)} = \frac{\mathbf{P}(A \mid H_i) \cdot \mathbf{P}(H_i)}{\sum_{j \in I} \mathbf{P}(A \mid H_j) \cdot \mathbf{P}(H_j)}.$$

   První rovnost je definice podmíněné pravděpodobnosti. Druhá rovnost je též definice podmíněné pravděpodobnosti a věta o úplné pravděpodobnosti. ♡

   * Příklad – střelci a kanci. 18 střelců střílelo na kance. 5 z nich se strefí s pstí 0.8, 7 s pstí 0.6, 4 s pstí 0.5 a 2 s pstí 0.4. Náhodně vybraný střelec minul. Jaká je pravděpodobnost, že střelec patřil k první skupině?

   Označme $A =$ „střelec minul“, $H_i =$ „střelec byl z $i$té skupiny“. Pak známe pravděpodobnosti $\mathbf{P}(A^c \mid H_i)$, kde A^c je doplněk jevu A, tedy že se střelec strefil. Jevy H_i jsou disjunktní a pokrývají celé Ω. Pak pravděpodobnost $\mathbf{P}(H_i \mid A)$ vypočteme pomocí Bayesovy věty

   $$\mathbf{P}(H_i \mid A) = \frac{0.2 \cdot \frac{5}{18}}{0.2 \cdot \frac{5}{18} + 0.4 \cdot \frac{7}{18} + 0.5 \cdot \frac{4}{18} + 0.6 \cdot \frac{2}{18}} = \frac{1}{7}$$

7. Náhodná veličina.

   * Mějme $(\Omega, \mathbf{A}, \mathbf{P})$. *Náhodná veličina* je funkce $X : (\Omega, \mathbf{A}, \mathbf{P}) \to (\mathbf{R_1}, B)$, kde B je třída množin (typicky intervalů) v $\mathbf{R_1}$. Nebo zjednodušeně $X : \Omega \to \mathbf{R}$.

8. Nezávislé náhodné veličiny.

   * *Náhodné veličiny* jsou nezávislé, jestliže pro všechny hodnoty a, b platí $\mathbf{P}(X = a \;\&\; Y = b) = \mathbf{P}(X = a) \cdot \mathbf{P}(Y = b)$.

9. Diskrétní náhodná veličina a její charakteristiky.

∗ Diskrétní náhodná veličina je náhodná veličina, která může nabývat pouze spočetně mnoha hodnot. Její charakteristiky jsou:

- Vektor hodnot a pravděpodobností.

  ∗ Vektor hodnot x_i a pravděpodobností p_i, se kterými těchto hodnot nabývá. Někdy se tento vektor nazývá *rozdělení pravděpodobnosti*.

- Distribuční funkce.

  ∗ *Distribuční funkce* náhodné veličiny je funkce

  $$F(x) = \mathbf{P}(X \leq x),$$

  kde $x \in \mathbf{R}$. Distribuční funkce je neklesající. Pro diskrétní náhodnou veličinu je po částech konstantní. Platí $F(-\infty) = 0$ a $F(\infty) = 1$.

- Střední hodnota.

  ∗ *Střední hodnota* náhodné veličiny X dané vektorem hodnot $\left\{\binom{x_i}{p_i}\right\}$, kde $P(X = x_i) = p_i$, je

  $$EX = \sum_{\omega \in \Omega} X(\omega)\mathbf{P}(\omega) = \sum_i x_i p_i.$$

- Rozptyl.

  ∗ *Rozptyl* náhodné veličiny X je definován jako

  $$\mathrm{var}X = E(X - EX)^2 = EX^2 - (EX)^2.$$

- Směrodatná odchylka.

  ∗ *Směrodatná odchylka* náhodné veličiny X je definována jako $\sqrt{\mathrm{var}X}$.

- Kovariace.

  ∗ *Kovariance* náhodných veličin X a Y je

  $$\mathrm{cov}(X, Y) = E(X - EX)(Y - EY).$$

- Korelace.

