

Statistika pro informatiku

Souhrn látky

únor 2014

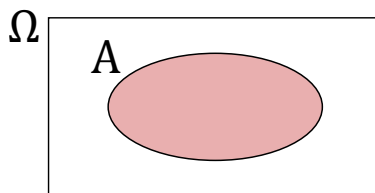
Obsah

1	Základy statistiky a pravděpodobnosti	2
1.1	Pravděpodobnost jevu a jeho doplňku	2
1.2	Sjednocení jevů	2
1.3	Průnik jevů	3
1.4	Nezávislost jevů	3
1.5	Podmíněná pravděpodobnost	4
1.6	Bayesova věta	4
1.7	Shrnutí	4
2	Vlastnosti	5
2.1	Střední hodnota EX	5
2.2	Rozptyl	5
2.3	Distribuční funkce	6
2.4	Hustota	6
2.5	Kovariance	6
2.6	Korelační koeficient	6
3	Rozdělení pravděpodobnosti	6
3.1	Diskrétní (nespojité) rozdělení	6
3.2	Spojité rozdělení	7
4	Entropie	7
4.1	Sdružená entropie	7
4.2	Podmíněná entropie	8
4.3	Vzájemná informace	8
4.4	Kódování	8
4.5	Logaritmus	9
5	Náhodné procesy	9
6	Markovovy řetězce	9
6.1	Stacionární distribuce	10
7	Systémy hromadné obsluhy	10
8	Ostatní	10
8.1	Řady	10

1 Základy statistiky a pravděpodobnosti

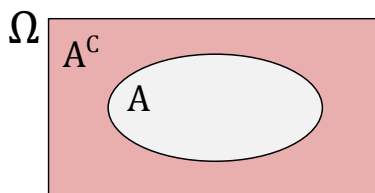
1.1 Pravděpodobnost jevu a jeho doplňku

$$\mathbb{P}(A) = \frac{\text{size}(A)}{\text{size}(\Omega)}$$



Obrázek 1: Vennův diagram základní pravděpodobnosti jevu

$$\mathbb{P}(A^C) = 1 - \mathbb{P}(A)$$

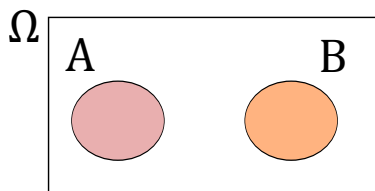


Obrázek 2: Vennův diagram doplňku jevu

1.2 Sjednocení jevů

Pro disjunktní jevy platí

$$\mathbb{P}(A \cup B) = \mathbb{P}(A) + \mathbb{P}(B).$$

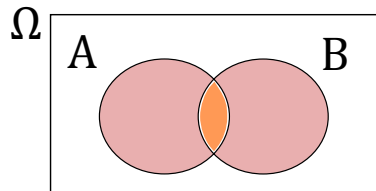


Obrázek 3: Dva disjunktní jevy

Jinak obecně platí (pro nedisjunktní jevy):

$$\mathbb{P}(A \cup B) = \mathbb{P}(A) + \mathbb{P}(B) - \mathbb{P}(A \cap B).$$

Oblast průniku by byla započítána dvakrát, proto je potřeba ji odečíst.



Obrázek 4: Sjednocení nedisjunktních jevů

1.3 Průnik jevů

$$\begin{aligned}\mathbb{P}(A \cap B) &= \mathbb{P}(A|B) \mathbb{P}(B) \\ \mathbb{P}(A \cap B) &= \mathbb{P}(B|A) \mathbb{P}(A) \\ \mathbb{P}(A \cap B \cap C \dots) &= \mathbb{P}(A) \mathbb{P}(B|A) \mathbb{P}(C|A \cap B) \dots\end{aligned}$$

Obecně zapsáno:

$$\mathbb{P}(\text{intersection}) = \mathbb{P}(\text{event}|\text{condition}) * \mathbb{P}(\text{condition})$$

