Informační entropie

Andrew Kozlík

KA MFF UK

Informační entropie

Definice

Nechť $X : \Omega \to A$ je diskrétní náhodná veličina.
Potom *entropie* náhodné veličiny X je definována jako

$$\mathrm{H}(X) := -\sum_{a \in A} \Pr(X{=}a) \cdot \log_2 \Pr(X{=}a),$$

kde předepisujeme $0 \cdot \log_2 0 := 0$.

Ospravedlnění předpisu pro 0:

- Limitním chováním: $\lim_{z \to 0_+} z \log_2 z = 0$.
- Selským rozumem: Je-li $X{=}a$ nemožný jev, pak by neměl zvyšovat míru nejistoty.

Vyloučení nemožných jevů

Úmluva

Nechť $X : \Omega \to A$ je diskrétní náhodná veličina.
Budeme předpokládat, že $\Pr(X{=}a) \neq 0$ pro všechna $a \in A$.

Ospravedlnění

Pokud X nesplňuje předpoklad, pak lze definovat novou náhodnou veličinu X', která jej splňuje a přitom $\mathrm{H}(X) = \mathrm{H}(X')$:

- Zadefinujeme $A' := \{\, a \in A : \Pr(X{=}a) \neq 0 \,\}$.
- Zadefinujeme zúžení $X' := X|_{A'}$.

Shannonova věta o kódování zdroje

- Entropie diskrétní náhodné veličiny je dolní mez na průměrnou délku kteréhokoliv jejího prostého kódování.

- Vždy existuje prosté kódování, kterým se lze přiblížit k této dolní mezi libovolně blízko.
  (Je však třeba kódovat více výsledků najednou.)

Vlastnosti entropie

$$\mathrm{H}(X) = -\sum_{a \in A} \Pr(X{=}a) \cdot \log_2 \Pr(X{=}a)$$

Pozorování

- Má-li X rovnoměrné rozdělení, pak

$$\mathrm{H}(X) = -\sum_{a \in A} \frac{1}{|A|} \log_2 \frac{1}{|A|} = \log_2 |A|.$$

  V případě $|A| = 2^k$ tak dostáváme $\mathrm{H}(X) = k$.

- Existuje-li $a_0 \in A$ takové, že $\Pr(X{=}a_0) = 1$, pak $H(X) = -1 \cdot \log_2 1 = 0$.
  Jinými slovy, je-li výsledek jistý, pak je entropie nulová.

Huffmanovo kódování náhodné veličiny

Máme náhodnou veličinu $X : \Omega \to A$.

- Prvkům z A přiřazujeme kódy z $\{0,1\}^*$.
- Vlastnosti Huffmanova kódování:
  - Je prosté.
  - Průměrná délka zakódovaného prvku leží v intervalu $[\mathrm{H}(X), \mathrm{H}(X)+1)$.

Příklad

Mějme $A = \{1,2,3,4,5,6\}$ a rozdělení pravděpodobnosti:

a	1	2	3	4	5	6
$\Pr(X{=}a)$	0,05	0,10	0,15	0,12	0,13	0,45

Huffmanovo kódování

- Sestrojíme binární strom.
- Listy budou prvky z A.
- Váha listu $x \in A$ bude $\Pr(X{=}a)$.
- Váha vnitřního vrcholu bude součet vah jeho podvrcholů.
- Na začátku máme množinu vrcholů A bez hran.
  (Každý z nich je kořen jednoprvkového stromu.)
- Zvolíme dva nejlehčí kořeny a připojíme je pod nový kořenový vrchol. Opakujeme dokud nezůstane jediný kořen.
- Pro každý vrchol označíme jednu vycházející hranu symbolem 0 a druhou symbolem 1.
- Huffmanův kód $h(a)$ určíme tak, že přečteme symboly na cestě od kořene k listu a.

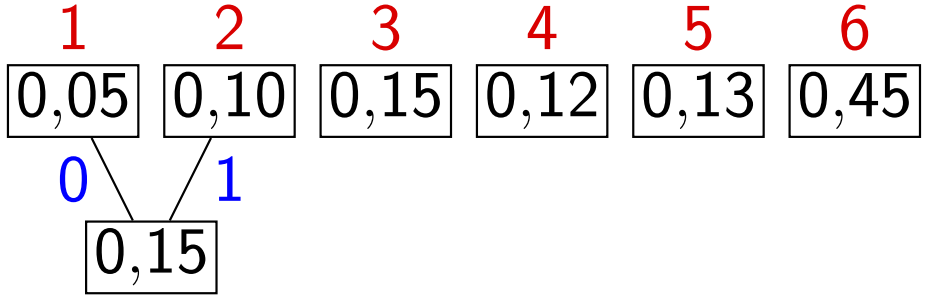
Huffmanův algoritmus.

