Entropie je míra neurčitosti (chaosu).
Geometricky říká jak daleko jsme od čistoty.

Příklad:

Házíme mincí: panna – pravděp. p ··· 1
orel – " – $1-p$ ··· 0

$$\vec{p} = (p, 1-p)$$

Předpokládejme, že pravděp. vektor je racionální

$$\vec{p} = \left(\frac{k}{n}, \frac{n-k}{n}\right)$$

Jaká je pravděpodobnost, že při n nezávislých hodech padne nejdříve $k \times$ panna a pak $n-k$ orel

neboli

$$P = P(\underbrace{1 \dots 1}_{k} 0 \dots 0)$$

$$P = p^k (1-p)^{n-k} =$$

$$= 2^{\log\left[p^k (1-p)^{n-k}\right]} =$$

$$= 2^{k \log p + (n-k)\log(1-p)} =$$

$$= 2^{n\left[\frac{k}{n}\log\frac{k}{n} + \frac{n-k}{n}\log\frac{n-k}{n}\right]}$$

$$= 2^{-n H\left(\frac{k}{n}\right)} = 2^{-n H(p)}$$

Rekurzivní vlastnost entropie:

coarse graining

Je-li $\vec{p} = (p_1, \dots, p_n)$ pravděpodobnostní vektor, pak

$$\boxed{H(p_1, \dots, p_n) = H(p_1+p_2, p_3, \dots, p_n) + (p_1+p_2) \cdot H\left(\frac{p_1}{p_1+p_2}, \frac{p_2}{p_1+p_2}\right)}$$

Dk: LHS:

$$-(p_1+p_2)\log(p_1+p_2) - p_3 \log p_3 - \dots - p_n \log p_n$$
$$+(p_1+p_2)\left[-\frac{p_1}{p_1+p_2}\log\frac{p_1}{p_1+p_2} - \frac{p_2}{p_1+p_2}\log\frac{p_2}{p_1+p_2}\right]$$
$$= -(p_1+p_2)\log(p_1+p_2) - p_3\log p_3 - \dots - p_n \log p_n$$
$$- p_1 \log p_1 + p_1 \log(p_1+p_2) - p_2 \log p_2$$
$$+ p_2 \log(p_1+p_2) =$$
$$= H(p_1, \dots, p_n)$$

Důsledek:

$$H(p_1+p_2, p_3, \dots, p_n) \leq H(p_1, p_2, \dots, p_n)$$

rovnost nastane $\Leftrightarrow$ $p_1 = p_2$

$$\vec{p} = \left(\frac{500}{1000}, \frac{500}{1000}\right) \Rightarrow P = 2^{-1000\,1/2}$$

Například $\vec{p} = \left(\frac{501}{1000}, \frac{499}{1000}\right)$

$n = 1000$ hodů

$$P = 2^{-n H\left(\frac{501}{1000}\right)}$$

Příklad: Máme dvě hrací kostky s pravděpodobnostmi

$$\vec{p} = \left(2^{-3}, 2^{-3}, 2^{-3}, 2^{-3}, 2^{-3}, 3 \cdot 2^{-3}\right)$$

$$\vec{q} = \left(2^{-2}, 2^{-2}, 2^{-3}, 2^{-3}, 2^{-3}, 2^{-3}\right)$$

Preferujete-li symetrii, kterou kostku si vyberete?
Je výhodnější kostku vylosovat?

Řešení: $H(\vec{p}) = 2{,}406 = 15/8 + \frac{3}{8} \log 3/8$

$$H(\vec{q}) = 2{,}5$$

Lepší je druhá!
Losujeme-li, máme šanci 50 na 50
a dostaneme rozdělení

$$\vec{r} = \frac{1}{2}\vec{p} + \frac{1}{2}\vec{q} = \left(\frac{3}{16}, \frac{3}{16}, \frac{2}{16}, \frac{2}{16}, \frac{2}{16}, \frac{4}{16}\right)$$

$$H(\vec{r}) = \frac{25 - 3\log 3}{8} = 2{,}531.$$

(Později uvidíme odhad pro směs)

Cvičení: X je náh. veličina s konečně hodnotami v $\mathcal{X}$. (ale $X = x$)
Pak
$$H(g(X)) \leq H(X)$$
kdy platí rovnost?

