Teorie informace

Mirko Navara
Centrum strojového vnímání
katedra kybernetiky FEL ČVUT
Karlovo náměstí, budova G, místnost 104a
http://cmp.felk.cvut.cz/˜navara/psi

3. 1. 2017

Obsah

1 Informace

Teorie pravděpodobnosti popisovala systém, v němž bude proveden náhodný pokus.
Jeho výsledky (i dílčí) popíše **informace.**
Informací může být určení výsledku náhodného pokusu, tj. hodnoty náhodné veličiny,

- předpověď,
- stanovení výsledku,
- sdělení výsledku.

Předpokládáme **diskrétní** rozdělení.

Příklad 1.1 (**provozování loterie**). *V loterii bude* $\doteq 500\,000$ *losů po 4 \$, z nichž 1 vyhraje 1 M\$.*
Certifikované losovací zařízení s 500 000 *výsledky není.*
(Např. ruleta o průměru 1 km.*)*
Losy očíslujeme a necháme ***nezávisle*** *vylosovat číslice.*
Problém: *Kolik máme zaplatit za vylosování číslic?*
(1. číslice má jen 5 stejně pravděpodobných hodnot.)
Řešení: *Podle hry větší-menší vylosujeme* 19 *nezávislých binárních číslic, čímž rozlišíme* $2^{19} = 524\,288$ *možností.*
Máme za vylosování každé číslice zaplatit stejně?

Ano! *(Mj. proto, že všechny hody mincí mají stejnou pracnost.)*
K výběru 1 *z* 2^k *(*$k \in \mathbb{N}$*) stejně pravděpodobných možností potřebujeme informaci*

$$I(2^k) := k\,I(2) = k \text{ b}\,, \qquad I(2) = 1 \text{ b } \textit{(bit)}.$$

Rychlejší řešení: *Vždy vylosujeme ze* 3 *možností, postačí* 12 *nezávislých pokusů,* $3^{12} = 531\,441$.
Kolik informace $I(3)$ *dává výběr ze* 3 *stejně pravděpodobných možností?*

$$\begin{aligned}
3^5 = 243 &\doteq 256 = 2^8\,, & I(3) &\doteq \frac{8}{5} = 1.6\,,\\
3^{12} = 531\,441 &\doteq 524\,288 = 2^{19}\,, & I(3) &\doteq \frac{19}{12} = 1.583\,,\\
&\dots\\
3^{\log_3 M} = M &= 2^{\log_2 M}\,, & I(3) &= \frac{\log_2 M}{\log_3 M} = \log_2 3 \doteq 1.585\,.
\end{aligned}$$

Obecně výběr 1 z n stejně pravděpodobných možností nese informaci

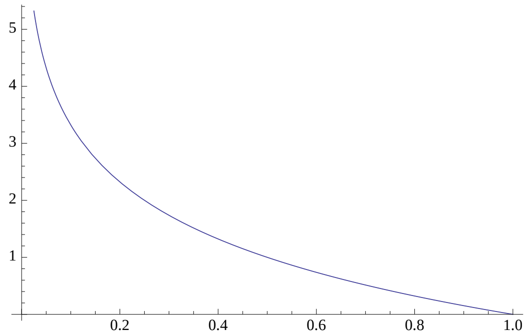
$$I(n) := \log_2 n\,.$$

I = **Hartleyho míra informace**. Je to jediná funkce $I\colon \mathbb{N} \to \mathbb{R}$ s vlastnostmi:

- $I(m \cdot n) = I(m) + I(n)$,
- I je neklesající,
- $I(2) = 1$.

Základ logaritmu je 2, čímž je dána jednotka informace (b = bit).
Speciálně pro $n = 2^k$: $\quad I(2^k) = k\,.$
Nadále základ neoznačujeme,

$$\log = \log_2\,.$$



Funkce $-\log p$ pro $p \in (0,1\rangle$

Derivace:

$$(\log x)' = \left(\frac{\ln x}{\ln 2}\right)' = \frac{1}{x\,\ln 2} \doteq \frac{1}{0.693\,x} \doteq \frac{1.443}{x}\,.$$

Kolik informace dostaneme, když se dovíme, že správná z n stejně pravděpdobobných možností patří do skupiny s $k < n$ prvky?
Chybí informace $I(k)$, abychom určili jediný výsledek, tj. abychom měli celkem $I(n)$; nyní máme

$$I(n) - I(k) = \log n - \log k = \log\frac{n}{k}\,.$$

Záleží jen na poměru $\frac{n}{k}$, tj. na pravděpdobnosti $p := \frac{k}{n}$, což dovoluje zobecnění

$$I\left(\frac{n}{k}\right) = I\left(\frac{1}{p}\right) := \log\frac{n}{k} = \log\frac{1}{p} = -\log p\,.$$

I pro iracionální $p \in (0,1\rangle$ definujeme jako jedinou spojitou funkci s výše uvedenými hodnotami pro racionální argumenty:

$$I\left(\frac{1}{p}\right) := -\log p\,.$$

(Nepravděpodobnější výsledek $\Rightarrow$ větší informace.)

