Pravděpodobnost

Klasická pravděpodobnost. Pravděpodobnost $P(A)$ jevu A v konečném základním prostoru $\Omega = \{\omega_1, \dots, \omega_n\}$ je $P(A) = \frac{m}{n}$, kde m je počet elementárních jevů příznivých jevu A.

Některé vlastnosti axiomatické pravděpodobnosti:

(i) $0 \le P(A) \le 1$,

(ii) $P(\Omega) = 1$, $P(\varnothing) = 0$,

(iii) pro neslučitelné jevy $A_1, A_2, \dots$ platí

$$P(A_1 \cup \dots \cup A_n) = P(A_1) + \dots + P(A_n),$$

(iv) $P(A) = 1 - P(\overline{A})$,

(v) $P(A \cup B) = P(A) + P(B) - P(A \cap B)$,

(vi) $P(A) \le P(B)$ pro $A \subseteq B$,

(vii) $P(B - A) = P(B) - P(A)$ pro $A \subseteq B$,

Podmíněná pravděpodobnost. Pravděpodobnost jevu A za podmínky, že nastal jev B, je $P(A|B) = \frac{P(A\cap B)}{P(B)}$, pokud $P(B) \neq 0$.

Pravděpodobnost průniku. Platí $P(A_1 \cap A_2 \cap \dots \cap A_m) = P(A_1)P(A_2|A_1)P(A_3|A_1 \cap A_2) \dots P(A_m|A_1 \cap \dots \cap A_{m-1})$.

Nezávislost náhodných jevů. Náhodné jevy $A_1, \dots, A_m$ jsou vzájemně nezávislé, pokud platí

$$\begin{aligned} P(A_i \cap A_j) &= P(A_i)P(A_j), \text{ pro } i \neq j, \qquad \text{(párově nezávislé)} \\ P(A_i \cap A_j \cap A_k) &= P(A_i)P(A_j)P(A_k), \text{ pro } i \neq j, j \neq k, \\ &\vdots \\ P(A_1 \cap A_2 \cap \dots \cap A_m) &= P(A_1)P(A_2)\dots P(A_m). \end{aligned}$$

Úplná pravděpodobnost. Pokud $B_1, \dots, B_n$ je úplný systém neslučitelných jevů a $P(B_i) \neq 0$ pro všechna $i = 1, \dots, n$, pak pro libovolný jev A platí

$$P(A) = \sum_{i=1}^{n} P(B_i)P(A|B_i).$$

Bayesův vzorec. Pokud $B_1, \dots, B_n$ je úplný systém neslučitelných jevů, $P(B_i) \neq 0$ pro všechna $i = 1, \dots, n$ a $P(A) > 0$, pak

$$P(B_i|A) = \frac{P(B_i)P(A|B_i)}{\sum\limits_{j=1}^{n} P(B_j)P(A|B_j)}$$

Náhodné veličiny

Funkční charakteristiky.

Distribuční funkce	$F(x) = P(X \le x)$.	$0 \le F(x) \le 1$, F je neklesající a zprava spojitá, $\lim_{x\to\infty} F(x) = 1$, $\lim_{x\to-\infty} F(x) = 0$, $P(a < X \le b) = F(b) - F(a)$, $P(X = a) = F(a) - \lim_{x\to a_-} F(x)$.
Pravděpodobnostní funkce	$p(x) = P(X = x)$.	$\sum_{x\in M} p(x) = 1$, $F(x) = \sum_{t\le x, t\in M} p(t)$.
Hustota pravděpodobnosti	$f(x):\ F(x) = \int_{-\infty}^{x} f(t)\,\mathrm{d}t$.	$\int_{-\infty}^{\infty} f(x)\,\mathrm{d}x = 1$, $f(x) = F'(x)$, $P(a \le X \le b) = \int_a^b f(x)\,\mathrm{d}x$.

Číselné charakteristiky.