  ∗ *Korelace* náhodných veličin X a Y je

  $$\mathrm{cor}(X, Y) = \frac{\mathrm{cov}(X, Y)}{\sqrt{\mathrm{var}X}\sqrt{\mathrm{var}Y}}.$$

10. Vlastnosti střední hodnoty.

   * Nechť X je náhodná veličina a $g(x)$ je funkce. Pak $Y = g(X)$ je také náhodná veličina a její střední hodnota je $\sum_i g(x_i)p_i$.
   * Nechť X a $Y = a + bX$ jsou náhodné veličiny, kde reálná čísla a, b se nazývají po řadě *posun* a *změna měřítka*. Pak

$$EY = E(a + bX) = a + bEX.$$

Důkaz:

$$E(a + bX) = \sum_\omega (a + bX)(\omega)\mathbf{P}(\omega) = \sum_\omega (a + b \cdot X(\omega))\mathbf{P}(\omega) =$$

$$= \sum_\omega a \cdot \mathbf{P}(\omega) + \sum_\omega (b \cdot X(\omega))\mathbf{P}(\omega) = a + b \sum_\omega X(\omega)\mathbf{P}(\omega) = a + bEX.$$

♡

   * Tedy nutně platí:

$$Ea = a \ , \quad E(bX) = bEX \quad \text{a} \quad E(X + Y) = EX + EY.$$

11. Vlastnosti rozptylu.

   * Platí var$X = EX^2 - (EX)^2$.
     Důkaz: var$X = E(X - EX)^2 = E(X^2 - 2XEX + (EX^2) = EX^2 - 2EX \cdot EX + E(EX)^2 = EX^2 - (EX)^2$. Využili jsme, že EX je konstanta, tedy ji můžeme vytýkat a střední hodnota konstanty je konstanta. ♡
   * Platí var$(a) = 0$.
     Důkaz: $Ea^2 - (Ea)^2 = 0$. ♡
   * Platí var$(a + X) =$ var$X$.
     Důkaz:

$$\text{var}(a + X) = E(a + X)^2 - (E(a + X))^2 =$$

$$= E(a^2 + 2aX + X^2) - (a^2 + 2aEX + (EX^2)) = EX^2 - (EX)^2 = \text{var}X.$$

♡

   * Platí var$(bX) = a^2$var(X).
     Důkaz:

$$\text{var}(bX) = E(bX)^2 - (E(bx))^2 = b^2EX^2 - b^2(EX)^2 = b^2\text{var}X$$

♡

12. Vlastnosti kovariance a korelace

    * $\text{cov}(X, X) = E(X - EX)(X - EX) = \text{var}X$.
    * $-1 \leq \text{cor}(X, Y) \leq 1$
    * Jsou-li X, Y nezávislé, pak $\text{cov}(X, Y) = 0$ a $\text{cor}(X, Y) = 0$.

$$E(X - EX)(Y - EY) = EXY - EX \cdot EY = 0.$$

    * $\text{cor}(X, X) = 1$.
    * $\text{cor}(X, -X) = 1$.

13. Spojitá náhodná veličina a její charakteristiky.

    * Spojitá náhodná veličina je náhodná veličina, která může nabývat všech hodnot ze spojitého intervalu. Její charakteristiky jsou:

- Distribuční funkce.

    * *Distribuční funkce* náhodné veličiny je funkce

$$F(x) = \mathbf{P}(X \leq x),$$

kde $x \in \mathbf{R}$. Distribuční funkce je neklesající. Pro diskrétní náhodnou veličinu je po částech konstantní. V $-\infty$ má distribuční funkce nulovou hodnotu a v nekonečnu 1.

- Hustota.

    * *Hustota* spojité náhodné veličiny X je

$$f(x) = F'(x),$$

kde $F(x)$ je distribuční funkce této náhodné veličiny. Platí, že má pro všechny hodnoty x nezáporné funkční hodnoty a $\int_{-\infty}^{\infty} f(x)dx = 1$.

