

Statistika pro informatiku

Souhrn látky

únor 2014

Obsah

1	Základy statistiky a pravděpodobnosti	3
1.1	Rozlišení pojmů	3
1.2	Pravděpodobnost jevu a jeho doplňku	3
1.3	Sjednocení jevů	3
1.4	Podmíněná pravděpodobnost	4
1.5	Nezávislost jevů (průnik)	4
1.6	Bayesova věta	4
2	Vlastnosti	5
2.1	Střední hodnota	5
2.2	Rozptyl	5
2.3	Distribuční funkce	5
2.4	Hustota	6
3	Rozdělení pravděpodobnosti	6
3.1	Diskrétní (nespojité) rozdělení	6
3.1.1	Bernoulliho rozdělení	6
3.1.2	Binomické rozdělení	6
3.1.3	Geometrické rozdělení	7
3.1.4	Poissonovo rozdělení	7
3.2	Spojité rozdělení	7
3.2.1	Rovnoměrné rozdělení	7
3.2.2	Exponenciální rozdělení	7
3.2.3	Normální (gaussovo) rozdělení	8
4	Entropie	8
4.1	Vzájemná informace	9
4.2	Kódování	9
4.3	Suma	9
4.4	Logaritmus	10
5	Náhodné procesy	10
6	Markovovy řetězce	10
6.1	Stacionární distribuce	11

Rejstřík

stacionarita procesu, 10

1 Základy statistiky a pravděpodobnosti

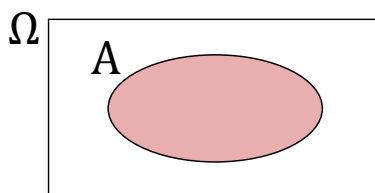
1.1 Rozlišení pojmů

Statistika TODO

Pravděpodobnost TODO

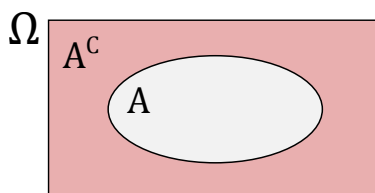
1.2 Pravděpodobnost jevu a jeho doplňku

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



Obrázek 1: Vennův diagram základní pravděpodobnosti jevu

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

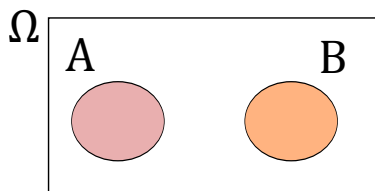


Obrázek 2: Vennův diagram doplňku jevu

1.3 Sjednocení jevů

Pro disjunktní jevy platí

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

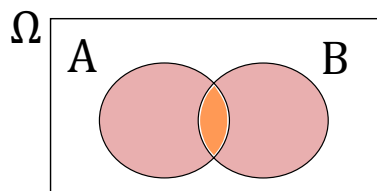


Obrázek 3: Disjunktní jevy

Pro jevy platí

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

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

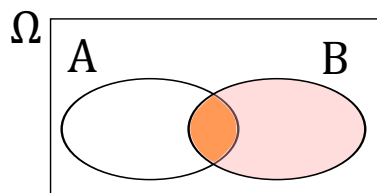


Obrázek 4: Jevy

1.4 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.“



Obrázek 5: Podmíněná pravděpodobnost

1.5 Nezávislost jevů (průnik)

Pro **nezávislé** jevy platí

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

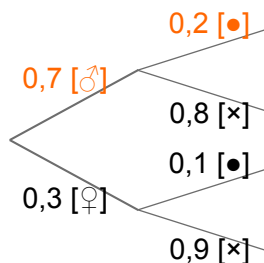
Jinak platí

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

1.6 Bayessova věta

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

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



Obrázek 6: Bayessova věta pomocí stromu

2 Vlastnosti

2.1 Střední hodnota

Pro diskrétní veličiny

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

Pro spojité veličiny

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

(P a f jsou funkce hustoty.)