1	2	3	4	5	6
0,05	0,10	0,15	0,12	0,13	0,45

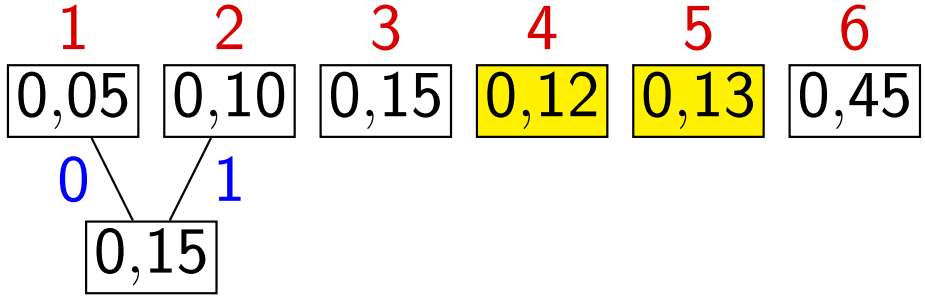
Huffmanův algoritmus.

1	2	3	4	5	6
0,05	0,10	0,15	0,12	0,13	0,45

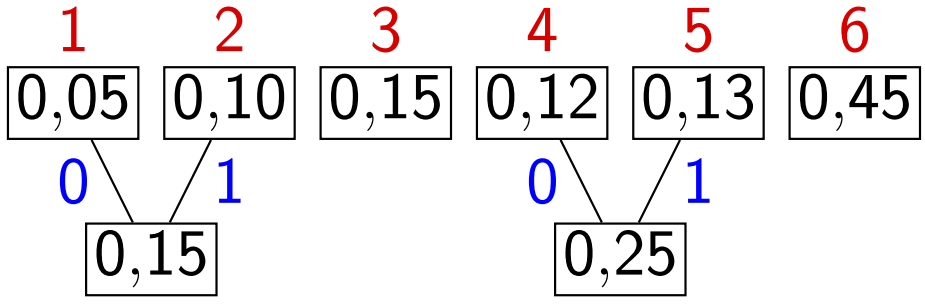
Huffmanův algoritmus.



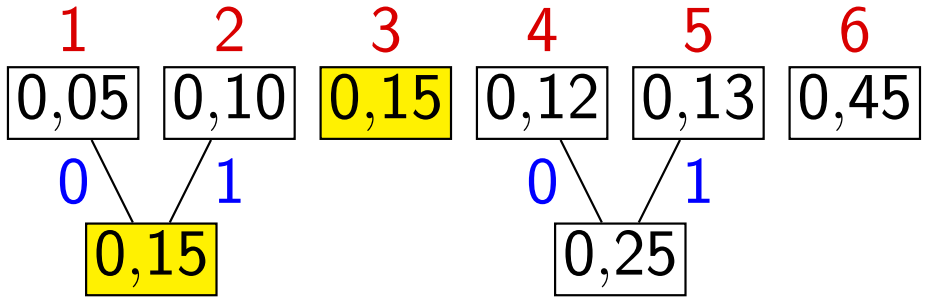
Huffmanův algoritmus.



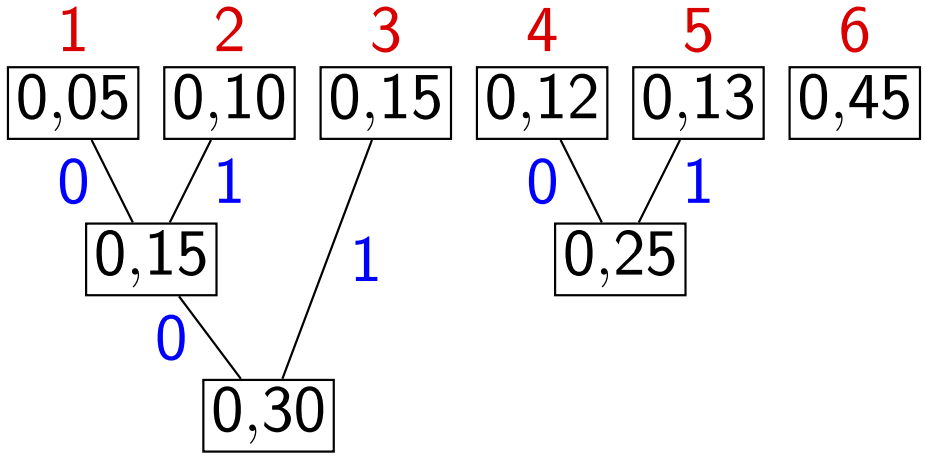
Huffmanův algoritmus.