Řešení: · dokázat elementárně
· důsledek rekurze
· rovnost $\Leftarrow$ g je prostá!

Např.: $H(X) = H(e^X)$.

Existují další entropie. Nejslavnější

Rényi: $q > 0$ $\quad \vec{p} = (p_1, \ldots, p_n)$
$$\frac{1}{1-q} \ln \left(\sum_{i=1}^{N} p_i^q \right)$$

Havrda-Charvát:
$$S_q^{HC} = \frac{1}{1-q} \left[\sum_{i=1}^{N} p_i^q - 1 \right]$$
pro $q \to 1$ dostaneme Shannon.

Informační divergence

Vyhodnocuje vzdálenost 2 pravděpodobnostních distribucí

Mnoho aplikací: například cena za špatně zvolený kód.

Definice: Máme dva pravděpodobnostní vektory $\vec{p}$ a $\vec{q}$ stejné délky

$$\vec{p} = (p_1, p_2, \dots, p_n)$$

$$\vec{q} = (q_1, q_2, \dots, q_n).$$

Informační divergence rozdělení $\vec{p}$ a $\vec{q}$ je číslo

$$D(\vec{p} \,\|\, \vec{q}) = \begin{cases} \infty & \text{existuje-li } i \text{ tak, že } q_i = 0 \text{ a } p_i \neq 0 \\ \sum_{i=1}^{n} p_i \log \frac{p_i}{q_i} \end{cases}$$

"$0 \cdot \log \frac{0}{0} = 0$"

Též: relativní entropie; Kullback-Leiblerova distance

Důležitá vlastnost: $D(\vec{p} \| \vec{p}) = 0$.

<u>Příklad</u>: $\vec{p} = (1/2, 1/2)$ – fair coin

$\vec{q} = (1, 0)$ – obě strany mince jsou stejné

$$D(\vec{p} \| \vec{q}) = \infty$$

máme-li jednu stranu, můžeme házet samé panny a prohlásit to za náhodu

$$D(\vec{q} \| \vec{p}) = 1 \cdot \log 2 + 0 \cdot \log 0 = 1$$

<u>Příklad</u>: $D(\vec{p} \| \vec{q})$

$\vec{p} = (p_1, \dots, p_n)$

$\vec{q} = (1/n, 1/n, \dots, 1/n)$ – uniformní

$$D(\vec{p} \| \vec{q}) = p_1 \log n p_1 + p_2 \log n p_2 + \dots + p_n \log n p_n =$$

$$= p_1 \log p_1 + p_2 \log p_2 + \dots + p_n \log p_n + (p_1 + \dots + p_n) \log n = -H(\vec{p}) + \log n.$$

$$\boxed{\begin{aligned}&\underline{\text{Věta}}:\ D(\vec{p}\,\|\,\vec{q}) \geq 0\\ &\text{a}\ \ D(\vec{p}\,\|\,\vec{q}) = 0 \iff \vec{p} = \vec{q}\end{aligned}}$$

<u>Dk</u>: Funkce $f(t) = t \log t$ je ryze konvexní na $(0, \infty)$ (neboť $f''(t) > 0, \forall t > 0$)

Vyloučíme nekonečnou hodnotu

$$D(\vec{p}\,\|\,\vec{q}) = \sum_{i=1}^{n} p_i \log \frac{p_i}{q_i} =$$

$$= \sum_i q_i \frac{p_i}{q_i} \log \frac{p_i}{q_i} =$$

$$= \sum_{i=1}^{n} q_i f\left(\frac{p_i}{q_i}\right) \geq$$

$$f\left(\sum_{i=1}^{n} q_i \frac{p_i}{q_i}\right) = f\left(\sum_{i=1}^{n} p_i\right) =$$

$$= 1 \log 1 = 0.$$

Druhá část z ryzí konvexity. Je-li $\vec{q}$ pak $D(\vec{p}\,\|\,\vec{q}) = 0$

$\Rightarrow \vec{p} = \vec{q}$ (jinak je divergence ∞)