2 Entropie

Příklad 2.1.

hodnota	0	1
pravděpodobnost	$\frac{1}{3}$	$\frac{2}{3}$
informace	$I(3) = \log 3 \doteq 1.585$	$I\left(\frac{3}{2}\right) = \log \frac{3}{2} = \log 3 - 1 \doteq 0.585$

Cena "předplacené" informace je střední hodnota

$$h\left(\frac{1}{3}, \frac{2}{3}\right) := \frac{1}{3} \log 3 + \frac{2}{3} \log \frac{3}{2} = \log_2 3 - \frac{2}{3} \doteq 0.918\,.$$

Obecněji:

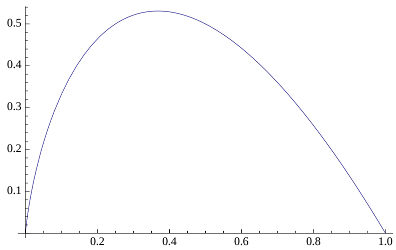
hodnota	x_1	x_2
pravděpodobnost	p	$1-p$
informace	$I\left(\frac{1}{p}\right) = -\log p$	$I\left(\frac{1}{1-p}\right) = -\log(1-p)$

$$h(p, 1-p) := \underbrace{-p \log p}_{\iota(p)} \underbrace{-(1-p) \log(1-p)}_{\iota(1-p)}\,,$$

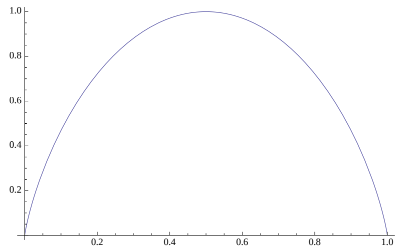
kde

$$\iota(p) := -p \log p\,, \qquad p > 0\,,$$
$$\iota(0) := 0\,.$$

$h(p, 1-p) =$ entropie alternativního rozdělení.



Funkce $-p \log p$ pro $p \in \langle 0, 1\rangle$



Funkce $h(p, 1-p)$ pro $p \in \langle 0, 1\rangle$

Obecně:

hodnota	x_1	x_2	...
pravděpodobnost	p_1	p_2	...
informace	$I\left(\frac{1}{p_1}\right) = -\log p_1$	$I\left(\frac{1}{p_2}\right) = -\log(p_2)$	...

Entropie rozdělení s pravděpodobnostmi $\boldsymbol{p} = (p_1, p_2, \ldots)$, $\sum_i p_i = 1$:

$$h(\boldsymbol{p}) = h(p_1, p_2, \ldots) := \sum_i \iota(p_i) = -\sum_i p_i \log p_i\,.$$

(Výsledků je spočetně mnoho, nutno předpokládat konvergenci sumy.)
Entropie náhodné veličiny X s **různými** hodnotami $x_1, x_2, \ldots$ a pravděpodobnostní funkcí p_X:

$$H(X) := h(p_X(x_1), p_X(x_2), \ldots) = \sum_x \iota(p_X(x)) = -\sum_x p_X(x) \log p_X(x) =$$
$$= -\mathrm{E} \log p_X(X)\,.$$

$\neq -\mathrm{E} \log X$

Příklad 2.2.

hodnota	x	1	4	9
pravděpodobnost	$p_X(x)$	$\frac{1}{2}$	$\frac{1}{4}$	$\frac{1}{4}$
informace	$-\log p_X(x)$	1	2	2

$$H(X) = h\left(\frac{1}{2}, \frac{1}{4}, \frac{1}{4}\right) = \frac{1}{2} \cdot 1 + \frac{1}{4} \cdot 2 + \frac{1}{4} \cdot 2 = \frac{3}{2}\,.$$

Příklad 2.3.

hodnota	x	-6	0	0.001	1000
pravd.	$p_X(x)$	0.1	0.2	0.2	0.5
inf.	$-\log p_X(x)$	$\log 10 \doteq 3.322$	$\log 5 \doteq 2.322$	$\log 5 \doteq 2.322$	$\log 2 = 1$

$$\begin{aligned} H(X) = h(0.1, 0.2, 0.2, 0.5) &= 0.1 \cdot \log 10 + 0.2 \cdot \log 5 + 0.2 \cdot \log 5 + 0.5 \cdot 1 = \\ &= 0.5 \cdot \log 5 + 0.6 \doteq 1.761\,. \end{aligned}$$

2.1 Entropie jako střední hodnota

Nahradili jsme hodnoty náhodné veličiny X jejich pravděpodobnostmi (jen na těch záleží!), tím jsme dostali náhodnou veličinu $W = p_X(X)$ s hodnotami v $\langle 0, 1\rangle$.
Její rozdělení popisuje *pravděpodobnosti pravděpodobností* náhodné veličiny X.

$$p_W(w) = P[p_X(X) = w]$$

= pravděpodobnost, že X nabývá hodnoty, *jejíž pravděpodobnost* je $w \in \langle 0, 1\rangle$.