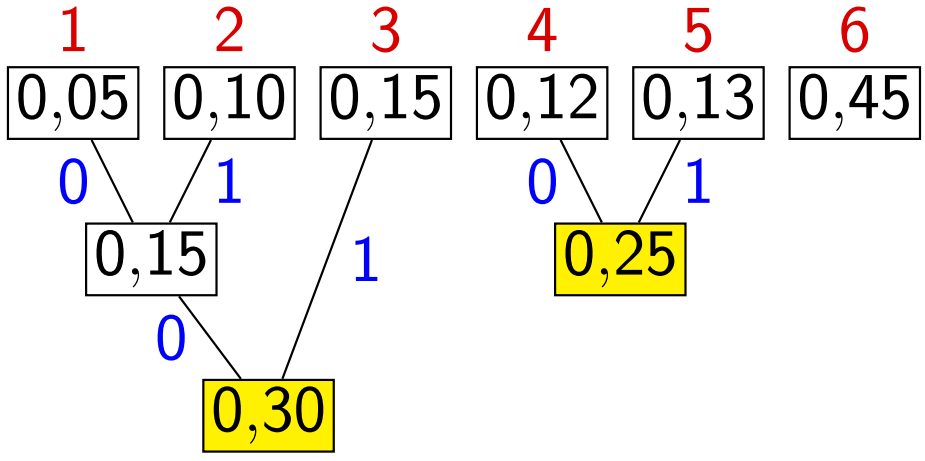
Huffmanův algoritmus.



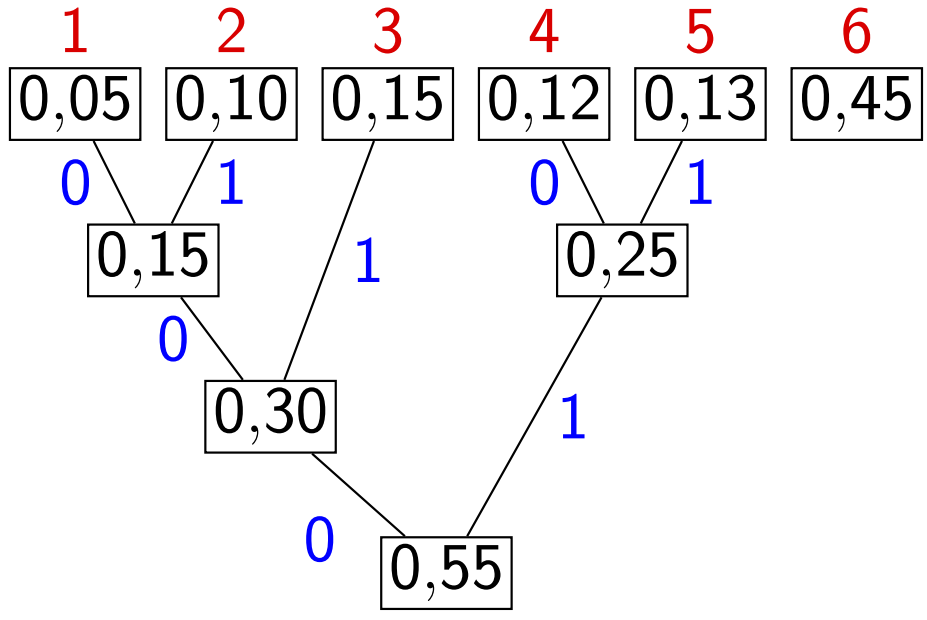
Huffmanův algoritmus.