1.4 Nezávislost jevů

U nezávislých jevů platí

$$\begin{aligned}\mathbb{P}(A|B) &= \mathbb{P}(A) \\ \mathbb{P}(B|A) &= \mathbb{P}(B),\end{aligned}$$

a proto tedy:

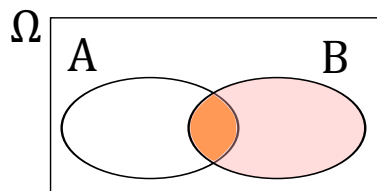
$$\boxed{\mathbb{P}(A \cap B) = \mathbb{P}(A) * \mathbb{P}(B)}.$$

1.5 Podmíněná pravděpodobnost

$$\mathbb{P}(A|B) = \frac{\mathbb{P}(A \cap B)}{\mathbb{P}(B)}, \mathbb{P}(B) \neq 0$$

„Pravděpodobnost jevu A za podmínky, že jsme v B a že jev B nastal.“

$$\mathbb{P}(A|B) = \mathbb{P}(B|A) * \mathbb{P}(A)$$

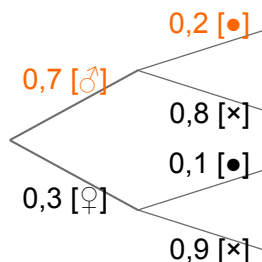


Obrázek 5: Podmíněná pravděpodobnost

1.6 Bayessova věta

$$\mathbb{P}(A|B) = \frac{\mathbb{P}(B|A) \mathbb{P}(A)}{\mathbb{P}(B)}$$

$$\mathbb{P}(\sigma \cap \bullet) = \mathbb{P}(\bullet|\sigma) * \mathbb{P}(\sigma) = \mathbb{P}(\sigma) * \mathbb{P}(\bullet|\sigma) = 0,7 * 0,2 = \underline{\underline{0,14}}$$



Obrázek 6: Bayessova věta pomocí stromu

1.7 Shrnutí

Jev	Sjednocení (\cup)	Průnik (\cap)
Disjunktní jevy	$\mathbb{P}(A \cup B) = \mathbb{P}(A) + \mathbb{P}(B)$	$\mathbb{P}(A \cap B) = \emptyset$
Nedisjunktní jevy	$\mathbb{P}(A \cup B) = \mathbb{P}(A) + \mathbb{P}(B) - \mathbb{P}(A \cap B)$	$\mathbb{P}(A \cap B) = \mathbb{P}(A) + \mathbb{P}(B) - \mathbb{P}(A \cup B)$
Závislé jevy	$\mathbb{P}(A \cup B) = \mathbb{P}(A) + \mathbb{P}(B) - \mathbb{P}(A \cap B)$	$\mathbb{P}(A \cap B) = \mathbb{P}(A B) * \mathbb{P}(B)$
Nezávislé jevy	$\mathbb{P}(A \cup B) = \mathbb{P}(A) + \mathbb{P}(B) - \mathbb{P}(A \cap B)$	$\mathbb{P}(A \cap B) = \mathbb{P}(A) * \mathbb{P}(B)$

Tabulka 1: Shrnutí operací nad různými jevy

2 Vlastnosti

2.1 Střední hodnota $\mathbb{E}X$

Pro diskrétní veličiny

$$\mathbb{E}X = \sum_i p_i x_i = \sum_i x_i * \mathbb{P}(X = x_i)$$

Pro spojitě veličiny

$$\mathbb{E}X = \int_{-\infty}^{+\infty} x * f_x(x) dx$$

(P a f jsou funkce hustoty.)

