Statistika pro informatiku

Souhrn látky

únor 2014

Obsah

1 Základy statistiky a pravděpodobnosti

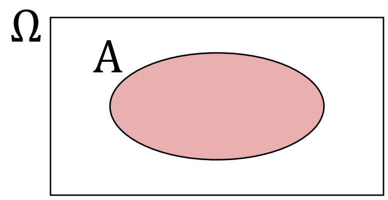
1.1 Rozlišení pojmů

Statistika TODO

Pravděpodobnost TODO

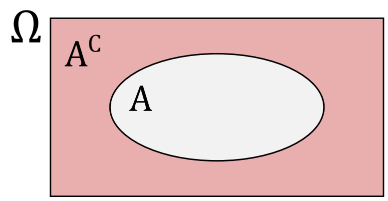
1.2 Pravděpodobnost jevu a jeho doplňku

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



Obrázek 1: Vennův diagram základní pravděpodobnosti jevu

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

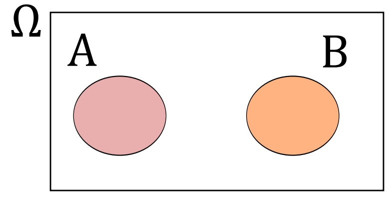


Obrázek 2: Vennův diagram doplňku jevu

1.3 Sjednocení jevů

Pro disjunktní jevy platí

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

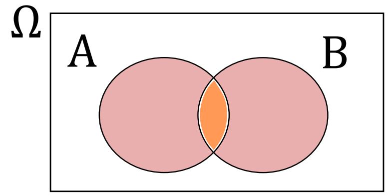


Obrázek 3: Disjunktní jevy

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.

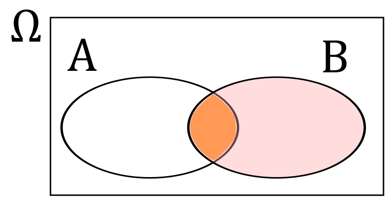


Obrázek 4: Jevy

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

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

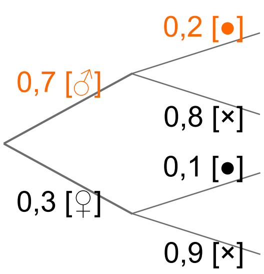
Jinak platí

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

1.6 Bayessova věta

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

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



Obrázek 6: Bayessova věta pomocí stromu

2 Vlastnosti

2.1 Střední hodnota

Pro diskrétní veličiny

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

Pro spojité veličiny

$$\mathbb{E}X = \int_{-\infty}^{+\infty} x * f_x\left(x\right)\mathrm{d}x$$

(P a f jsou funkce hustoty.)

Pro libovolné náhodné veličiny platí:

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

2.2 Rozptyl

$$varX = EX^2-\left(EX\right)^2$$

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

$$F = \mathbb{P}(X \leq x_i) = \int_{-\infty}^{x} f_x(u)\,\mathrm{d}u \;\forall x \in \mathbb{R}$$

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

$$varX = p(1-p)$$

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

$$varX = np(1-p)$$

3.1.3 Geometrické rozdělení

$$X \sim geom\left(p\right)$$

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

Hustota

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

Distribuční funkce

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

Funkce přežití

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

Střední hodnota

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

Rozptyl

$$varX = \frac{1}{p}\left(\frac{1}{p} - 1\right)$$

3.1.4 Poissonovo rozdělení

TODO

3.2 Spojité rozdělení

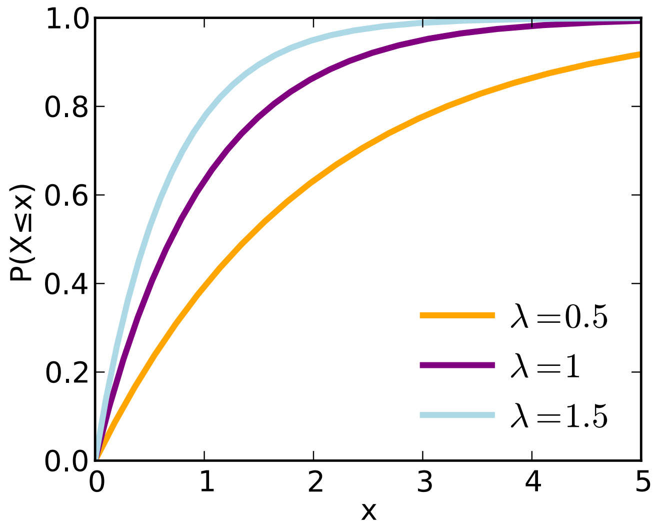
Spojité náhodné veličiny nabývají na rozdíl od diskrétních veličin nějakého intervalu.

3.2.1 Rovnoměrné rozdělení

TODO

3.2.2 Exponenciální rozdělení

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



Obrázek 7: Graf distribuční funkce exponenciálního rozdělení[1]

3.2.3 Normální (gaussovo) rozdělení

TODO

4 Entropie

Pro entropii platí aditivita entropie

$$H\left(X,Y\right)=H\left(X\right)+H\left(Y|X\right).$$

4.1 Diskrétní veličiny

$$H\left(X\right)=-\sum p_i\log p_i$$

4.2 Spojité veličiny

$$H\left(X\right)=-\int f\left(x\right)\log f\left(x\right)\mathrm{d}x$$

4.3 Sdružená entropie

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

4.4 Podmíněná entropie

$$H\left(X|Y\right)=-\sum_{i,j} p\left(x_i,y_i\right)\log p\left(y_i|x_i\right)$$

5 Náhodné procesy

Značení procesu

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

Střední hodnota

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

Autokorelační funkce

$$\begin{aligned}R_x\left(t_1,\ t_2\right) &= \mathbb{E}X\left(t_1\right)*\overline{X\left(t_2\right)}\ \text{v}\ \mathbb{C}\\ &= \mathbb{E}X\left(t_1\right)*X\left(t_2\right)\ \text{v}\ \mathbb{R}\end{aligned}$$

Reference

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