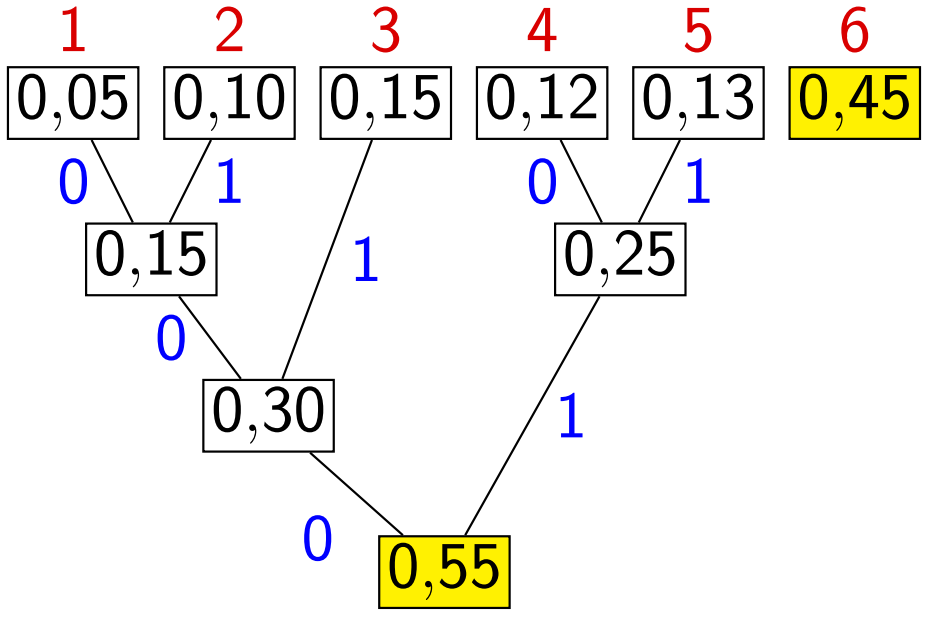
Huffmanův algoritmus.



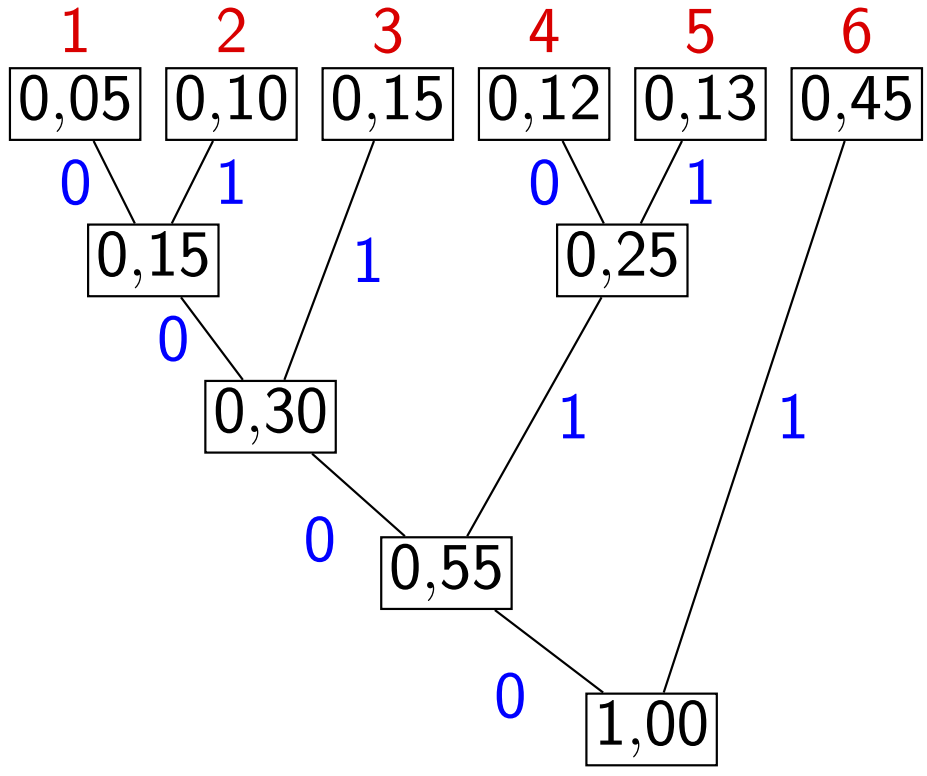
Huffmanův algoritmus.