Podmíněná entropie

Značení:

X n.v. s konečným oborem hodnot $\mathcal{X}$

Y n.v. s konečným oborem hodnot $\mathcal{Y}$

$P : \mathcal{X} \times \mathcal{Y} \mapsto \langle 0,1)$ sdružená pravděp. funkce

$$P(x,y) = P[X=x, Y=y]$$

$$P(x|y) = \frac{1}{P(y)} P(x,y)$$

Podíváme se nad sdruženou entropii

$$H(X,Y) = -\sum_{\substack{x\in\mathcal{X}\\ y\in\mathcal{Y}}} P(x,y)\log P(x,y)$$

např. jsou-li X a Y nezávislé pak

$$H(X,Y) = -\sum_{\substack{x\in\mathcal{X}\\ y\in\mathcal{Y}}} P(x)P(y)\log P(x)P(y)$$

$$= -\sum_{\substack{x\in\mathcal{X}\\ y\in\mathcal{Y}}} P(x)P(y)\log P(x)$$

$$-\sum_{\substack{x\in\mathcal{X}\\ y\in\mathcal{Y}}} P(x)P(y)\log P(y)$$

$$= H(X) + H(Y).$$

"Chaos dvou nezávislých n.v je roven součtu chaosů."

Jaký je vztah $H(X,Y)$, $H(X)$, $H(Y)$ v obecném případě?
Odpověď: liší se o podmíněnou entropii.

<u>Definice</u>: Pro $x \in X$ s $P(x) \neq 0$ definujeme <u>podmíněnou entropii</u>

$$H(Y/X=x) = -\sum_{y \in Y} P(y/x) \cdot \log P(y/x)$$

jedná se o entropii rozdělení

$$\left(\frac{P(x,y)}{P(x)} \right)_{y \in Y}$$

<u>Střední podmíněná entropie</u>

$$H(Y/X) = \sum_{\substack{x \in X \\ P(x) \neq 0}} P(x) H(Y/X=x)$$

- je to konvexní kombinace entropií $H(Y/X=x)$. Tedy kombinace entropií řádkových vektorů v matici přechodu $X \to Y$.

Ekvivalentní vyjádření

$$H(Y|X) = -\sum_{x\in\mathcal{X}} P(x) \sum_{y\in\mathcal{Y}} P(y|x) \log P(y|x)$$

$$= -\sum_{x\in\mathcal{X}} \sum_{y\in\mathcal{Y}} P(x) \frac{P(x,y)}{P(x)} \log \frac{P(x,y)}{P(x)}$$

$$= -\sum_{x\in\mathcal{X}} \sum_{y\in\mathcal{Y}} P(x,y) \log \frac{P(x,y)}{P(x)} =$$

$$-\sum_{x\in\mathcal{X}} \sum_{y\in\mathcal{Y}} P(x,y) \cdot \log P(y|x).$$

<u>Věta</u>: Řetězové pravidlo

$$\boxed{H(X,Y) = H(X) + H(Y|X)}$$

$H(X)$: informace o X (závisí na distribuci X)

$H(Y|X)$: informace o Y skrze X (závisí na přechodové matici)

Důkaz řetězového pravidla:

$$H(X,Y) = -\sum_{\substack{x\in\mathcal{X}\\ y\in\mathcal{Y}}} P(x,y)\log P(x,y)$$

$$= -\sum_{\substack{x\in\mathcal{X}\\ y\in\mathcal{Y}}} P(x,y)\log P(x)\cancel{\log} P(y|x) \quad .$$

$$= -\sum_{\substack{x\in\mathcal{X}\\ y\in\mathcal{Y}}} P(x,y)\log P(x)$$

$$-\sum_{\substack{x\in\mathcal{X}\\ y\in\mathcal{Y}}} P(x,y)\log P(y|x)$$

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

Důsledek: • $H(X) \leq H(X,Y)$

neboli sdružením nemůžeme snížit entropii.