- Střední hodnota.

    * *Střední hodnota* spojité náhodné veličiny X je

$$EX = \int_{R_1} xf(x)dx.$$

14. Čebyševova nerovnost (znění a důkaz).

* Nechť náhodná veličina X má konečný rozptyl var$X$. Pak

$$\forall \varepsilon > 0 \quad \mathbf{P}[|X - EX| \geq \varepsilon] \leq \frac{\mathrm{var}X}{\varepsilon^2}.$$

Důkaz:

$$\mathrm{var}X = E(X - EX)^2 = \sum_i p_i(x_i - EX)^2 \geq \sum_{i|(x_i - EX)^2 \geq \varepsilon^2} p_i(x_i - EX)^2 \geq$$

$$\geq \sum_{i|(|(x_i - EX)| \geq \varepsilon)} p_i\varepsilon^2 = \varepsilon^2 \mathbf{P}[|X - EX| \geq \varepsilon^2]$$

♡

15. Čebyševova věta.

* Nechť $X_1, \ldots, X_n$ jsou nezávislé veličiny s konečným rozptylem σ^2 a střední hodnotou μ. Potom

$$\forall \varepsilon > 0 \quad \mathbf{P}[|\overline{X_n} - \mu| \geq \varepsilon] \to 0 \text{ pro } n \to \infty$$

16. Centrální limitní věta.

* Nechť $X_1, X_2, \ldots, X_n$ jsou nezávislé stejně rozdělené náhodné veličiny s konečnou střední hondotou. Potom

$$\mathbf{P}\left[\frac{\sum_{i=1}^n X_i - E\sum_{i=1}^n X_i}{\sqrt{\mathrm{var}\sum_{i=1}^n X_i}} \leq x\right] \to \Phi(x).$$

* Nechť $Y \sim Bi(n, p)$. Potom

$$\mathbf{P}\left[\frac{Y - EY}{\sqrt{\mathrm{var}Y}} \leq x\right] \to \Phi(x).$$

* Přitom platí, že

$$\Phi(x) = \int_{-\infty}^x \frac{1}{\sqrt{2\pi}} e^{-\frac{t^2}{2}} dt.$$

17. Zákon velkých čísel.

* Mějme nezávislé stejně rozdělené náhodné veličiny $X_1, \ldots, X_n$ s konečnou střední hodnotou $EX_n = \mu$. Potom

$$\mathbf{P}\left[\omega \in \Omega : \lim_{n\to\infty} \frac{1}{n} \sum_{j=1}^n X_j(\omega) = \mu\right] = 1$$

Rozdělení – diskrétní rozdělení

1. Rovnoměrné rozdělení
   * $R(M)$
   * X nabývá hodnot z množiny $\{1, \ldots, M\}$.
   * $\mathbf{P}[X = k] = \frac{1}{M}$

$$EX = \frac{M+1}{2}$$

   * Všechny hodnoty mají stejnou pravděpodobnost. Např. kolik padne na kostce při jednom hodu.
2. Alternativní rozdělení
   * $Alt(p)$
   * X nabývá hodnot z množiny $\{0, 1\}$.
   * $\mathbf{P}[X = 0] = 0$ a $\mathbf{P}[X = 1] = 1$

$$EX = p \qquad \text{var}X = p(1-p)$$

   * Jsou jen dvě hodnoty, které mohou nastat, jedna má pst p, druhá $(1-p)$. Např. zda padne panna nebo orel při jednom hodu mincí.
3. Binomické rozdělení
   * $Bi(n, p)$
   * X nabývá hodnot z množiny $\{0, 1, \ldots, n\}$.