Střední hodnota	$\mathrm{E}(X) = \begin{cases} \sum\limits_M x \cdot p(x), & \text{pro } X \text{ diskrétní}, \\ \int\limits_{-\infty}^{\infty} x \cdot f(x)\,\mathrm{d}x, & \text{pro } X \text{ spojitou}. \end{cases}$	$\mathrm{E}(X \pm Y) = \mathrm{E}(X) \pm \mathrm{E}(Y)$, $\mathrm{E}(a + bX) = a + b \cdot \mathrm{E}(X)$, $\mathrm{E}(X \cdot Y) = \mathrm{E}(X) \cdot \mathrm{E}(Y)$ pro X, Y nezávislé.
Rozptyl a směrodatná odchylka	$\mathrm{D}(X) = \mathrm{E}\left[(X - \mathrm{E}(X))^2\right]$, $\sigma(X) = \sqrt{\mathrm{D}(X)}$.	$\mathrm{D}(X) \ge 0$, $\mathrm{D}(a + bX) = b^2 \cdot \mathrm{D}(X)$, $\mathrm{D}(X) = \mathrm{E}(X^2) - \mathrm{E}^2(X)$, $\mathrm{D}(X + Y) = \mathrm{D}(X) + \mathrm{D}(Y)$ pro X, Y nezávislé.
α-kvantil x_α (spojité náh. vel.)	$F(x_\alpha) = \alpha$ pro $F(x)$ rostoucí, případně $\int\limits_{-\infty}^{x_\alpha} f(x)\,\mathrm{d}x = \alpha$.	

Vybraná rozdělení pravděpodobnosti

Binomické rozdělení. $X \sim \mathrm{Bi}(n, \pi), \quad n \in \mathbb{N}, 0 < \pi < 1$

$$p(x) = \binom{n}{x} \pi^x (1-\pi)^{n-x}, \quad x = 0, 1, \ldots, n.$$

Aproximace:
(a) Pro $n > 30$ a $\pi < \frac{1}{10}$ lze použít $\mathrm{Bi}(n,\pi) \approx \mathrm{Po}(n\pi)$.
(b) Pro $n\pi(1-\pi) > 9$ lze použít $\mathrm{Bi}(n,\pi) \approx \mathrm{N}(n\pi, n\pi(1-\pi))$.

$\mathrm{E}(X) = n\pi$,
$\mathrm{D}(X) = n\pi(1-\pi)$,
$(n+1)\pi - 1 \leq \mathrm{Mod}(X) \leq (n+1)\pi$.

Hypergeometrické rozdělení. $X \sim \mathrm{Hg}(N, M, n), \quad N, M, n \in \mathbb{N}$

$$p(x) = \frac{\binom{M}{x}\binom{N-M}{n-x}}{\binom{N}{n}}, \quad \max\{0, n-N+M\} \leq x \leq \min\{n, M\}, x \in \mathbb{N}_0.$$

$\mathrm{E}(X) = n\pi, \quad \text{kde } \pi = \frac{M}{N}$,
$\mathrm{D}(X) = n\pi(1-\pi)\frac{N-n}{N-1}$,
$a - 1 \leq \mathrm{Mod}(X) \leq a, \quad \text{kde } a = \frac{(M+1)(n+1)}{N+2}$.

Poissonovo rozdělení. $X \sim \mathrm{Po}(\lambda), \quad \lambda > 0$

$$p(x) = \frac{\lambda^x}{x!} e^{-\lambda}, \quad x = 0, 1, 2, \ldots.$$

$\mathrm{E}(X) = \mathrm{D}(X) = \lambda$,
$\lambda - 1 \leq \mathrm{Mod}(X) \leq \lambda$.

Rovnoměrné spojité rozdělení. $X \sim \mathrm{Ro}(a, b), \quad a < b, a, b \in \mathbb{R}$

$$f(x) = \frac{1}{b-a}, \quad a \leq x \leq b.$$

$$F(x) = \begin{cases} 0, & x < a, \\ \frac{x-a}{b-a}, & a \leq x < b, \\ 1, & x \geq b. \end{cases}$$

$\mathrm{E}(X) = x_{0.5} = \frac{a+b}{2}$,
$\mathrm{D}(X) = \frac{(b-a)^2}{12}$,
$\mathrm{Mod}(X) = \langle a, b \rangle$.