Pro libovolné náhodné veličiny platí:

$$\begin{aligned}\mathbb{E}(aX + Y) &= a\mathbb{E}(X) + \mathbb{E}(Y) \quad (\text{linearita}) \\ \mathbb{E}(X \pm Y) &= \mathbb{E}(X) \pm \mathbb{E}(Y) \\ \mathbb{E}(X + Y) &= \mathbb{E}(\max\{X, Y\}) + \mathbb{E}(\min\{X, Y\}) \\ \mathbb{E}X^2 &= \sum_i p_i x_i^2 \quad (\text{pro diskrétní jevy}) \\ \mathbb{E}(\max\{X, Y\}) &= \mathbb{E}(X) + \mathbb{E}(Y) - \mathbb{E}(\min\{X, Y\}) \\ \mathbb{E}(XY) &= \mathbb{E}X * \mathbb{E}Y \quad (\text{platí jen pro nezávislé jevy})\end{aligned}$$

2.2 Rozptyl

Pro diskrétní náhodnou veličinu jej můžeme definovat vztahem:

$$\sigma^2 = \sum_{i=1}^n [x_i - E(X)]^2 p_i = \sum_{i=1}^n x_i^2 p_i - [E(X)]^2$$

Pro spojitou náhodnou veličinu definujeme rozptyl vztahem:

$$\sigma^2 = \int_{-\infty}^{+\infty} [x_i - E(X)]^2 f(x) dx = \int_{-\infty}^{+\infty} x_i^2 f(x) dx - [E(X)]^2$$

Dále platí:

$$\begin{aligned}\text{var}(X) &= E[(X - E(X))^2] = E(X^2) - (E(X))^2 \\ \text{var}(X) &= \text{cov}(X, X) \\ \text{var}(aX) &= a^2 \text{var}(X) \\ \text{var}(X + a) &= \text{var}(X) \\ \text{var}(X \pm Y) &= \text{var}(X) + \text{var}(Y) \pm 2\text{cov}(X, Y)\end{aligned}$$

2.3 Distribuční funkce

- Funkce je zprava spojitá.

Distribuční funkce pro diskrétní veličiny

$$F = \mathbb{P}(X \leq x_i) = \sum_{x_i \leq x} p_x(x_i)$$

Distribuční funkce pro spojité veličiny

$$F = \mathbb{P}(X \leq x_i) = \int_{-\infty}^x f_x(u) du \quad \forall x \in \mathbb{R}$$

(X je náhodná veličina, x_i je číslo)

2.4 Hustota

Funkce hustoty pro diskrétní veličiny

$$p(X) = \mathbb{P}(X = x)$$

Funkce hustoty pro spojité veličiny

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

2.5 Kovariance

$$\text{cov}(X, Y) = E[(X - E(X)) * (Y - E(Y))] = E(XY) - E(X)E(Y)$$

Platí, že pokud

$$\text{cov}(X, Y) = 0$$

pak

$$E(XY) = E(X) * E(Y)$$

a X a Y jsou nezávislé.

2.6 Korelační koeficient

$$\rho(X, Y) = \frac{\text{cov}(X, Y)}{\sigma_x * \sigma_y} = \frac{\text{cov}(X, Y)}{\sqrt{\text{var}(X)} * \sqrt{\text{var}(Y)}}$$

3 Rozdělení pravděpodobnosti

Distribuční funkce	$X \leq k$
Hustota	$X = k$
Funkce přežití	$X > k$

Tabulka 2: Funkce

3.1 Diskrétní (nespojité) rozdělení

Diskrétní veličiny mohou nabývat pouze spočetného počtu hodnot (i nekonečného).

Rozdělení	Funkce hustoty	Distribuční funkce	$\mathbb{E}X$	$varX$
Bernoulliho , $X \sim Be(p)$	$\mathbb{P}(0) = 1 - p, \mathbb{P}(1) = p$	\times	p	$p(1 - p)$
Binomické , $X \sim Bi(n, p)$	$\binom{n}{k} p^k (1 - p)^{n-k}$	$I_{1-p}(n - k, 1 + k)$	$\mathbb{E}X = n * p$	$varX = np(1 - p)$
Geometrické , $X \sim geom(p)$	$(1 - p)^{k-1} * p$	$\mathbb{P}(T \leq n) = 1 - (1 - p)^n$ $\mathbb{P}(T > n) = (1 - p)^n$	$\mathbb{E}X = \frac{1}{p}$	$varX = \frac{1-p}{p^2}$
Poissonovo , $X \sim Pois(\lambda)$	$\frac{\lambda^k}{k!} e^{-\lambda}$	$Q(\lfloor k + 1 \rfloor, \lambda)$	λ	λ