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

$$EX = np \qquad \text{var}X = np(1-p)$$

   * Součet alternativních rozdělení, neboli n nezávislých dichotomických pokusů. Např. n hodů mincí a počet orlů.
4. Geometrické rozdělení
   * $Ge(p)$
   * X nabývá hodnot z množiny $\{1, 2, \ldots\}$.

$$\mathbf{P}[X = k] = p(1-p)^{k-1}$$

$$EX = \frac{1}{p} \qquad \text{var}X = \frac{1}{p}(\frac{1}{p} - 1)$$

   * Geometrické rozdělení je jediné diskrétní rozdělení bez paměti.

$$\mathbf{P}[X > a + b \mid X > a] = \mathbf{P}[X > b]$$

* Čekání na první zdar. Např. počet hodů mincí, dokud nepadne panna. Tedy první zdar v k-tém pokuse.
* Obdobně počet nezdarů před prvním zdarem, což má pst

$$\mathbf{P}[X = k] = p(1-p)^k$$

* Střední hodnota a rozptyl lze vypočítat jako první a druhá derivace vytvořující funkce

$$A(x) = \sum_{i=0}^{\infty} pq^i x^i = \frac{p}{1-qx}$$

5. Negativně geometrické rozdělení
    * Před r-tým zdarem máme i nezdarů, resp. r-tý zdar v i-tém pokusu.
    * První má pst

$$\binom{r+i-1}{r-1} q^i p^r$$

a druhé má pst

$$\binom{i-1}{r-1} q^i p^r.$$

6. Hypergeometrické rozdělení
    * $Hyp(N, M, n)$
    * X nabývá hodnot z množiny $\{0, 1, \ldots, \min n, M\}$.

$$\mathbf{P}[X = k] = \frac{\binom{M}{k}\binom{N-M}{n-k}}{\binom{N}{n}}$$

$$EX = \frac{nM}{N}$$

$$\mathrm{var}X = \frac{nM}{N}\left(1 - \frac{M}{N}\right)\frac{N-n}{N-1}$$

* Např. N koulí, M z nich bílých, tahám n. X je počet bílých vytažených koulí (tahy bez vracení).

7. Poissonovo rozdělení
    * $Pois(\lambda)$
    * X nabývá hodnot z množiny $\{0, 1, \ldots\}$.

$$\mathbf{P}[X = k] = e^{-\lambda}\frac{\lambda^k}{k!}$$

$$EX = \lambda$$

$$\mathrm{var}X = \lambda$$

* Počet událostí za jednotku času.

Rozdělení – spojitá rozdělení

1. Rovnoměrné rozdělení
   - $R(a, b)$
   - X nabývá hodnot z množiny $\{a, \ldots, b\}$.
   - $f(x) = \frac{1}{b-a}$ pro $x \in [a, b]$, jinde 0.

$$EX = \frac{a+b}{2}$$

$$\text{var}X = \frac{(b-a)^3}{12}$$

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

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

   - $EX = \mu$ a $\text{var}X = \sigma^2$

3. Normované normální rozdělení
   - $N(0, 1)$

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

   - $EX = 0$ a $\text{var}X = 1$
   - $f(-x) = f(x)$ a $\Phi(-x) = 1 - \Phi(x)$, kde Φ je distribuční funkce.
   - $\Phi(u_\alpha) = \alpha$ se nazývá α-*kvantil*.
   - Všechny hodnoty mají stejnou pravděpodobnost.

4. Exponenciální rozdělení
   - $Exp(\lambda)$
   - X nabývá hodnot z množiny $\{0, \ldots\}$.
   - $f(x) = \lambda e^{-\lambda x}$ pro $x > 0$, jinde 0.

$$EX = \frac{1}{\lambda}$$

$$\text{var}X = \frac{1}{\lambda^2}$$

   - Exponenciální rozdělení je jediné spojité rozdělení bez paměti.

$$\mathbf{P}[X > a + b \mid X > a] = \mathbf{P}[X > b]$$

   - Doba čekání na určitou událost.