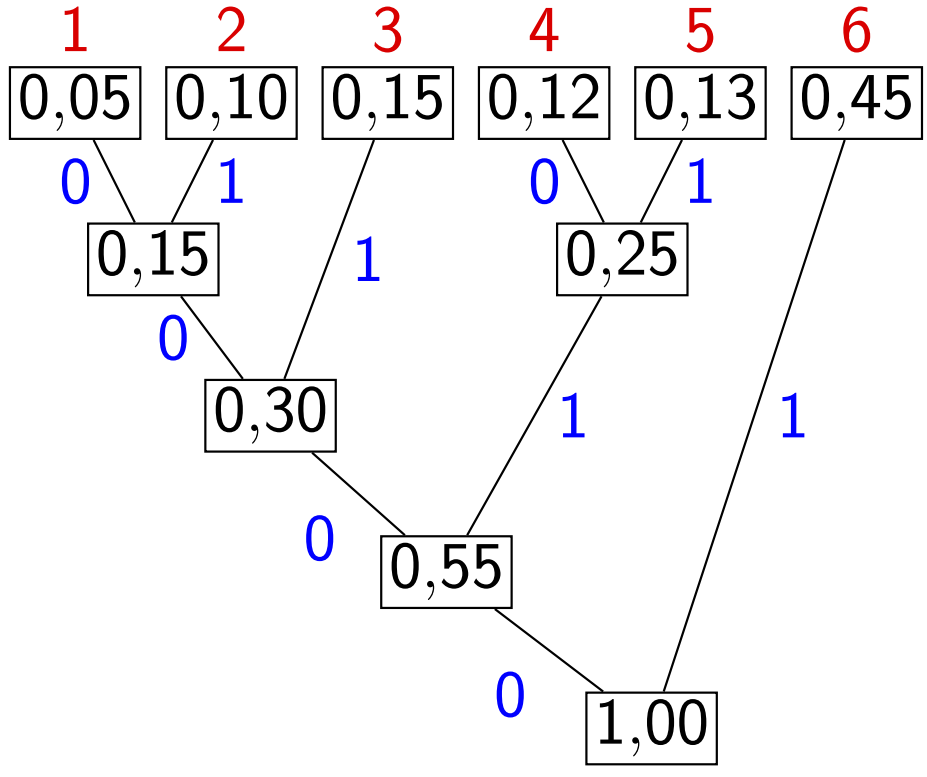
Huffmanův algoritmus.



Huffmanův algoritmus.



Huffmanův algoritmus.



a	$h(a)$
1	0000
2	0001
3	001
4	010
5	011
6	1

Entropie a střední délka kódu

a	$\Pr(X{=}a)$	$h(a)$	$\log_2 \Pr(X{=}a)$
1	0,05	0000	−4,32
2	0,10	0001	−3,32
3	0,15	001	−2,74
4	0,12	010	−3,06
5	0,13	011	−2,94
6	0,45	1	−1,15

- Délka kódu $\ell(h(a))$ je přibližně rovna $-\log_2 \Pr(X{=}a)$.
- Srovnání entropie a střední délky kódu:

$$\mathrm{H}(X) = -\sum_{a \in A} \Pr(X{=}a) \cdot \log_2 \Pr(X{=}a) \approx 2{,}23$$

$$\mathrm{E}\,\ell(h(X)) = \sum_{a \in A} \Pr(X{=}a) \cdot \ell(h(a)) = 2{,}25.$$

Věta (Jensenova nerovnost)

Nechť

- $f : I \to \mathbb{R}$ *je ryze konkávní funkce na intervalu* I,
- $x_1, \ldots, x_n \in I$,
- $\lambda_1, \ldots, \lambda_n \in (0, \infty)$,
- $\sum_{i=1}^{n} \lambda_i = 1$.

Potom

$$\sum_{i=1}^{n} \lambda_i f(x_i) \leq f\left(\sum_{i=1}^{n} \lambda_i x_i \right),$$

přičemž rovnost nastává právě tehdy, když $x_1 = x_2 = \cdots = x_n$.

Důkaz.

- V případě $n = 2$ se jedná o definici ryze konkávní funkce.
- Dále postupujeme indukcí podle n.

Věta (o maximální entropii)

Nechť $X : \Omega \to A$ je diskrétní náhodná veličina.
Potom

$$\mathrm{H}(X) \leq \log_2 |A|,$$

přičemž rovnost nastává, právě když X má rovnoměrné rozdělení.

Důkaz.

Využijeme Jensenovu nerovnost s $f = \log_2$.

$$\mathrm{H}(X) = \sum_{a \in A} \underbrace{\Pr(X{=}a)}_{\text{„}\lambda_i\text{“}} \cdot \log_2 \underbrace{\frac{1}{\Pr(X{=}a)}}_{\text{„}x_i\text{“}} \leq \log_2 \sum_{a \in A} 1.$$

Ověření předpokladů:

- Funkce $\log_2$ je ryze konkávní na $(0, \infty)$.
- Podle úmluvy je $\Pr(X{=}a) \neq 0$ pro všechna $a \in A$.
- $\sum_{a \in A} \Pr(X{=}a) = 1$.