Pro libovolné náhodné veličiny platí:

$$\mathbb{E}(aX + Y) = a\mathbb{E}(X) + \mathbb{E}(Y) \text{ (linearita)}$$

$$\mathbb{E}(X + Y) = \mathbb{E}(X) + \mathbb{E}(Y)$$

$$\mathbb{E}(X + Y) = \mathbb{E}(\max\{X, Y\}) + \mathbb{E}(\min\{X, Y\})$$

$$\mathbb{E}X^2 = \sum_i p_i x_i^2 \text{ (pro diskrétní jevy)}$$

$$\mathbb{E}(\max\{X, Y\}) = \mathbb{E}(X) + \mathbb{E}(Y) - \mathbb{E}(\min\{X, Y\})$$

$$\mathbb{E}(XY) = \mathbb{E}X * \mathbb{E}Y \text{ (platí jen pro nezávislé jevy)}$$

2.2 Rozptyl

$$\sigma_x = \text{var}X = \mathbb{E}(X - \mathbb{E}X)^2 = \mathbb{E}X^2 - (\mathbb{E}X)^2$$

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

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

3 Rozdělení pravděpodobnosti

3.1 Diskrétní (nespojité) rozdělení

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

3.1.1 Bernoulliho rozdělení

$$X \sim Be(p)$$

Hustota

$$\mathbb{P}(1) = p, \mathbb{P}(0) = 1 - p$$

Střední hodnota

$$\mathbb{E}X = p$$

Rozptyl

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

3.1.2 Binomické rozdělení

$$X \sim Bi(n, p)$$

Hustota

$$\mathbb{P}_X(k) = \binom{n}{k} p^k (1 - p)^{n-k}$$

Střední hodnota

$$\mathbb{E}X = n * p$$

Rozptyl

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

3.1.3 Geometrické rozdělení

$$X \sim \text{geom}(p)$$

- „Počet hodů mincí než padne první panna. Tedy čekání na úspěch.“

Hustota

$$\mathbb{P}_x(k) = (1-p)^{k-1} * p$$

Distribuční funkce

$$\mathbb{P}(T \leq n) = 1 - (1-p)^n$$

Funkce přežití

$$\mathbb{P}(T > n) = (1-p)^n$$

Střední hodnota

$$\mathbb{E}X = \frac{1}{p}$$

Rozptyl

$$\text{var}X = \frac{1}{p} \left(\frac{1}{p} - 1 \right)$$

3.1.4 Poissonovo rozdělení

TODO

3.2 Spojité rozdělení

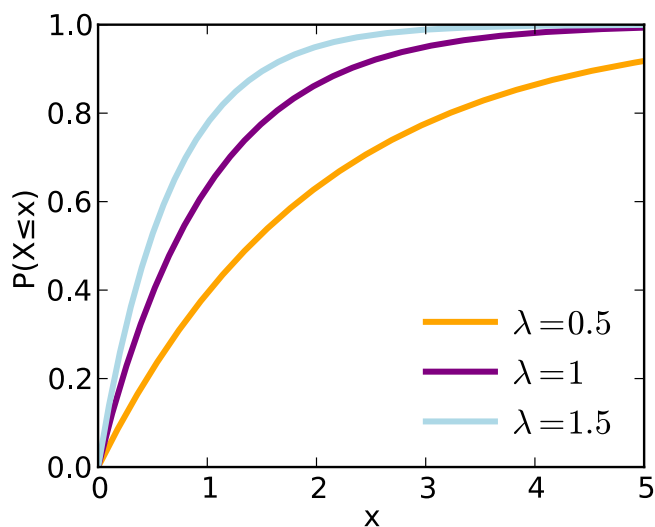
Spojitě náhodné veličiny nabývají na rozdíl od diskrétních veličin nějakého intervalu.

3.2.1 Rovnoměrné rozdělení

TODO

3.2.2 Exponenciální rozdělení

$$\begin{aligned} X &\sim \text{exp}(\lambda) \\ f(x) &= \lambda e^{-\lambda x}, x \geq 0 \\ P(X \leq x) &= 1 - e^{-\lambda x} \\ \mathbb{E}X &= \frac{1}{\lambda} \\ \text{var}X &= \frac{1}{\lambda^2} \end{aligned}$$



Obrázek 7: Graf distribuční funkce exponenciálního rozdělení[1]

3.2.3 Normální (gaussovo) rozdělení

TODO

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

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

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

Sdružená entropie

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

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

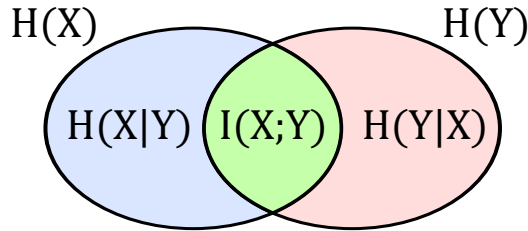
Podmíněná entropie

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

4.1 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 8: Vzájemná informace a entropie (tedy $H(X, Y)$)

4.2 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.

4.3 Suma

Výpočet sumy

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

4.4 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

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

Markovova podmínka

TODO

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

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

Reference

- [1] The Free Encyclopedia Wikipedia. Exponential distribution. 2014.