• X, Y nezávislé $\Rightarrow H(Y|X) = H(Y)$

↓

též elementárně z definice: všechny řádky přechodové matice budou stejné a to rovny prav. vektoru Y.

Příklad 1.3 Nechť (X, Y) je zdroj s následujícím rozdělením pravděpodobnosti $p(x, y)$:

$x \setminus y$	1	2	3	4
1	$\frac{1}{8}$	$\frac{1}{16}$	$\frac{1}{16}$	$\frac{1}{4}$
2	$\frac{1}{16}$	$\frac{1}{8}$	$\frac{1}{16}$	0
3	$\frac{1}{32}$	$\frac{1}{32}$	$\frac{1}{16}$	0
4	$\frac{1}{32}$	$\frac{1}{32}$	$\frac{1}{16}$	0

Pro marginální distribuce $p(x), p(y)$ na množinách $\mathcal{X} = \mathcal{Y} = \{1, 2, 3, 4\}$ platí

$$p(x) = \left(\frac{1}{2}, \frac{1}{4}, \frac{1}{8}, \frac{1}{8}\right) \quad \text{resp.} \quad p(y) = \left(\frac{1}{4}, \frac{1}{4}, \frac{1}{4}, \frac{1}{4}\right),$$

takže

$$H(X) = 7/4 \ bitu, \quad H(Y) = 2 \ bity.$$

Dále

$$\begin{aligned}
H(X \mid Y) &= \sum_{y \in \mathcal{Y}} H(X \mid Y = y)\, p(y) \\
&= \frac{1}{4} H\left(\frac{1}{2}, \frac{1}{4}, \frac{1}{8}, \frac{1}{8}\right) + \frac{1}{4} H\left(\frac{1}{4}, \frac{1}{2}, \frac{1}{8}, \frac{1}{8}\right) \\
&+ \frac{1}{4} H\left(\frac{1}{4}, \frac{1}{4}, \frac{1}{4}, \frac{1}{4}\right) + \frac{1}{4} H(1, 0, 0, 0) \\
&= \frac{1}{4} \cdot \frac{7}{4} + \frac{1}{4} \cdot \frac{7}{4} + \frac{1}{4} \cdot 2 + \frac{1}{4} \cdot 0 \\
&= \frac{11}{8} \ bitu.
\end{aligned}$$

→ pro [illegible] entropie X a Y
za [illegible]

Podobně $H(Y \mid X) = 13/8$ bitu a $H(X, Y) = 27/8$ bitu. Z tohoto příkladu je tedy vidět, že obecně

$$H(Y \mid X) \neq H(X \mid Y),$$

tj. že neurčitost zprávy Y při dané X je obecně jiná, než neurčitost zprávy X při dané Y.

Spočítat.

$$H(X, Y) = H(X) + H(Y \mid X) = 7/4 + 13/8 = \frac{14 + 13}{\text{známe } 8} = \frac{27}{8}$$

$$H(X, Y) = H(Y) + H(X \mid Y) = 2 + \frac{11}{8} = \frac{16 + 11}{8} = \frac{27}{8}$$

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

neboli $H(X/Y)$ [illegible] $H(Y/X)$ ale nejsou stejné!

Kdy je $H(Y|X)=0$?

$\Leftrightarrow H(X,Y)=H(X)$ tj. nedojde k růstu entropie

$\Leftrightarrow H(Y|X=x)=0$ pro každé $x\in\mathcal{X}$ s $P(x)>0$.

$\Leftrightarrow \left(\frac{P(x,y)}{P(x)}\right)_{y\in Y}$ je Diracův $\forall x\in\mathcal{X}$ s $P(x)>0$

$\Leftrightarrow$ existuje funkce $g:\mathcal{X}\to Y$ tak, že $Y=g(X)$

Obecné řetězové pravidlo pro n-rozměrné vektory

<u>Věta</u>: Ať $X_1,\dots,X_n$ jsou náhodné veličiny s konečně mnoha hodnotami a sdruženou pravděpodobnostní funkcí $P(x_1,x_2,\dots,x_n)$.