Příklad 2.4.

hodnota	x	A	B	C	D	E	F
pravd.	$p_X(x)$	0.1	0.1	0.15	0.2	0.2	0.25
informace (*přibl.*)	$-\log p_X(x)$	$\log 10$ 3.322	$\log 10$ 3.322	$\log \frac{20}{3}$ 2.737	$\log 5$ 2.322	$\log 5$ 2.322	2 2

$W = p_X(X)$:

hodnota	w	0.1	0.15	0.2	0.25
pravděpodobnost	$p_W(w)$	**0.2**	0.15	**0.4**	0.25
informace v x	$-\log w$	$\log 10 \doteq 3.322$	$\log \frac{20}{3} \doteq 2.737$	$\log 5 \doteq 2.322$	2

První řádek vznikl opsáním všech různých hodnot z 2. řádku první tabulky.
Ve skutečnosti potřebujeme jejich logaritmy, jsou v posledním řádku.
Druhý řádek obsahuje odpovídající pravděpodobnosti (každá byla vynásobena počtem svých výskytů ve 2. řádku první tabulky.

$$\begin{aligned} H(X) &= -\mathrm{E} \log p_X(X) = -\mathrm{E} \log W \doteq \\ &\doteq 0.2 \cdot 3.322 + 0.15 \cdot 2.737 + 0.4 \cdot 2.322 + 0.25 \cdot 2 \doteq 2.504\,. \end{aligned}$$

2.2 Vlastnosti entropie

$$0 \le H(X) \le \log n\,,$$

kde n je počet různých hodnot náhodné veličiny X (je-li konečný).
$0 = H(X)$ jen pro Diracovo rozdělení,
$H(X) = \log n$ (kde n je počet možných hodnot) jen pro rovnoměrné rozdělení:

Věta 2.1. *Maximální entropie se nabývá jen pro rovnoměrné rozdělení.*

Důkaz. Předpokládejme, že rozdělení $(p_1, p_2, ..., p_n)$ není rovnoměrné. Pak některá z pravděpodobností je větší a některá menší než $\frac{1}{n}$; BÚNO* $p_1 > \frac{1}{n}$, $p_2 < \frac{1}{n}$.
Změnou $p_1 := p_1 - \varepsilon$, $p_2 := p_2 + \varepsilon$ o malé $\varepsilon \ge 0$ se entropie změní o

$$\begin{aligned} \Delta(\varepsilon) &= h(p_1 - \varepsilon, p_2 + \varepsilon, p_3, ..., p_n) - h(p_1, p_2, p_3, ..., p_n) = \\ &= \iota(p_1 - \varepsilon) + \iota(p_2 + \varepsilon) - \iota(p_1) - \iota(p_2)\,, \end{aligned}$$

kde

$$\iota'(p) = (-p \log p)' = -\frac{1}{\ln 2} - \log p\,.$$

Funkce Δ je diferencovatelná, $\Delta(0) = 0$,

$$\Delta'(\varepsilon) = -\iota'(p_1 - \varepsilon) + \iota'(p_2 + \varepsilon) = \log(p_1 - \varepsilon) - \log(p_2 + \varepsilon) = \log \frac{p_1 - \varepsilon}{p_2 + \varepsilon}\,.$$

Pro dostatečně malé $\varepsilon > 0$ je

$$p_1 - \varepsilon > \frac{1}{n} > p_2 + \varepsilon\,,$$

$\Delta'(\varepsilon) > 0$ a $\Delta(\varepsilon) > 0$.
Tedy entropie $h(p_1, p_2, ..., p_n)$ není maximální. □

* *BÚNO = bez újmy na obecnosti*

Důsledek 2.1. *Entropie rozdělení s nekonečně mnoha hodnotami může být libovolně velká.*

Věta 2.2.

$$h(p_1, p_2, p_3, ...) = h(p_1 + p_2, p_3, ...) + (p_1 + p_2)\, h\left(\frac{p_1}{p_1 + p_2}, \frac{p_2}{p_1 + p_2}\right).$$

Důkaz.

$$\begin{aligned}
&h(p_1, p_2, p_3, ...) - h(p_1 + p_2, p_3, ...) = \\
&= \iota(p_1) + \iota(p_2) - \iota(p_1 + p_2) = \\
&= -p_1 \log p_1 - p_2 \log p_2 + (p_1 + p_2) \log(p_1 + p_2) = \\
&= -p_1 \log \frac{p_1}{p_1 + p_2} - p_2 \log \frac{p_2}{p_1 + p_2} = \\
&= (p_1 + p_2) \left(-\frac{p_1}{p_1 + p_2} \log \frac{p_1}{p_1 + p_2} - \frac{p_2}{p_1 + p_2} \log \frac{p_2}{p_1 + p_2} \right) = \\
&= (p_1 + p_2)\, h\left(\frac{p_1}{p_1 + p_2}, \frac{p_2}{p_1 + p_2}\right).
\end{aligned}$$

□

3 Informační divergence

Příklad 3.1 (chybný předpoklad rovnoměrného rozdělení). *Předpokládali jsme rozdělení $\boldsymbol{q} = \left(\frac{1}{2}, \frac{1}{2}\right)$ a odhadli entropii $h\left(\frac{1}{2}, \frac{1}{2}\right) = 1$.*
Ve skutečnosti bylo rozdělení $\boldsymbol{p} = \left(\frac{1}{3}, \frac{2}{3}\right)$ a entropie $h\left(\frac{1}{3}, \frac{2}{3}\right) \doteq 0.918$.
O kolik jsme nadhodnotili jeden znak?