Obrázek 7: Diskrétní rozdělení

3.2 Spojité rozdělení

Spojité náhodné veličiny nabývají na rozdíl od diskrétních veličin nějakého intervalu.

Rozdělení	Funkce hustoty	Distribuční funkce	$\mathbb{E}X$	$varX$
Rovnoměrné , $X \sim unif(a, b)$	$\frac{1}{b-a}; x \in [a, b]$	\times	$\frac{a+b}{2}$	$\frac{(b-a)^2}{12}$
Exponenciální , $X \sim Bi(n, p)$	$\lambda e^{-\lambda x}; x \in [0, +\infty)$	$1 - e^{-\lambda x}$	$\frac{1}{\lambda}$	$\frac{1}{\lambda^2}$
Normální (Gaussovo) , $X \sim geom(p)$	$\frac{1}{\sigma\sqrt{2\pi}} * e^{\frac{-(x-\mu)^2}{2\sigma^2}}$	\times	μ	σ^2

Obrázek 8: Spojité rozdělení

4 Entropie

Entropie diskrétní veličiny

$$H_b(X) = - \sum p_i \log_b p_i$$

Entropie spojité veličiny

$$H_b(X) = - \int_{-\infty}^{+\infty} f(x) \log_b f(x) dx$$

(b je základ abecedy pro kódová slova, nejčastěji používáme binární abecedu, tedy $b = 2$)

Aditivita entropie

$$H(X, Y) = H(X) + H(Y|X)$$

$$H(X, Y) = H(X) + H(Y) \text{ (speciálně jen pro nezávislé náhodné veličiny)}$$

4.1 Sdružená entropie

$$H(X, Y) = - \sum_{i,j} p_{i,j} \log p_{i,j}$$

(sdružená hustota $p_{i,j} = P(X = x_i, Y = Y_j)$)

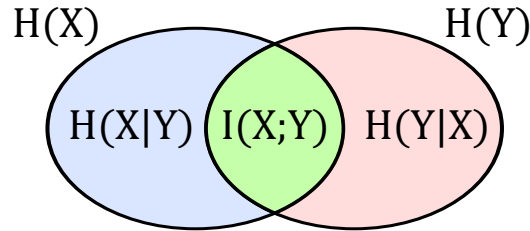
4.2 Podmíněná entropie

$$\begin{aligned}
 H(X|Y) &= - \sum_{i,j} p(x_i, y_j) \log p(y_j|x_i) \\
 \mathbb{P}(x_i|y_j) &= \frac{\mathbb{P}(x_i, y_j)}{\mathbb{P}(y_j)} \quad \begin{array}{l} \text{sdužená hustota} \\ \text{marginální hustota} \end{array} \\
 \mathbb{P}(X, Y) &= \mathbb{P}(X|Y) * \mathbb{P}(Y)
 \end{aligned}$$

4.3 Vzájemná informace

$$\begin{aligned}
 I(X; Y) &= H(X) - H(X|Y) \\
 I(X; Y) &= H(Y) - H(Y|X) \\
 I(X; Y) &= H(X) + H(Y) - H(X, Y) \\
 I(X; Y) &= I(Y; X) \\
 I(X; X) &= H(X)
 \end{aligned}$$

$$I(X, Y) = \sum_{i,j} p_{i,j} \log \frac{p_{i,j}}{p_i * p_j} = \dots = H(X) - H(X|Y)$$



Obrázek 9: Vzájemná informace a entropie (tedy $H(X, Y)$)

4.4 Kódování

Střední délka kódového slova

$$L(C) = \mathbb{E} \ell(X) = \sum_i \ell(x_i) * \mathbb{P}(X = x_i)$$

Kódování je optimální, pokud je entropie menší než střední délka kódového slova.