Pak

$$H(X_1,\dots,X_n)=$$

$$H(X_1)+H(X_2|X_1)+H(X_3|X_2,X_1)+\dots+H(X_n|X_{n-1},\dots,X_1).$$

Zebo-li

$$H(X_1, X_2, \dots, X_n) = \sum_{i=1}^{n} H(X_i / X_{i-1}, \dots, X_1)$$

<u>Důkaz</u>:

$$P(x_1, \dots, x_n) = P(x_1)\, P(x_2/x_1) \cdot P(x_3 / x_1, x_2) \cdots$$
$$= \prod_{i=1}^{n} P(x_i / x_{i-1}, \dots, x_1)$$

$$H(X_1, X_2, \dots, X_n) =$$
$$= -\sum_{x_1, \dots, x_n} P(x_1, x_2, \dots, x_n) \log P(x_1, x_2, \dots, x_n)$$
$$= -\sum_{x_1, \dots, x_n} P(x_1, x_2, \dots, x_n) \log \prod_{i=1}^{n} P(x_i / x_{i-1}, \dots x_1)$$
$$= -\sum_{i=1}^{n} \sum_{x_1, x_2, \dots, x_n} P(x_1, x_2, \dots, x_n) \log P(x_i / x_{i-1}, \dots, x_1)$$

sečíst přes $x_{i+1}, \dots x_n$ dá "marginální" rozdělení

$$= -\sum_{i=1}^{n} \sum_{x_1, x_2, \dots, x_i} P(x_1, x_2, \dots x_i) \log P(x_i / x_{i-1}, \dots x_1)$$
$$= \sum_{i=1}^{n} H(X_i / X_{i-1}, \dots, X_1).$$

Důsledek: $X_1, X_2, \ldots, X_n$ jsou nezávislé

právě tehdy když

$$(*) \quad H(X_1, \ldots, X_n) = H(X_1) + H(X_2) + \cdots + H(X_n)$$

(opačná implikace za domácí)

Jaký je obecně vztah mezi levou a pravou stranou rovnosti $(*)$.

Odpověď: Vzájemná informace

Vzájemná informace

opět X a Y s konečnými obory hodnot $\mathcal{X}$ a $\mathcal{Y}$

Definice: Vzájemná informace $I(X,Y)$ je číslo

$$I(X,Y) = \sum_{x,y} P(x,y) \log \frac{P(x,y)}{P(x)P(y)}$$

Z ... náhodný vektor s marginálními rozděleními, stejnými jako X a Y ale s nezávislými složkami

Pak $I(X,Y) = D((X,Y)||Z)$.

---

z vlastností informační divergence vyplývá, že

- $I(X,Y) \geq 0$
- $I(X,Y) = I(Y,X)$
- $I(X,Y) = 0 \Leftrightarrow X$ a Y jsou nezávislé.

BOBO

Dále platí:

$I(X,X) = H(X)$.

$$(X,X) \cdots P(x,y) = \begin{cases} 0 & x \neq y \\ P(x) & x = y \end{cases}$$

$$I(X,Y) = \sum_{x,y} P(x,y) \log \frac{P(x,y)}{P(x)P(y)}$$

$$= \sum_{x} P(x) \log \frac{1}{P(x)} = H(X).$$

$$\boxed{\begin{array}{l} \underline{\text{Tvrzení}} \\ I(X,Y) = H(X) - H(X|Y). \end{array}}$$

<u>Důkaz</u>: $I(X,Y) =$

$$= \sum_{x,y} P(x,y) \log \frac{P(x,y)}{P(x)P(y)} =$$

$$= \sum_{x,y} P(x,y) \log \frac{P(x|y)}{P(x)} =$$

$$= -\sum_{x,y} P(x,y) \log P(x) + \sum_{x,y} P(x,y) \log P(x|y)$$

$$= H(X) - H(X|Y)$$

<u>Důsledek</u>: (1) $H(X|Y) \leq H(X)$

podmíněná entropie $\leq$ entropie

(2) $H(X|Y) = H(X) \Leftrightarrow I(X,Y) = 0$

$\Leftrightarrow$ X a Y jsou nezávislé

(máme tedy škálu od determinismu k nezávislosti)

Identity:

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

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

$= H(X) + H(Y) - I(X,Y)$.