Normální rozdělení. $X \sim \mathrm{N}(\mu, \sigma^2), \quad \mu \in \mathbb{R}, \sigma^2 > 0$

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

$\mathrm{E}(X) = x_{0.5} = \mathrm{Mod}(X) = \mu$,
$\mathrm{D}(X) = \sigma^2$.

Platí:

$$F(x) = \Phi\left(\frac{x-\mu}{\sigma}\right), \qquad f(x) = \frac{1}{\sigma}\varphi\left(\frac{x-\mu}{\sigma}\right),$$

$$\Phi(-x) = 1 - \Phi(x), \qquad \varphi(-x) = \varphi(x)$$

Dvourozměrný diskrétní náhodný vektor

Diskrétní vektor $\boldsymbol{X} = (X, Y)$ s oborem hodnot $M = M_X \times M_Y \subseteq \mathbb{R}^2$ je popsán pravděpodobnostní funkcí $p(x, y)$, pro kterou platí $p(x, y) \geq 0$, $p(x, y) > 0$ na M a dále

$$\sum_{M_X}\sum_{M_Y} p(x, y) = 1.$$

Marginální pravděpodobnostní funkce náhodných veličin X a Y se dostanou

$$\begin{aligned} X: \quad p_X(x) &= \sum_{y \in M_Y} p(x, y), \\ Y: \quad p_Y(y) &= \sum_{x \in M_X} p(x, y). \end{aligned}$$

Náhodné veličiny X, Y jsou stochasticky nezávislé, pokud platí $p_X(x) \cdot p_Y(y) = p(x, y)$ pro všechna $(x, y) \in R^2$.

Kovariance

$$\mathrm{C}(X, Y) = \mathrm{E}[(X - \mathrm{E}(X))(Y - \mathrm{E}(Y))],$$

$\mathrm{C}(X, X) = \mathrm{D}(X)$,
$\mathrm{C}(X, Y) = \mathrm{E}(XY) - \mathrm{E}(X) \cdot \mathrm{E}(Y)$,
$\mathrm{D}(aX + bY) = a^2\,\mathrm{D}(X) + b^2\,\mathrm{D}(Y) + 2ab\,\mathrm{C}(X, Y)$,
$\mathrm{C}(X, Y) = 0$ pro X, Y stochasticky nezávislé.

Korelační koeficient

$$\rho(X, Y) = \frac{\mathrm{C}(X, Y)}{\sqrt{\mathrm{D}(X)}\sqrt{\mathrm{D}(Y)}},$$

$-1 \leq \rho(X, Y) \leq 1$,
$\rho(X, X) = 1$,
$\rho(X, Y) = \pm 1 \Rightarrow X, Y$ jsou lineárně závislé,
$\rho(X, Y) = 0$ pro X, Y stochasticky nezávislé.

Bodové a intervalové odhady parametrů normálního a alternativního rozdělení

výběrový průměr	$\overline{X} = \sum_{i=1}^{n} X_i,$	$\overline{X}$ je nestranným konzistentním odhadem střední hodnoty,
výběrový rozptyl	$S^2 = \frac{1}{n-1}\sum_{i=1}^{n}(X_i - \overline{X})^2$	S^2 je nestranným konzistentním odhadem rozptylu,
	$S_n^2 = \frac{1}{n}\sum_{i=1}^{n}(X_i - \overline{X})^2$	S_n^2 je konzistentním odhadem rozptylu, $S_n^2 = \overline{X^2} - \overline{X}^2$.
výběrová směrodatná odchylka	$S = \sqrt{S^2}, \quad S_n = \sqrt{S_n^2}.$	Pro náhodný výběr z $\mathrm{N}(\mu, \sigma^2)$ jsou odhady $\overline{X}, S^2$ navíc nejlepší, pro náhodný výběr z $\mathrm{A}(\pi)$ je $\overline{X}$ nejlepší také.