Sdružená entropie

Definice

Nechť $X : \Omega \to A$ a $Y : \Omega \to B$ jsou diskrétní náhodné veličiny.

- *Kartézský součin* veličin X a Y definujeme jako

$$(X, Y) : \Omega \to A \times B$$
$$\omega \mapsto (X(\omega), Y(\omega)).$$

- *Sdruženou entropii* veličin X a Y definujeme jako entropii jejich kartézského součinu a značíme ji $\mathrm{H}(X, Y)$.

Pozorování

Pro libovolná $a \in A$ a $b \in B$ platí

$$\begin{aligned}
&\Pr((X, Y){=}(a, b)) \\
&\quad = \Pr(\{\, \omega \in \Omega : X(\omega) = a, Y(\omega) = b \,\}) \\
&\quad = \Pr(\{\, \omega \in \Omega : X(\omega) = a \,\} \cap \{\, \omega \in \Omega : Y(\omega) = b \,\}) \\
&\quad = \Pr(X{=}a, Y{=}b).
\end{aligned}$$

Sdružená entropie

- Sdruženou entropii lze rozepsat

$$\mathrm{H}(X,Y) = -\sum_{(a,b)\in A\times B} \Pr((X,Y)=(a,b)) \log_2 \Pr((X,Y)=(a,b))$$
$$= -\sum_{a\in A}\sum_{b\in B} \Pr(X=a, Y=b) \log_2 \Pr(X=a, Y=b).$$

- Kartézský součin a sdruženou entropii lze přirozeným způsobem rozšířit na více veličin.

- V zápisu sdružené entropie nezáleží na pořadí veličin, ani na jejich případném uzávorkování, např.:

$$\mathrm{H}(X,Y,Z) = \mathrm{H}(X,(Y,Z)) = \mathrm{H}((Z,X),Y).$$

Věta (o sdružené entropii)

Nechť $X : \Omega \to A$ a $Y : \Omega \to B$ jsou diskrétní náhodné veličiny. Potom

$$\mathrm{H}(X, Y) \leq \mathrm{H}(X) + \mathrm{H}(Y),$$

přičemž rovnost nastává, právě když X a Y jsou nezávislé.

Důkaz.

Mapa důkazu:

1. Dokážeme $\mathrm{H}(X, Y) - \mathrm{H}(X) - \mathrm{H}(Y) \leq 0$.
2. Dokážeme, že jsou-li X a Y nezávislé, pak nastává rovnost.
3. Dokážeme, že nastává-li rovnost, pak

$$\Pr(X{=}a)\Pr(Y{=}b) = \Pr(X{=}a, Y{=}b)$$

pro všechna a a b taková, že

(a) $\Pr(X{=}a, Y{=}b) \neq 0$,
(b) $\Pr(X{=}a, Y{=}b) = 0$,

Příprava důkazu

Definici entropie rozepíšeme podle důsledku věty o úplné pravděpodobnosti:

$$\begin{aligned} \mathrm{H}(X) &= -\sum_{a\in A} \Pr(X{=}a)\, \log_2 \Pr(X{=}a) \\ &= -\sum_{a\in A} \sum_{b\in B} \Pr(X{=}a, Y{=}b)\, \log_2 \Pr(X{=}a). \end{aligned}$$

Stejně tak pro Y:

$$\mathrm{H}(Y) = -\sum_{a\in A} \sum_{b\in B} \Pr(X{=}a, Y{=}b) \log_2 \Pr(Y{=}b).$$

Příprava důkazu

$$\mathrm{H}(X,Y) - \mathrm{H}(X) - \mathrm{H}(Y)$$

$$= -\sum_{a\in A}\sum_{b\in B} \Pr(X{=}a,\, Y{=}b) \log_2 \Pr(X{=}a,\, Y{=}b)$$

$$+\sum_{a\in A}\sum_{b\in B} \Pr(X{=}a,\, Y{=}b) \log_2 \Pr(X{=}a)$$

$$+\sum_{a\in A}\sum_{b\in B} \Pr(X{=}a,\, Y{=}b) \log_2 \Pr(Y{=}b)$$

$$= \sum_{\substack{a\in A,\ b\in B \\ \Pr(X=a,Y=b)\neq 0}} \Pr(X{=}a,\, Y{=}b) \log_2 \frac{\Pr(X{=}a)\Pr(Y{=}b)}{\Pr(X{=}a,\, Y{=}b)}$$

Sčítance, ve kterých $\Pr(X{=}a,\, Y{=}b) = 0$, můžeme vynechávat.