$$h\left(\frac{1}{2}, \frac{1}{2}\right) - h\left(\frac{1}{3}, \frac{2}{3}\right) \doteq 1 - 0.918 = 0.082\,.$$

Příklad 3.2 (chybný předpoklad *nerovnoměrného* rozdělení). *Předpokládali jsme rozdělení $\boldsymbol{q} = \left(\frac{1}{3}, \frac{2}{3}\right)$ a odhadli entropii $h\left(\frac{1}{3}, \frac{2}{3}\right) \doteq 0.91$.*
Ve skutečnosti bylo rozdělení $\boldsymbol{p} = \left(\frac{1}{6}, \frac{5}{6}\right)$ a entropie $h\left(\frac{1}{6}, \frac{5}{6}\right) \doteq 0.65$.
O kolik jsme nadhodnotili jeden znak?
Odpověď:

$$h\left(\frac{1}{3}, \frac{2}{3}\right) - h\left(\frac{1}{6}, \frac{5}{6}\right) \doteq 0.918 - 0.65 = 0.268\,.$$

Špatně!
Správná odpověď:

hodnota	x_i	x_1	x_2
předp. pravd.	q_i	$\frac{1}{3}$	$\frac{2}{3}$
předp. informace	$-\log q_i$	$\log 3 \doteq 1.585$	$\log \frac{3}{2} \doteq 0.585$
skut. pravd.	p_i	$\frac{1}{6}$	$\frac{5}{6}$
skut. informace	$-\log p_i$	$\log 6 \doteq 2.585$	$\log \frac{6}{5} \doteq 0.263$
předp. − skut. inf.	$\log p_i - \log q_i = \log \frac{p_i}{q_i}$	-1	$\log \frac{5}{4} \doteq 0.322$

Střední chyba odhadu informace je

$$p_1 \log \frac{p_1}{q_1} + p_2 \log \frac{p_2}{q_2} \doteq \frac{1}{6} \cdot (-1) + \frac{5}{6} \cdot 0.322 \doteq 0.102\,.$$

Příklad 3.3 (chybný předpoklad *nerovnoměrného* rozdělení 2)**.** *Vyměníme rozdělení z předchozího příkladu: Předpokládali jsme rozdělení* $\boldsymbol{q} = \left(\frac{1}{6}, \frac{5}{6}\right)$ *a odhadli entropii* $h\left(\frac{1}{6}, \frac{5}{6}\right) \doteq 0.65$.
Ve skutečnosti bylo rozdělení $\boldsymbol{p} = \left(\frac{1}{3}, \frac{2}{3}\right)$ *a entropie* $h\left(\frac{1}{3}, \frac{2}{3}\right) \doteq 0.91$.
O kolik jsme "nadhodnotili" jeden znak?
Špatná odpověď: $h\left(\frac{1}{6}, \frac{5}{6}\right) - h\left(\frac{1}{3}, \frac{2}{3}\right) \doteq 0.65 - 0.918 = -0.268$.
Správná odpověď:

hodnota	x_i	x_1	x_2
předp. pravd.	q_i	$\frac{1}{6}$	$\frac{5}{6}$
předp. inf.	$-\log q_i$	$\log 6 \doteq 2.585$	$\log \frac{6}{5} \doteq 0.263$
skut. pravd.	p_i	$\frac{1}{3}$	$\frac{2}{3}$
skut. inf.	$-\log p_i$	$\log 3 \doteq 1.585$	$\log \frac{3}{2} \doteq 0.585$
předp. − skut. inf.	$\log p_i - \log q_i = \log \frac{p_i}{q_i}$	1	$\log \frac{4}{5} \doteq -0.322$

Střední chyba odhadu informace je

$$p_1 \log \frac{p_1}{q_1} + p_2 \log \frac{p_2}{q_2} \doteq \frac{1}{3} \cdot 1 - \frac{2}{3} \cdot 0.322 \doteq 0.119\,.$$

Odhad byl opět nadhodnocený! (Ne stejně.)

Obecně: Předpokládali jsme rozdělení $\boldsymbol{q} = (q_1, q_2, \ldots)$,
ve skutečnosti bylo $\boldsymbol{p} = (p_1, p_2, \ldots)$.

hodnota	x_i	x_1	x_2	$\ldots$
předp. pravd.	q_i	q_1	q_2	$\ldots$
předp. inf.	$-\log q_i$	$-\log q_1$	$-\log q_2$	$\ldots$
skut. pravd.	p_i	p_1	p_2	$\ldots$
skut. inf.	$-\log p_i$	$-\log p_1$	$-\log p_2$	$\ldots$
předp. − skut. inf.	$\log p_i - \log q_i = \log \frac{p_i}{q_i}$	$\log \frac{p_1}{q_1}$	$\log \frac{p_2}{q_2}$	$\ldots$

Střední chyba odhadu informace je

$$D(\boldsymbol{p}||\boldsymbol{q}) := \sum_i p_i \log \frac{p_i}{q_i} = \underbrace{\sum_i p_i \log p_i}_{-h(p_1,p_2,\ldots)} - \sum_i p_i \log q_i\,,$$

kde $0 \log 0 := 0, \quad 0 \log \frac{0}{0} := 0, \quad r \log \frac{r}{0} := \infty$ pro $r \in (0,1\rangle$.
$D(\boldsymbol{p}||\boldsymbol{q})$ = **Kullbackova–Leiblerova divergence** = **informační divergence** rozdělení $\boldsymbol{p}, \boldsymbol{q}$.