Intervalové odhady parametrů normálního rozdělení. Nechť $X_1, \ldots, X_n$ je náhodný výběr z $\mathrm{N}(\mu, \sigma^2)$ a $\alpha \in (0,1)$ je dané číslo. Symboly $t_\alpha(n)$, $\chi^2_\alpha(n)$ značí α-kvantily Studentova $t(n)$ a Pearsonova $\chi^2(n)$ rozdělení s n stupni volnosti.

parametr	**100(1 − α)% interval spolehlivosti**		
	pravostranný	oboustranný	levostranný
μ (σ^2 známý)	$\left(-\infty, \overline{X} + u_{1-\alpha}\cdot\frac{\sigma}{\sqrt{n}}\right\rangle$	$\left\langle\overline{X} - u_{1-\frac{\alpha}{2}}\cdot\frac{\sigma}{\sqrt{n}}, \overline{X} + u_{1-\frac{\alpha}{2}}\cdot\frac{\sigma}{\sqrt{n}}\right\rangle$	$\left\langle\overline{X} - u_{1-\alpha}\cdot\frac{\sigma}{\sqrt{n}}, \infty\right)$
μ (σ^2 neznámý)	$\left(-\infty, \overline{X} + t_{1-\alpha}(n-1)\cdot\frac{S}{\sqrt{n}}\right\rangle$	$\left\langle\overline{X} - t_{1-\frac{\alpha}{2}}(n-1)\cdot\frac{S}{\sqrt{n}}, \overline{X} + t_{1-\frac{\alpha}{2}}(n-1)\cdot\frac{S}{\sqrt{n}}\right\rangle$	$\left\langle\overline{X} - t_{1-\alpha}(n-1)\cdot\frac{S}{\sqrt{n}}, \infty\right)$
σ^2 (μ neznámý)	$\left\langle 0, \frac{(n-1)\cdot S^2}{\chi^2_\alpha(n-1)}\right\rangle$	$\left\langle\frac{(n-1)\cdot S^2}{\chi^2_{1-\frac{\alpha}{2}}(n-1)}, \frac{(n-1)\cdot S^2}{\chi^2_{\frac{\alpha}{2}}(n-1)}\right\rangle$	$\left\langle\frac{(n-1)\cdot S^2}{\chi^2_{1-\alpha}(n-1)}, \infty\right)$

Intervalový odhad parametru alternativního rozdělení. Nechť $X_1, \ldots, X_n$ je náhodný výběr z rozdělení $\mathrm{A}(\pi)$ a $\alpha \in (0,1)$ dané číslo. Označme $p = \overline{X}$ bodový odhad parametru π.

parametr	**100(1 − α)% interval spolehlivosti pro n > 30**		
	pravostranný	oboustranný	levostranný
π	$\left\langle 0, p + u_{1-\alpha}\cdot\sqrt{\frac{p\cdot(1-p)}{n}}\right\rangle$	$\left\langle p - u_{1-\frac{\alpha}{2}}\cdot\sqrt{\frac{p\cdot(1-p)}{n}}, p + u_{1-\frac{\alpha}{2}}\cdot\sqrt{\frac{p\cdot(1-p)}{n}}\right\rangle$	$\left\langle p - u_{1-\alpha}\cdot\sqrt{\frac{p\cdot(1-p)}{n}}, 1\right\rangle$

Testy statistických hypotéz

Jednovýběrové testy hypotéz o parametrech.

Tabulka 1: Testy hypotéz o parametrech normálního rozdělení na hladině významnosti α.