Krok 1 důkazu

Aplikujeme Jensenovu nerovnost s $f = \log_2$:

$$\mathrm{H}(X,Y) - \mathrm{H}(X) - \mathrm{H}(Y)$$

$$= \sum_{\substack{a\in A,\ b\in B \\ \Pr(X=a,Y=b)\neq 0}} \Pr(X{=}a,\, Y{=}b) \log_2 \frac{\Pr(X{=}a)\Pr(Y{=}b)}{\Pr(X{=}a,\, Y{=}b)}$$

$$\leq \log_2 \sum_{\substack{a\in A,\ b\in B \\ \Pr(X=a,Y=b)\neq 0}} \Pr(X{=}a)\Pr(Y{=}b)$$

$$\leq \log_2 \sum_{\substack{a\in A \\ b\in B}} \Pr(X{=}a)\Pr(Y{=}b)$$

$$= \log_2 \sum_{a\in A} \left(\Pr(X{=}a) \sum_{b\in B} \Pr(Y{=}b) \right)$$

$$= \log_2 \sum_{a\in A} \Pr(X{=}a) = \log_2 1 = 0.$$

Krok 2 důkazu

- Předpokládejme, že X a Y jsou nezávislé,
  tj. $\Pr(X{=}a, Y{=}b) = \Pr(X{=}a)\Pr(Y{=}b)$ pro všechna a a b.

- $\mathrm{H}(X, Y) - \mathrm{H}(X) - \mathrm{H}(Y)$

$$= \sum_{\substack{a\in A,\ b\in B \\ \Pr(X=a,Y=b)\neq 0}} \Pr(X{=}a, Y{=}b) \log_2 \frac{\Pr(X{=}a)\Pr(Y{=}b)}{\Pr(X{=}a, Y{=}b)}$$

$$= \sum_{\substack{a\in A,\ b\in B \\ \Pr(X=a,Y=b)\neq 0}} \Pr(X{=}a, Y{=}b) \log_2 1$$

$$= 0$$

Krok 3 důkazu, část (a)

- Předpokládejme, že $\mathrm{H}(X,Y) - \mathrm{H}(X) - \mathrm{H}(Y) = 0$.
- Potom v Jensenově nerovnosti nastává rovnost:

$$0 = \sum_{\substack{a\in A,\ b\in B \\ \Pr(X=a,Y=b)\neq 0}} \Pr(X{=}a, Y{=}b) \log_2 \frac{\Pr(X{=}a)\Pr(Y{=}b)}{\Pr(X{=}a, Y{=}b)}$$

$$\leq \log_2 \sum_{\substack{a\in A,\ b\in B \\ \Pr(X=a,Y=b)\neq 0}} \Pr(X{=}a)\Pr(Y{=}b) \leq 0.$$

- Podle věty o Jensenově nerovnosti je hodnota zlomku stejná pro všechna a a b, přes která sčítáme.
- Označme tuto hodnotu c. Potom:

$$0 = \sum_{\substack{a\in A,\ b\in B \\ \Pr(X=a,Y=b)\neq 0}} \Pr(X{=}a, Y{=}b) \log_2 c.$$

- Zřejmě musí být $c = 1$.

Krok 3 důkazu

- Dokázali jsme, že

$$\frac{\Pr(X{=}a)\Pr(Y{=}b)}{\Pr(X{=}a, Y{=}b)} = 1$$

pro všechna a a b taková, že $\Pr(X{=}a, Y{=}b) \neq 0$.

- Zbývá dokázat rovnost i pro případy $\Pr(X{=}a, Y{=}b) = 0$.

- Ukážeme, že

$$\Pr(X{=}a)\Pr(Y{=}b) = 0$$

pro všechna a a b taková, že $\Pr(X{=}a, Y{=}b) = 0$.

Krok 3 důkazu, část (b)

Číslo 1 rozepíšeme dvěma způsoby:

$$1 = \sum_{a\in A,\ b\in B} \Pr(X{=}a)\Pr(Y{=}b)$$

$$1 = \sum_{\substack{a\in A,\ b\in B \\ \Pr(X=a,Y=b)\neq 0}} \Pr(X{=}a, Y{=}b) = \sum_{\substack{a\in A,\ b\in B \\ \Pr(X=a,Y=b)\neq 0}} \Pr(X{=}a)\Pr(Y{=}b)$$

Využíváme výsledek z části (a)

Odečtením rovnic dostaneme:

$$0 = \sum_{\substack{a\in A,\ b\in B \\ \Pr(X=a,Y=b)=0}} \Pr(X{=}a)\Pr(Y{=}b).$$

Čili $\Pr(X{=}a)\Pr(Y{=}b) = 0$ pro všechna a a b taková, že $\Pr(X{=}a, Y{=}b) = 0$.