3.1 Informační divergence jako střední hodnota

Máme náhodnou veličinu X, skutečnou pravděpodobnostní funkci p_X a předpokládanou pravděpodobnostní funkci q_X. (To zde *není* kvantilová funkce.)
Dva pravděpodobnostní modely budeme rozlišovat indexy, např

$$\mathrm{E}_p X = \sum_x x\, p_X(x)\,, \qquad H_p(X) = -\mathrm{E}_p \log p_X(X) = h(p_X(x_1), p_X(x_2), \ldots)\,,$$

$$\mathrm{E}_q X = \sum_x x\, q_X(x)\,, \qquad H_q(X) = -\mathrm{E}_q \log q_X(X) = h(q_X(x_1), q_X(x_2), \ldots)\,.$$

Při výsledku x je předpokládaná informace $-\log q_X(x)$ a skutečná $-\log p_X(x)$. Střední hodnota rozdílu při **skutečném** rozdělení p je

$$D(\boldsymbol{p}_X||\boldsymbol{q}_X) = \mathrm{E}_p \log \frac{p_X(X)}{q_X(X)} = \mathrm{E}_p(\log p_X(X) - \log q_X(X)) =$$
$$= \underbrace{\mathrm{E}_p \log p_X(X)}_{-H_p(X)} \underbrace{-\mathrm{E}_p \log q_X(X)}_{\neq H_q(X)}\,.$$

Příklad 3.4 (chybný předpoklad *rovnoměrného* rozdělení – pokr.)**.** *Proč nám to prvně vyšlo správně? Protože* $q_X(X) = \frac{1}{2}$ *byla konstanta,*

$$H_q(X) = -\mathrm{E}_q \log q_X(X) = -\log \frac{1}{2} = -\mathrm{E}_p \log q_X(X)\,.$$

3.2 Vlastnosti informační divergence

Může být $D(\boldsymbol{p}||\boldsymbol{q}) \neq D(\boldsymbol{q}||\boldsymbol{p})$.

Věta 3.1. *$D(\boldsymbol{p}||\boldsymbol{q}) \geq 0$; rovnost nastává, právě když $\boldsymbol{p} = \boldsymbol{q}$.*

Důkaz. Předpokládejme $\boldsymbol{p} \neq \boldsymbol{q}$, BÚNO $p_1 < q_1, p_2 > q_2$. Výraz

$$V(q_1, q_2, q_3, \ldots) := \sum_i p_i \log q_i$$

se změnou $q_1 := q_1 - \varepsilon$, $q_2 := q_2 + \varepsilon$ o malé $\varepsilon \geq 0$ změní o

$$\begin{aligned}\Delta(\varepsilon) &= V(q_1 - \varepsilon, q_2 + \varepsilon, q_3, \ldots) - V(q_1, q_2, q_3, \ldots) = \\ &= p_1 \log(q_1 - \varepsilon) + p_2 \log(q_2 + \varepsilon) - p_1 \log q_1 - p_2 \log q_2 .\end{aligned}$$

Funkce Δ je diferencovatelná, $\Delta(0) = 0$,

$$\Delta'(\varepsilon) = -\frac{1}{\ln 2} \frac{p_1}{q_1 - \varepsilon} + \frac{1}{\ln 2} \frac{p_2}{q_2 + \varepsilon} .$$

Pro dostatečně malé $\varepsilon > 0$ je

$$\frac{p_1}{q_1 - \varepsilon} < 1 < \frac{p_2}{q_2 + \varepsilon} ,$$

$\Delta'(\varepsilon) > 0$ a $\Delta(\varepsilon) > 0$.
Tedy $V(q_1, q_2, q_3, \ldots)$ není maximální.
Uvedený argument nelze použít jedině pro $\boldsymbol{p} = \boldsymbol{q}$, pak nabývá výraz maxima, a to

$$\begin{aligned}V(p_1, p_2, p_3, \ldots) &:= \sum_i p_i \log p_i = -h(p_1, p_2, p_3, \ldots) , \\ D(\boldsymbol{p}||\boldsymbol{p}) &= 0 .\end{aligned}$$