Test $\mathbf{H_0}$ proti $\mathbf{H}$	**Testové kritérium**	**Kritický obor**
nulová hypotéza $\mathrm{H}_0: \quad \mu = \mu_0$ alternativa $\mathrm{H}_1: \quad \mu \neq \mu_0$ $\mathrm{H}_2: \quad \mu > \mu_0$ $\mathrm{H}_3: \quad \mu < \mu_0$	$T = \frac{\overline{X} - \mu_0}{\sigma}\cdot\sqrt{n}$ pro σ^2 známé	$W_1 = \left(-\infty, -u_{1-\frac{\alpha}{2}}\right) \cup \left(u_{1-\frac{\alpha}{2}}, \infty\right)$ $W_2 = (u_{1-\alpha}, \infty)$ $W_3 = (-\infty, -u_{1-\alpha})$
	$T = \frac{\overline{X} - \mu_0}{S}\cdot\sqrt{n}$ pro σ^2 neznámé	$W_1 = \left(-\infty, -t_{1-\frac{\alpha}{2}}(n-1)\right) \cup \left(t_{1-\frac{\alpha}{2}}(n-1), \infty\right)$ $W_2 = (t_{1-\alpha}(n-1), \infty)$ $W_3 = (-\infty, -t_{1-\alpha}(n-1))$
nulová hypotéza $\mathrm{H}_0: \quad \sigma^2 = \sigma_0^2$ alternativa $\mathrm{H}_1: \quad \sigma^2 \neq \sigma_0^2$ $\mathrm{H}_2: \quad \sigma^2 > \sigma_0^2$ $\mathrm{H}_3: \quad \sigma^2 < \sigma_0^2$	$T = \frac{(n-1)\cdot S^2}{\sigma_0^2}$ pro μ neznámé	$W_1 = \left\langle 0, \chi^2_{\frac{\alpha}{2}}(n-1)\right) \cup \left(\chi^2_{1-\frac{\alpha}{2}}(n-1), \infty\right)$ $W_2 = \left(\chi^2_{1-\alpha}(n-1), \infty\right)$ $W_3 = \left\langle 0, \chi^2_\alpha(n-1)\right)$

Tabulka 2: Test hypotézy o parametru π alternativního rozdělení na hladině významnosti α. Statistika $p = \overline{X}$ značí nestranný odhad parametru.

Test $\mathbf{H_0}$ proti $\mathbf{H}$	**Testové kritérium**	**Kritický obor**
nulová hypotéza $\mathrm{H}_0: \quad \pi = \pi_0$ alternativa $\mathrm{H}_1: \quad \pi \neq \pi_0$ $\mathrm{H}_2: \quad \pi > \pi_0$ $\mathrm{H}_3: \quad \pi < \pi_0$	$T = \frac{p - \pi_0}{\sqrt{\pi_0\cdot(1-\pi_0)}}\cdot\sqrt{n}$	$W_1 = \left(-\infty, -u_{1-\frac{\alpha}{2}}\right) \cup \left(u_{1-\frac{\alpha}{2}}, \infty\right)$ $W_2 = (u_{1-\alpha}, \infty)$ $W_3 = (-\infty, -u_{1-\alpha})$

Dvouvýběrové testy hypotéz o parametrech.

Tabulka 3: Test shody rozptylů dvou normálních rozdělení na hladině významnosti α. Statistiky S_x^2, S_y^2 značí nestranné odhady rozptylů příslušných náhodných výběrů.

Test $\mathbf{H_0}$ proti $\mathbf{H}$	**Testové kritérium**	**Kritický obor**
nulová hypotéza $H_0: \quad \sigma_x^2 = \sigma_y^2$ alternativa $H_1: \quad \sigma_x^2 \neq \sigma_y^2$ $H_2: \quad \sigma_x^2 > \sigma_y^2$ $H_3: \quad \sigma_x^2 < \sigma_y^2$	$F = \dfrac{S_x^2}{S_y^2}$	$W_1 = \left\langle 0, F_{\frac{\alpha}{2}}(n_x-1, n_y-1)\right) \cup \left(F_{1-\frac{\alpha}{2}}(n_x-1, n_y-1), \infty\right)$ $W_2 = (F_{1-\alpha}(n_x-1, n_y-1), \infty)$ $W_3 = \langle 0, F_{\alpha}(n_x-1, n_y-1))$

Tabulka 4: Test shody středních hodnot dvou normálních rozdělení na hladině významnosti α. Statistiky S_x^2, S_y^2, S_d^2 značí nestranné odhady rozptylů příslušných náhodných výběrů.