Podmíněná entropie

Definice

Nechť

- $X : \Omega \to A$ a $Y : \Omega \to B$ jsou diskrétní náhodné veličiny a
- $E \subseteq \Omega$ je jev takový, že $\Pr(E) \neq 0$.

Entropii veličiny X podmíněnou jevem E definujeme jako

$$\mathrm{H}(X \mid E) := -\sum_{a \in A} \Pr(X{=}a \mid E) \log_2 \Pr(X{=}a \mid E).$$

Podmíněnou entropii $\mathrm{H}(X \mid Y)$ definujeme jako

$$\mathrm{H}(X \mid Y) := \sum_{b \in B} \Pr(Y{=}b)\, \mathrm{H}(X \mid Y{=}b).$$

Tvrzení

Nechť $X : \Omega \to A$ *a* $Y : \Omega \to B$ *jsou diskrétní náhodné veličiny.*
Potom $\mathrm{H}(X \mid Y) = \mathrm{H}(X, Y) - \mathrm{H}(Y)$.

Důkaz.

Definici podmíněné entropie můžeme rozepsat

$$\begin{aligned}
&\mathrm{H}(X \mid Y) \\
&\quad = -\sum_{b \in B} \Pr(Y{=}b) \sum_{a \in A} \Pr(X{=}a \mid Y{=}b) \log_2 \Pr(X{=}a \mid Y{=}b) \\
&\quad = -\sum_{a \in A} \sum_{b \in B} \Pr(X{=}a, Y{=}b) \log_2 \Pr(X{=}a \mid Y{=}b).
\end{aligned}$$

Pokračování důkazu

- Máme tedy

$$\mathrm{H}(X \mid Y) = -\sum_{a\in A}\sum_{b\in B} \Pr(X{=}a,\, Y{=}b)\log_2 \Pr(X{=}a \mid Y{=}b).$$

- V důkazu věty o sdružené entropii jsme měli

$$\mathrm{H}(Y) = -\sum_{a\in A}\sum_{b\in B} \Pr(X{=}a,\, Y{=}b)\log_2 \Pr(Y{=}b).$$

- Spojením rovností dostáváme

$$\begin{aligned}
&\mathrm{H}(X \mid Y) + \mathrm{H}(Y) \\
&\quad = -\sum_{a\in A}\sum_{b\in B} \Pr(X{=}a,\, Y{=}b)\log_2 \underbrace{\Pr(X{=}a \mid Y{=}b)\Pr(Y{=}b)}_{\Pr(X=a,\,Y=b)} \\
&\quad = \mathrm{H}(X, Y).
\end{aligned}$$

Pozorování

Podle poslední věty a posledního tvrzení platí:

$$0 \leq \mathrm{H}(X \mid Y) = \mathrm{H}(X,Y) - \mathrm{H}(Y) \leq \mathrm{H}(X),$$

Rovnost nastává, právě když X a Y jsou nezávislé.

Důsledek

Nechť $X : \Omega \to A$ a $Y : \Omega \to B$ jsou diskrétní náhodné veličiny.
Potom:

1. $\mathrm{H}(X \mid Y) \leq \mathrm{H}(X)$,
   přičemž rovnost nastává, právě když X a Y jsou nezávislé.
2. $\mathrm{H}(Y) \leq \mathrm{H}(X,Y)$.

Definice

Nechť $X : \Omega \to A$ a $Y : \Omega \to B$ jsou diskrétní náhodné veličiny. *Vzájemnou informaci* definujeme jako

$$I(X;Y) = \mathrm{H}(X) + \mathrm{H}(Y) - \mathrm{H}(X,Y).$$

Tvrzení

Nechť $X : \Omega \to A$ a $Y : \Omega \to B$ jsou diskrétní náhodné veličiny. Potom

1. $I(X;Y) = I(Y;X)$;
2. $I(X;Y) = \mathrm{H}(X) - \mathrm{H}(X \mid Y) = \mathrm{H}(Y) - \mathrm{H}(Y \mid X)$;
3. $I(X;Y) \geq 0$,
   přičemž rovnost nastává, právě když X a Y jsou nezávislé.