Statistika – pojmy

1. Bodový odhad.

   * Mějme náhodný výběr rozsahu n (posloupnost n nezávislých stejně rozdělených veličin) $X_1, X_2, \ldots, X_n$ z rozdělení, které závisí na parametru Θ. Najít bodový odhad znamená najít takovou funkci náhodných veličin $X_1, X_2, \ldots, X_n$ (též statistiku), která je v nějakém smyslu blízko určené hodnotě Θ. Označme tento odhad $T = T(X_1, X_2, \ldots, X_n)$. Protože je T funkce náhodných veličin, je také náhodnou veličinou. Zároveň se T nazývá *bodovým odhadem.*

2. Vychýlení.

   * Vektor konstatnt $\mathbf{b} = ET - \Theta$ se nazývá *vychýlení.*

3. Nestranný odhad parametru Θ.

   * Odhad T parametru Θ je *nestranný (nevychýlený)*, jestliže je nulové vychýlení, neboli

$$ET = \Theta.$$

4. Konzistentní odhad.

   * Odhad $T = T_n$ je *konzistentní* odhad parametru Θ , jestliže platí

$$\forall \varepsilon > 0 \qquad lim_{n\to\infty}\mathbf{P}[|T_n - \Theta| < \varepsilon] = 1.$$

   * Např. pokud $lim_{n\to\infty}ET_n = \Theta$ a $lim_{n\to\infty}\text{var}T_n = 0$, pak T_n je konzistentním odhadem Θ.
   * Odhad může být konzistentní a zároveň nemusí být nestranný. Např. náhodný výběr $X_1, \ldots, X_n$ z $N(\mu, \sigma^2)$. Pro odhad rozptylu se používá statistika

$$\hat{\sigma}^2 = \frac{1}{n}\sum_{i=1}^{n}(X_i - \overline{X})^2.$$

   To není nestranný odhad, neboť $E\hat{\sigma}^2 = \sigma^2$. Ale je konzistentní, neboť $lim_{n\to\infty}\text{var}\hat{\sigma}^2 = 0.$⟨1⟩

5. Intervalový odhad, interval spolehlivosti.

   * Platí-li pro statistiky

$$T_L = T_L(X_1, \ldots, X_n), \quad T_U = T_U(X_1, \ldots, X_n)$$

⟨1⟩ To by chtělo podrobnější odvození.

vztah

$$P[T_L \leq \Theta \leq T_U] = 1 - \alpha,$$

říkáme, že (T_L, T_U) tvoří *interval spolehlivosti (intervalový odhad, konfidenční interval)* pro parametr Θ s koeficientem spolehlivosti $1 - \alpha$.

6. Chyba prvního druhu.

   * $P[$ zamítneme $H_0 \mid H_0$ platí $]$, neboli, že zamítneme platnou hypotézu. Značí se α.

7. Chyba druhého druhu.

   * $P[$nezamítneme $H_0 \mid H_0$ neplatí$]$, neboli, že nezamítneme neplatnou hypotézu. Značí se β.

8. Síla testu.

   * $1-\beta$. Někdy při hledání opptimálního testu stanovíme hladinu významnosti α (omezíme pravděpodobnost chyby 1. druhu) a mezi α-testy hledáme ten s největší sílou (nejmenší chybou 2. druhu).

9. Hladina testu.

   * Hladina testu je maximální dovolená chyba prvního druhu.

10. P-hodnota testu.

    * Nejmenší hladina, při které bychom ještě hypotézu zamítli.
    * Pravděpodobnost, s jakou testovací statistika nabývá „horších hodnot“ (více svědčících proti hypotéze). Hypotézu H_0 zamítáme na hladině α, právě když p-hodnota je menší než α.

11. Kritický obor

    * Množina výsledků pokusu, při kterých budeme hypotézu zamítat. Značí se W.

12. Na příkladu falšené mince popsat testování hypotéz.

    * Viz papíry.