□

4 Sdružená entropie

Sdružená entropie $H(X, Y)$ náhodných veličin X, Y je entropie náhodného vektoru (X, Y), počítaná podle stejného principu:
X nabývá hodnot $x_1, x_2, \ldots$,
Y nabývá hodnot $y_1, y_2, \ldots$,
(X, Y) nabývá hodnot

$$\begin{aligned}&(x_1, y_1), (x_1, y_2), \ldots , \\ &(x_2, y_1), (x_2, y_2), \ldots , \\ &\ldots ,\end{aligned}$$

zajímají nás jen jejich pravděpodobnosti $p_{X,Y}(x_i, y_j)$, z nichž vypočteme

$$\begin{aligned}H(X, Y) &:= \sum_i \sum_j \iota(p_{X,Y}(x_i, y_j)) = -\sum_i \sum_j p_{X,Y}(x_i, y_j) \log p_{X,Y}(x_i, y_j) = \\ &= \sum_x \sum_y \iota(p_{X,Y}(x, y)) = -\sum_x \sum_y p_{X,Y}(x, y) \log p_{X,Y}(x, y) = \\ &= -\mathrm{E} \log p_{X,Y}(X, Y) .\end{aligned}$$

Pro X, Y **nezávislé** je

$$p_{X,Y}(x,y) = p_X(x) \cdot p_Y(y)\,,$$
$$-\log p_{X,Y}(x,y) = -\log p_X(x) - \log p_Y(y)\,,$$

pro všechna x, y, tedy i pro střední hodnoty

$$H(X,Y) = -\mathrm{E}\log p_{X,Y}(X,Y) = -\mathrm{E}\log p_X(X) - \mathrm{E}\log p_Y(Y) =$$
$$= H(X) + H(Y)\,.$$

Co znamená zápis $H(\cdot)$?

p. argumentů	typ argumentů	význam	příklad
1	náhodná veličina	entropie náh. veličiny	$H(X)$
>1	náhodné veličiny	entropie náh. vektoru (sdr. entropie náh. veličin)	$H(X,Y)$

Co znamená zápis $h(\cdot)$?

p. argumentů	typ argumentů	význam	příklad
>1	čísla $\in \langle 0,1\rangle$ se součtem 1	entropie rozdělení	$h\left(\frac{1}{2},\frac{1}{3},\frac{1}{6}\right)$
1	číslo $\in \langle 0,1\rangle$	zkratka pro entropii alternativního rozdělení	$h\left(\frac{1}{3}\right) = h\left(\frac{1}{3},\frac{2}{3}\right) = h\left(\frac{2}{3}\right)$

Příklad 4.1 (poklad na ostrově)**.**
$p_{X,Y}(x,y)$:

$x \backslash y$	A	B	C	$p_X(x)$
1	$\frac{1}{6}$	$\frac{1}{6}$	0	$\frac{1}{3}$
2	$\frac{1}{6}$	$\frac{1}{6}$	0	$\frac{1}{3}$
3	0	$\frac{1}{6}$	$\frac{1}{6}$	$\frac{1}{3}$
$p_Y(y)$	$\frac{1}{3}$	$\frac{1}{2}$	$\frac{1}{6}$	1

$H(X) = \log 3 \doteq 1.585$,
$H(Y) = \frac{1}{3}\log 3 + \frac{1}{2} + \frac{1}{6}\log 6 \doteq 1.459$,
$H(X,Y) = \log 6 \doteq 2.585 <$
$< H(X) + H(Y) \doteq 1.585 + 1.459 = 3.044$. *Závislost X, Y nás připravila o* $H(X) + H(Y) - H(X,Y) \doteq 3.044 - 2.585 = 0.459$. *Jak vyčíslit ztrátu informace z jedné proměnné, když jsme se dověděli druhou?*

4.1 Podmíněná entropie

Pokud nastal jev $Y = y$, aktualizujeme pravděpodobnosti hodnot veličiny X pomocí podmíněných pravděpodobností

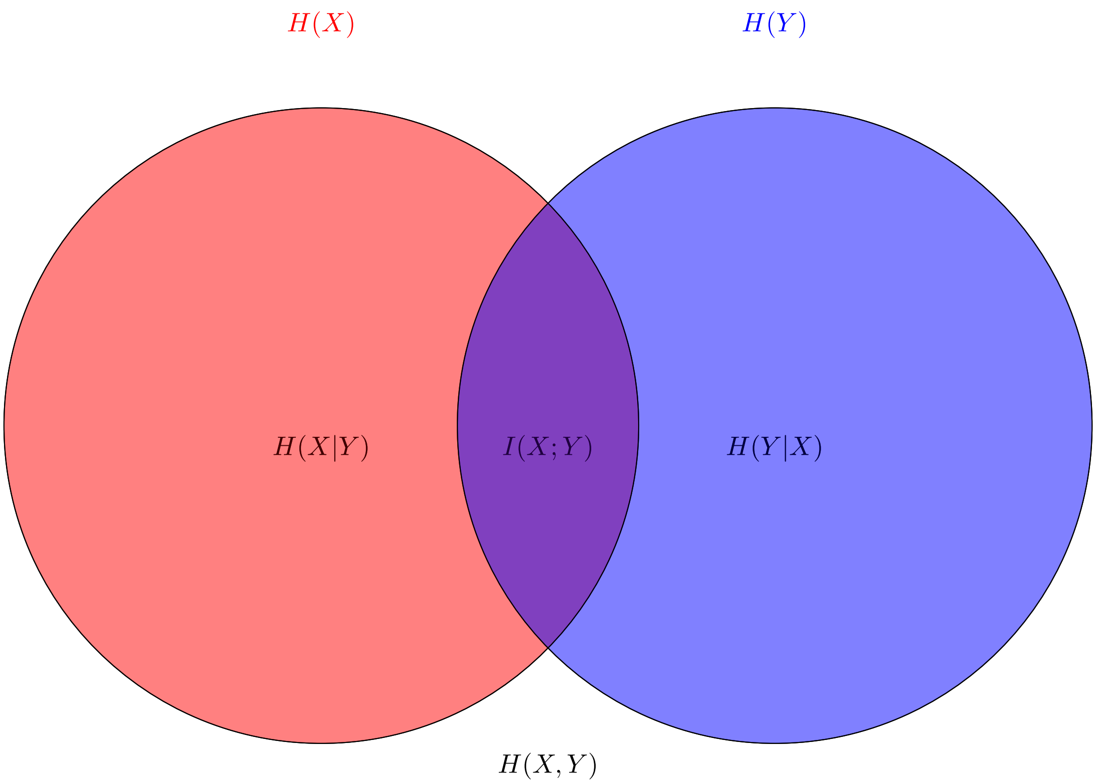
$$p_{X|Y}(x|y) := P[X = x | Y = y] = \frac{P[X = x, Y = y]}{P[Y = y]} = \frac{p_{X,Y}(x,y)}{p_Y(y)}\,.$$