<u>Důsledek</u>

je již dříve

Obecněji platí pro náhodné veličiny $X_1, \dots, X_n$ (s konečně mnoha hodnotami), že

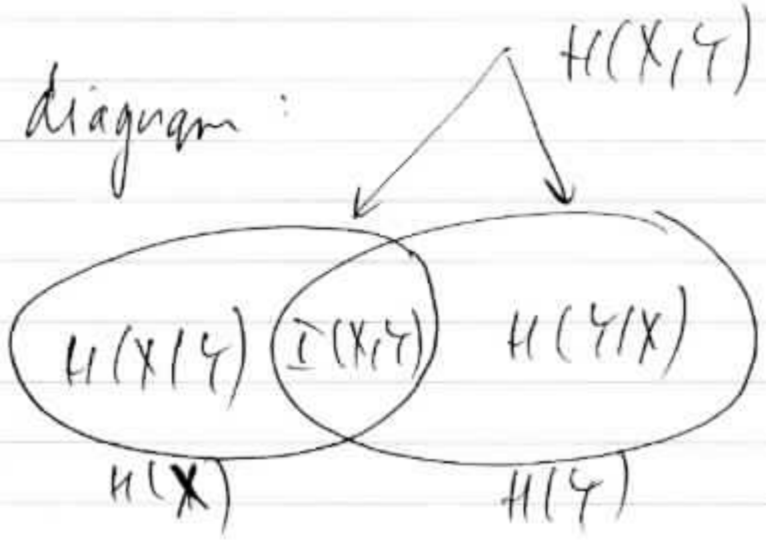
$$H(X_1, \dots, X_n) \le H(X_1) + \dots + H(X_n)$$

přičemž rovnost nastává $\Leftrightarrow$ $X_1, \dots, X_n$ jsou nezávislé.

viz výše.

Dk: Řetězové pravidlo + podmíněná entropie je vždy menší

Diagram:



4. Uvažujme abecedu $\Lambda = \Omega = \{0, 1\}$. Náhodná veličina X má binomické rozdělení $\mathrm{Bi}(1, 3/4)$, kde $P[X = 1] = 3/4$. Šum je vyjádřen veličinou Z mající rozdělení $\mathrm{Bi}(1, 1/10)$, kde $P[Z = 1] = 1/10$. Položme $Y = X \oplus Z$, přičemž $\oplus$ je součet modulo 2. Určete vzájemnou informaci veličin X a Y.

Řešení:

Platí $I(X; Y) = H(Y) - H(Y|X)$. Stanovme nejprve $H(Y|X)$. Z definice Y dostaneme následující tabulku podmíněných pravděpodobností:

$p_{Y\|X}(y\|x)$	$y = 0$	$y = 1$
$x = 0$	$9/10$	$1/10$
$x = 1$	$1/10$	$9/10$

přechodová matice

plyne díky permutaci

Z té plyne $H(Y|X) = \frac{1}{4}H(Y|X = 0) + \frac{3}{4}H(Y|X = 1) = H(Y|X = 0) \doteq 0.47$. Dále spočítáme p_Y jako marginální rozdělení pomocí p_{XY}:

$p_{XY}(x, y)$	$y = 0$	$y = 1$
$x = 0$	$9/40$	$1/40$
$x = 1$	$3/40$	$27/40$

$y = \left(\frac{12}{40}, \frac{28}{40}\right)$

Dostaneme $H(Y) \doteq 0.88$ a tudíž $I(X; Y) \doteq 0.41$. Jelikož $H(X) = 0.81$, lze usoudit, že při přenosu se ztratí část informace o X díky šumu Z.