13. Rozdíl mezi párovým a dvouvýběrovým testem.

    * *Párový test* je test hypotézy o hodnotě rozdílu středních hodnot μ_1 a μ_2 složek náhodných vektorů v náhodném výběru $(X_1, Y_1), \ldots, (X_n, Y_n)$ z dvourozměrného normálního rozdělení. (Uvnitř každé dvojice nemusí jít o nezávislé veličiny.)

$$\begin{aligned} H_0 &: \quad \mu_1 - \mu_2 = d \\ H_1 &: \quad \mu_1 - \mu_2 \neq d \end{aligned}$$

Můžeme zavést náhodnou veličinu $Z_i = X_i - Y_i$ pro $i = 1, \ldots, n$. Pokud neznáme rozptyl, můžeme použít t-test pro

$$T = \frac{\overline{Z} - d}{S_Z/\sqrt{n}},$$

kde $\overline{Z}$ je výběrový průměr a S_Z^2 je výběrový rozptyl (odhadnutý). Hypotézu H_0 zamítneme na hladině významnosti α, pokud

$$T \in W = (-\infty, t_{\frac{\alpha}{2}}(n-1)) \cup (t_{1-\frac{\alpha}{2}}(n-1)).$$

Kde $t_{1-\frac{\alpha}{2}}(n-1)$ je $(1-\frac{\alpha}{2})$-kvantil studentova rozdělení s $(n-1)$ stupni volnosti.

* *Dvouvýběrový test* je testem hypotézy o hodnotě rozdílu středních μ_1 a μ_2 ve dvou nezávislých náhodných výběrech

  $X_1, \ldots, X_n$ z normálního rozdělení $N(\mu_1, \sigma^2)$
  $Y_1, \ldots, Y_m$ z normálního rozdělení $N(\mu_2, \sigma^2)$

  se stejným (i když klidně neznámým) rozptylem σ^2.

$$\begin{array}{lcl} H_0 & : & \mu_1 - \mu_2 = d \\ H_1 & : & \mu_1 - \mu_2 \neq d \end{array}$$

K testování můžeme použít statistiku (opět t-test)

$$T = \frac{\overline{X} - \overline{Y} - d}{\sqrt{(n-1)S_X^2 + (m-1)S_Y^2}} \sqrt{\frac{nm(n+m-2)}{n+m}},$$

kde S_X^2 a S_Y^2 jsou příslušné výběrové rozptyly (odhadnuté). Hypotézu H_0 zamítneme na hladině významnosti α, pokud

$$T \in W = (-\infty, t_{\frac{\alpha}{2}}(n+m-2)) \cup (t_{1-\frac{\alpha}{2}}(n+m-2)).$$

Kde $t_{1-\frac{\alpha}{2}}(n+m-2)$ je $(1-\frac{\alpha}{2})$-kvantil studentova rozdělení s $(n+m-2)$ stupni volnosti.

14. Model lineární regrese.

    * Lineární regrese představuje aproximaci daných (naměřených) hodnot (X_i, Y_i) přímkou. Mějme nezávislé náhodné veličiny $Y_1, \ldots, Y_n$ z $N(\beta_0 + \beta_1 x_i, \sigma^2)$, kde x_i jsou dané nestejně velké konstanty. Rozptyly Y_i jsou tedy stejné (σ^2), ale střední hodnoty lze vyjádřit jako linerání funkci známých konstant x_i $(\beta_0 + \beta_1 x_i)$ pomocí neznámých parametrů β_0, β_1.

15. Reziduum.

    * *Reziduum* je hodnota $r_i = Y_i - \beta_0 - \beta_1 x_i$.

Poznámky ke statistice – výpisky Zvára str. 140 – 162[2]