V tomto modelu udává střední hodnotu informace veličiny X **podmíněná entropie**

$$H(X|Y=y) := \sum_x \iota(p_{X|Y}(x|y)) = -\sum_x p_{X|Y}(x|y)\,\log p_{X|Y}(x|y)\,.$$

Její střední hodnota v závislosti na Y je **střední podmíněná entropie**

$$H(X|Y) := \sum_y p_Y(y)\,H(X|Y=y) = -\sum_x\sum_y p_Y(y)\,p_{X|Y}(x|y)\,\log p_{X|Y}(x|y) =$$
$$= -\sum_x\sum_y p_{X,Y}(x,y)\,\log\frac{p_{X,Y}(x,y)}{p_Y(y)} =$$
$$= -\sum_x\sum_y p_{X,Y}(x,y)\,\log p_{X,Y}(x,y) + \sum_x\sum_y p_{X,Y}(x,y)\,\log p_Y(y) =$$
$$= -\mathrm{E}\log p_{X,Y}(X,Y) + \mathrm{E}\log p_Y(Y) = H(X,Y) - H(Y)\,.$$



Obdobně

$$H(Y|X=x) = -\sum_y p_{Y|X}(y|x)\,\log p_{Y|X}(y|x)\,,$$
$$H(Y|X) = \sum_x p_X(x)\,H(Y|X=x) = -\sum_x \sum_y p_{X,Y}(x,y)\,\log p_{Y|X}(y|x) =$$
$$= -\mathrm{E}\log p_{X,Y}(X,Y) + \mathrm{E}\log p_X(X) = H(X,Y) - H(X)\,.$$

5 Vzájemná informace

Ze sdruženého rozdělení $p_{X,Y}$ vypočítáme marginální pravděpodobnosti

$$p_X(x) = \sum_y p_{X,Y}(x,y)\,,$$
$$p_Y(y) = \sum_x p_{X,Y}(x,y)\,.$$

Z nich vypočítáme sdružené rozdělení

$$q_{X,Y}(x,y) := p_X(x)\cdot p_Y(y)\,,$$

které má stejné marginální pravděpodobnosti $q_X = p_X$, $q_Y = p_Y$, ale odpovídá **nezávislým** náhodným veličinám. Při rozdělení $q_{X,Y}$ by sdružená entropie vyšla $H(X)+H(Y)$. Skutečné rozdělení $p_{X,Y}$ vede na sdruženou entropii

$$H(X,Y) = -\sum_x \sum_y p_{X,Y}(x,y)\,\log p_{X,Y}(x,y) = -\mathrm{E}\log p_{X,Y}(X,Y)\,.$$

Ztrátu informace v důsledku závislosti X, Y vyjadřuje **vzájemná informace** $I(X;Y)$ náhodných veličin X, Y, což je informační divergence $D(\boldsymbol{p}_{X,Y}||\boldsymbol{q}_{X,Y})$:

$$
\begin{aligned}
I(X;Y) := D(\boldsymbol{p}_{X,Y}||\boldsymbol{q}_{X,Y}) &= \mathrm{E}\log p_{X,Y}(X,Y) - \mathrm{E}\log q_{X,Y}(X,Y) = \\
&= -H(X,Y) - \mathrm{E}\log p_X(X)\cdot p_Y(Y) = \\
&= -H(X,Y) - \mathrm{E}\log p_X(X) - \mathrm{E}\log p_Y(Y) = \\
&= -H(X,Y) + H(X) + H(Y)\,.
\end{aligned}
$$

5.1 Shrnutí vlastností vzájemné informace a podmíněné entropie

Vzájemná informace vyjadřuje, kolik informace o jedné proměnné nese druhá proměnná.