Test $\mathbf{H_0}$ proti $\mathbf{H}$	**Testové kritérium**	**Kritický obor**
nulová hypotéza $H_0: \quad \mu_x = \mu_y$ alternativa $H_1: \quad \mu_x \neq \mu_y$ $H_2: \quad \mu_x > \mu_y$ $H_3: \quad \mu_x < \mu_y$	pro $\sigma_x^2 = \sigma_y^2$: $T = \dfrac{\overline{X} - \overline{Y}}{S}\sqrt{\dfrac{n_x \cdot n_y}{n_x + n_y}},$ kde $S = \sqrt{\frac{(n_x-1)S_x^2 + (n_y-1)S_y^2}{n_x+n_y-2}}$	$W_1 = \left(-\infty, -t_{1-\frac{\alpha}{2}}(\nu)\right) \cup \left(t_{1-\frac{\alpha}{2}}(\nu), \infty\right)$ $W_2 = (t_{1-\alpha}(\nu), \infty)$ $W_3 = (-\infty, -t_{1-\alpha}(\nu))$ $\nu = n_x + n_y - 2$
	pro $\sigma_x^2 \neq \sigma_y^2$: $T = \dfrac{\overline{X} - \overline{Y}}{\sqrt{\frac{S_x^2}{n_x} + \frac{S_y^2}{n_y}}}$	$W_1 = \left(-\infty, -t_{1-\frac{\alpha}{2}}(\nu)\right) \cup \left(t_{1-\frac{\alpha}{2}}(\nu), \infty\right)$ $W_2 = (t_{1-\alpha}(\nu), \infty)$ $W_3 = (-\infty, -t_{1-\alpha}(\nu))$ $\nu \approx \dfrac{\left(\frac{S_x^2}{n_x} + \frac{S_y^2}{n_y}\right)^2}{\frac{1}{n_x-1}\left(\frac{S_x^2}{n_x}\right)^2 + \frac{1}{n_y-1}\left(\frac{S_y^2}{n_y}\right)^2}$
	pro závislé výběry: $T = \dfrac{\overline{D}}{\sqrt{S_d^2}} \cdot \sqrt{n}$ kde $D_i = X_i - Y_i$	$W_1 = \left(-\infty, -t_{1-\frac{\alpha}{2}}(n-1)\right) \cup \left(t_{1-\frac{\alpha}{2}}(n-1), \infty\right)$ $W_2 = (t_{1-\alpha}(n-1), \infty)$ $W_3 = (-\infty, -t_{1-\alpha}(n-1))$

Tabulka 5: Test shody parametrů π_x, π_y dvou alternativních rozdělení na hladině významnosti α. Statistiky $p_x = \overline{X}$, $p_y = \overline{Y}$ značí nestranné odhady parametrů příslušných náhodných výběrů.

Test $\mathbf{H_0}$ proti $\mathbf{H}$	**Testové kritérium**	**Kritický obor**
nulová hypotéza $H_0: \quad \pi_x = \pi_y$ alternativa $H_1: \quad \pi_x \neq \pi_y$ $H_2: \quad \pi_x > \pi_y$ $H_3: \quad \pi_x < \pi_y$	$U = \dfrac{p_x - p_y}{\sqrt{\frac{p_x \cdot (1-p_x)}{n_x} + \frac{p_y \cdot (1-p_y)}{n_y}}}$	$W_1 = \left(-\infty, -u_{1-\frac{\alpha}{2}}\right) \cup \left(u_{1-\frac{\alpha}{2}}, \infty\right)$ $W_2 = (u_{1-\alpha}, \infty)$ $W_3 = (-\infty, -u_{1-\alpha})$

Pearsonův χ^2 test dobré shody. $H_0: X \sim F(x; \widehat{\boldsymbol{\theta}})$ proti $H: X \nsim F(x; \boldsymbol{\theta})$ pro libovolné $\boldsymbol{\theta} = (\theta_1, \dots, \theta_m) \in \boldsymbol{\Theta}$.
Testová statistika

$$T = \sum_{i=1}^{k} \frac{(n_i - np_i)^2}{np_i} \overset{\cdot}{\sim} \chi^2(k-m-1),$$

kritický obor $W = (\chi^2_{1-\alpha}(k-m-1), \infty)$.

Podmínky použitelnosti:

(i) $n \cdot p_i \geq 1$ pro všechny třídy,

(ii) $n \cdot p_i \geq 5$ pro 80 % tříd.