1. Test nezávislosti v normálním rozdělení.
   * Pro testování nezávislosti složek náhodných vektorů $(X_1, Y_1), \ldots, (X_n, Y_n)$ v náhodném výběru z dvourozměrného normálního rozdělení vektoru (X, Y) a hypotézy

     $$\begin{array}{lcl} H_0 & : & X, Y \text{ nezávislé} \\ H_1 & : & X, Y \text{ závislé} \end{array}$$

     se používá test založený na statistice

     $$T = \frac{r}{\sqrt{1-r^2}}\sqrt{n-2},$$

     kde r je výběrový korelační koeficient. Hypotézu H_0 zamítneme na hladině významnosti α, pokud

     $$|T| > t_{1-\frac{\alpha}{2}}(n-2).$$

     Kde $t_{1-\frac{\alpha}{2}}(n-2)$ je $(1-\frac{\alpha}{2})$-kvantil studentova rozdělení s $(n-2)$ stupni volnosti.
2. *Výběrový průměr* naměřených hodnot x_i je jejich aritmetický průměr

   $$\overline{x} = \frac{1}{n}\sum_{i=1}^{n} x_i.$$

   Přičemž platí, že $\overline{(a + xb)} = a + b\overline{x}$.
3. *Medián* je prostřední hodnota v setříděném souboru naměřených hodnot. Pokud jich je sudo, tak je to aritmetický průměr prostředních dvou.
4. *(Výběrový) p-tý kvantil (percentil)* je v uspořádaném souboru hodnot definovaný $x_p = x_{\lfloor np \rfloor + 1}$ pro $np \neq \lfloor np \rfloor$ a aritemtický průměr $1/2 x_{np} + x_{np+1}$ pro $np = \lfloor np \rfloor$.
5. *Výběrový rozptyl* je definován jako

   $$s_x^2 = \frac{1}{n-1}\sum_{i=1}^{n}(x_i - \overline{x})^2 = \frac{1}{n-1}\sum_{i=1}^{n}(x_i^2 - \overline{x}^2).$$

6. *Směrodatná odchylka* s_x je definována jako odmocnina z rozptylu.
7. Krabicový diagram (box plot) znázorňuje medián, 25%ní a 75%ní kvantil a horní měření a dolní měření, resp. 10%ní a 90%ní kvantil.

---

[2] *Pravděpodobnost a matematická statistika – Karel Zvára a Josef Štěpán*

8. *Populace* neboli *základní soubor* je statistický soubor, na kterém děláme všechna měření. Jeho velikost je N. Pokud měříme hodnotu zvoleného číselného znaku X, tak naměřené hodnoty označujeme $x_1, x_2, \ldots, x_n$. Jejich průměr $\overline{x}$ značíme μ. Populační rozptyl značíme σ^2.
9. *Výběrový soubor (výběr)* je výběr z populace, aby dobře reprezentoval populaci. Jeho velikost se značí n.
10. Abychom vybrali náhodný výběrový soubor, použijeme *náhodný výběr bez vracení*. Každý takový soubor má pravděpodobnost $\frac{1}{\binom{N}{n}}$.
11. *Výběrový průměr* n náhodných veličin $X_1, \ldots, X_n$ definujeme

$$\overline{X} = \frac{1}{n} \sum_{j=1}^{n} X_j.$$

12. Pro výběrový průměr spočítaný z prvků náhodného výběru z konečné populace lze dokázat

$$E\overline{X} = \mu \qquad \mathrm{var}\overline{X} = \frac{N-n}{N-1} \cdot \frac{\sigma^2}{n}.$$

Výraz $\frac{N-n}{N-1}$ se nazývá *konečnostní násobitel* a pro $n << N$ je vliv konečnostního násobitele zanedbatelný a $\mathrm{var}\overline{X} = \frac{\sigma^2}{n}$. Zřejmě pro $n = 1$ platí $EX = \mu$ a $\mathrm{var}X = \sigma^2$.
13. V případě výběru z konečné populace je střední hodnota výběrového průměru $\overline{X}$ rovna populačnímu průměru μ, tedy odhadovanému parametru. Uvedenou vlastnost formulujeme, že $\overline{X}$ je *nestranným odhadem* parametru μ.
14. Nebo můžeme výběrový soubor vybírat s vracením. Naměřené hodnoty $X_1, \ldots, X_n$ na takto náhodně vybraných prvcích jsou *nezávislé* náhodné veličiny. Při výběru bez vracení by obecně nezávislé nebyly, ale pro dostatečně velikou populaci (nekonečnou) ano.
15. Pro výběrový průměr spočítaný z náhodného výběru rozsahu n z rozdělení s konečnou střední hodnotou a konečným rozptylem platí:

$$E\overline{X} = \mu \qquad \mathrm{var}\overline{X} = \frac{\sigma^2}{n}.$$

16. Jako odhad rozptylu σ^2 se používá výběrový rozptyl

$$S^2 = \frac{1}{n-1} \sum_{i=1}^{N} (X_i - \overline{X})^2.$$

17. Pro náhodný výběr rozsahu n platí $ES^2 = \sigma^2$.

18. Je-li $X_1, \ldots, X_n$ náhodný výběr z rozdělení $N(\mu, \sigma^2)$, potom $\overline{X}$ a S^2 jsou nezávislé veličiny a platí
$$\frac{(N-1)S^2}{\sigma^2} = \frac{1}{\sigma^2}\sum_{i=1}^{n}(X_i - \overline{X})^2 \sim \chi^2(n-1).$$
19. Mějme náhodnou veličinu T dánu podílem nezávislých náhodných veličin
$$T = \frac{Z}{\sqrt{\frac{X}{n}}}.$$
Kde $Z \sim N(0,1)$ a $X \sim \chi^2(n)$. Potom hustota *Studentova t-rozdělení* s n stupni volnosti je $F_T(t)$ a je to nějaký hnusný vzorec.
20. Můžeme nahlédnout, že náhodná veličina
$$T = \frac{\overline{X} - \mu}{S}\sqrt{n}$$
má rozdělení $t(n-1)$.

Výběr z normálního rozdělení se známou střední hodnotou

Předpokládejme, že náhodné veličiny $X_1, \ldots, X_n$ tvoří náhodný výběr rozsahu n z normálního rozdělení $N(\mu, \sigma^2)$ (tedy celá populace se řídí tímto normálním rozdělením). Předpokládejme navíc, že známe rozptyl σ^2. Jakou informaci o hodnotě μ můžeme získat z tohoto náhodného výběru?

Již víme, že v tomto případě má výběrový průměr normální rozdělení $N(\mu, \frac{\sigma^2}{n})$. Je tedy nestranným odhadem parametru μ (neboť náš výběrový průměr má též střední hodnotu μ).

Provedeme-li normování této náhodné veličiny, dostaneme náhodnou veličinu s normovaným náhodným rozdělením $N(0,1)$. Ta bude vypadat
$$\frac{\overline{X} - \mu}{\sqrt{\frac{\sigma^2}{n}}}.$$
(Pomocí jednoduchých vztahů lze dokázat, že bude mít střední hodnotu nulovou a rozptyl roven 1). Pro libovolné $\alpha \in (0,1)$ pak bude platit:

$$\begin{aligned} 1 - \alpha &= \mathbf{P}(|Z| < z(\alpha/2)) && \text{Plyne z CLV.} \quad (1) \\ &= \mathbf{P}\Big(\Big|\tfrac{\overline{X}-\mu}{\sqrt{\frac{\sigma^2}{n}}}\Big| < z(\alpha/2)\Big) && \text{Plyne z pozorování.} \quad (2) \\ &= \mathbf{P}\Big(\overline{X} - \tfrac{\sigma}{\sqrt{n}}z(\alpha/2) < \mu < \overline{X} + \tfrac{\sigma}{\sqrt{n}}z(\alpha/2)\Big) && (2) \end{aligned}$$

Našli jsme interval s náhodnými konci, který s předem danou pravděpodobností pokrývá neznámý parametr μ. Říkáme, že
$$\left(\overline{X} - \frac{\sigma}{\sqrt{n}}z(\alpha/2) < \mu < \overline{X} + \frac{\sigma}{\sqrt{n}}z(\alpha/2)\right)$$
je *interval spolehlivosti* (neboli *konfidenční interval*) pro parametr μ s *koeficientem spolehlivosti* 1 - α.