4.5 Logaritmus

$$\begin{aligned}\log_a (x_1 * x_2) &= \log_a x_1 + \log_a x_2 \\ \log_a \left(\frac{x_1}{x_2} \right) &= \log_a x_1 - \log_a x_2 \\ \log_a x^r &= r * \log_a x \\ \log_a a &= 1 \\ \log_a 1 &= 0\end{aligned}$$

5 Náhodné procesy

Značení procesu

$$X(t, \omega) = X_t = X(t)$$

Střední hodnota

$$\begin{aligned}\eta_x(t) &= \mathbb{E}X(t) = \int x(t) * f_{X_t}(x) dx \\ \mathbb{E}X(t) &= \sum x_i(t) \mathbb{P}(X_t = x_i(t))\end{aligned}$$

Pokud je střední hodnota **nezávislá** na t , **je proces stacionární** v průměru. Tedy

$$\eta(t) = \eta_x \forall t$$

Střední hodnota integrálu

$$\mathbb{E} \int_a^b X(t) dt = \int_a^b \mathbb{E}X(t) dt$$

Autokorelační funkce

$$\begin{aligned}R_{xx}(t_1, t_2) &= \mathbb{E} [X(t_1) * \overline{X(t_2)}] \text{ v } \mathbb{C} \\ &= \mathbb{E} [X(t_1) * X(t_2)] = \sum_i x_i(t) * P(X(t_1) * X(t_2) = x_i(t)) \text{ v } \mathbb{R}\end{aligned}$$

V ukázkovém příkladu se autokorelační funkce spočítala pomocí následujícího „triku“ ($0 \leq t_1 \leq t_2 \leq 2$)

$$\begin{aligned}R_x(t_1, t_2) &= \mathbb{E} [X(t_1) * X(t_2)] = 0 * \mathbb{P}(X(t_1) * X(t_2) = 0) + 1 * \mathbb{P}(X(t_1) * X(t_2) = 1) \\ &= \mathbb{P}(X(t_1) * X(t_2) = 1) = \mathbb{P}(X(t_1) = 1, X(t_2) = 1) = \\ &= \boxed{\mathbb{P}(t_1 \leq A, t_2 \leq A) = \mathbb{P}(A \geq t_1, t_2 \geq A) = P(A \geq t_2)}\end{aligned}$$

6 Markovovy řetězce

Matice přechodů

$$\mathbb{P}_{i,j} = p(i, j) = P(X_{n+1} = s_j | X_n = s_i) = p_{i,j}$$

Řetězec zapisujeme

- diagramem,
- maticí přechodů.

6.1 Stacionární distribuce

- Po čase se některé Markovovy řetězce ustálí v nějakém stavu.
- Rovnovážná distribuce

$$\pi : \pi = \pi * \mathbb{P}$$

- Detailní rovnováha

$$\pi_i p_{i,j} = \pi_j p_{j,i} \quad \forall i, j$$

7 Systémy hromadné obsluhy

Modely:

- $M/M/1$
- $M/M/m$
- $M/G/\infty$

Míra vytížení:

$$\rho = \frac{\lambda}{\mu}$$

(musí být < 1 , aby systém pracoval)

Little's law (průměrný počet zákazníků ve frontě):

$$N = \lambda * T$$

8 Ostatní

8.1 Řady

Výpočet řad

$$\begin{aligned} \sum_{n=0}^{\infty} r^n &= \frac{1}{1-r} \\ \sum_{n=0}^{\infty} n * r^n &= \frac{r}{(1-r)^2} \end{aligned}$$

Nekonečná řada:

$$\frac{a_1}{1-q}$$

(a_1 – počátek, q – kvocient)