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

Důsledky:

$$
\begin{aligned}
I(X;Y) &= I(Y;X)\,,\\
H(X) &= I(X;Y) + H(X|Y)\,,\\
H(Y) &= I(X;Y) + H(Y|X)\,,\\
0 &\le I(X;Y) \le \min(H(X),H(Y))\,,\\
I(X;Y) = 0 &\iff X,Y \text{ jsou nezávislé},\\
I(X;Y) = H(X) &\iff \forall y\ \exists! x : p_{X|Y}(x|y) = 1 \text{ (podm. rozdělení jsou Diracova)},\\
I(X;X) &= H(X)\,.
\end{aligned}
$$

Vše lze beze změny uplatnit i na náhodné vektory.
Řetězcové pravidlo:

$$
\begin{aligned}
H(X_1,X_2,\ldots,X_n) = H(X_1|X_2,\ldots,X_n) &+ H(X_2|X_3,\ldots,X_n) + \ldots + \\
&+ H(X_{n-1}|X_n) + H(X_n)\,.
\end{aligned}
$$

Příklad 5.1 (poklad na ostrově – pokr.)**.**
$p_{X,Y}(x,y)$:

x \ y	A	B	C	$p_X(x)$
1	$\frac{1}{6}$	$\frac{1}{6}$	0	$\frac{1}{3}$
2	$\frac{1}{6}$	$\frac{1}{6}$	0	$\frac{1}{3}$
3	0	$\frac{1}{6}$	$\frac{1}{6}$	$\frac{1}{3}$
$p_Y(y)$	$\frac{1}{3}$	$\frac{1}{2}$	$\frac{1}{6}$	1

$H(X) \doteq 1.585$,
$H(Y) \doteq 1.459$,
$H(X,Y) \doteq 2.585$.
Model s nezávislými veličinami a stejnými marginálními rozděleními:
$p_X(x)\cdot p_Y(y)$:

x \ y	A	B	C	$p_X(x)$
1	$\frac{1}{9}$	$\frac{1}{6}$	$\frac{1}{18}$	$\frac{1}{3}$
2	$\frac{1}{9}$	$\frac{1}{6}$	$\frac{1}{18}$	$\frac{1}{3}$
3	$\frac{1}{9}$	$\frac{1}{6}$	$\frac{1}{18}$	$\frac{1}{3}$
$p_Y(y)$	$\frac{1}{3}$	$\frac{1}{2}$	$\frac{1}{6}$	1

$I(X;Y) = D(\boldsymbol{p}_{X,Y}||\boldsymbol{p}_X\cdot\boldsymbol{p}_Y) = 2\cdot\frac{1}{6}\log\frac{3}{2} + \frac{1}{6}\log 3 \doteq 0.459$, *též*
$I(X;Y) = H(X) + H(Y) - H(X,Y) \doteq 1.585 + 1.459 - 2.585 = 0.459$.

Kdybychom věděli, že $Y = y$, pak by X přispívalo podm. entropií $H(X|Y = y)$:

$p_{X|Y}(x|y)$:

x \ y	A	B	C
1	$\frac{1}{2}$	$\frac{1}{3}$	0
2	$\frac{1}{2}$	$\frac{1}{3}$	0
3	0	$\frac{1}{3}$	1
$p_Y(y)$	$\frac{1}{3}$	$\frac{1}{2}$	$\frac{1}{6}$
$H(X\|Y = y)$	1	$\log 3$	0

střední hodnota $H(X|Y) = \frac{1}{3} + \frac{1}{2}\log 3 \doteq 1.126$, *též*
$H(X|Y) = H(X,Y) - H(Y) \doteq 2.585 - 1.459 = 1.126$.

Kdybychom věděli, že $X = x$, pak by Y přispívalo podm. entropií $H(Y|X = x)$:

$p_{Y|X}(y|x)$:

x \ y	A	B	C	$p_X(x)$	$H(Y\|X = x)$
1	$\frac{1}{2}$	$\frac{1}{2}$	0	$\frac{1}{3}$	1
2	$\frac{1}{2}$	$\frac{1}{2}$	0	$\frac{1}{3}$	1
3	0	$\frac{1}{2}$	$\frac{1}{2}$	$\frac{1}{3}$	1

střední hodnota $H(Y|X) = 1$, *též*
$H(Y|X) = H(X,Y) - H(X) \doteq 2.585 - 1.585 = 1$.

Literatura

[Roman] Roman, S.: *Coding and Information Theory.* Graduate Texts in Mathematics, Vol. 134, Springer, 1992.

[Moser] Moser, S.M.: *Information Theory.* National Chiao Tung University, Hsinchu, Taiwan, 2012.

[MacKay] MacKay, D.J.C.: *Information Theory, Inference, and Learning Algorithms.* Cambridge University Press, 2003.

[Cover, Thomson] Cover, T.M., Thomson, J.: *Elements of Information Theory.* Wiley, 1991.

[Vajda] Vajda, I.: *Teorie informace.* ČVUT, Praha, 2004.

[MN: PMS] Navara, M.: *Pravděpodobnost a matematická statistika.* ČVUT, Praha, 2007.

[Apl. mat.] kol.: *Aplikovaná matematika.* Oborové encyklopedie SNTL, Praha, 1978.

[Enc. Math.] Hazewinkel, M.: *Encyclopaedia of Mathematics.* Kluwer Academic Publishers, 1995.