

Informace a entropie náhodných veličin

Kapitola 3 po konec oddílu 3.1

Podle skript *Teorie informace – text k přednášce*
Štěpán Holub a Michal Kupsa

SkriptKompletMoodleB.pdf, strany 14–22

Obsah přednášky

1. Informační obsah a entropie

Informace o jednom výsledku a její průměrná hodnota.

2. Podmíněná informace a entropie

Kolik zbývá zjistit o Y , když známe X ?

3. Vzájemná informace a součtové vztahy

Jak vyjádřit informaci, kterou mají dvě veličiny společnou?

4. Řetězové pravidlo

Rozklad informace o celém vektoru na postupné přírůstky.

Předpoklady a značení

Pracujeme na jednom pravděpodobnostním prostoru $(\Omega, \mathcal{F}, \mathbb{P})$.
Všechny veličiny mají **konečné obory hodnot**.

$$X : \Omega \rightarrow A, \quad Y : \Omega \rightarrow B, \quad P_X(a) = \mathbb{P}(X = a).$$

Rozlišujeme dva nosiče:

$$\begin{aligned} s(P_X) &= \{a \in A : P_X(a) > 0\} \subseteq A, \\ s(X) &= \{\omega \in \Omega : X(\omega) \in s(P_X)\} \subseteq \Omega. \end{aligned}$$

- $\mathbb{P}(s(X)) = 1$. Výrok platící na nosiči platí **skoro jistě**.
- $\log = \log_2$ a informaci udáváme v **bitech**.
- V součtech pro entropii vynecháváme nulové pravděpodobnosti.

Informační obsah jevu

Definice 3.1

Pro jev $U \in \mathcal{F}$ s $\mathbb{P}(U) > 0$ definujeme

$$\mathfrak{I}(U) = -\log \mathbb{P}(U).$$

Čím méně pravděpodobný jev nastal, tím větší je jeho informační obsah.

$\mathbb{P}(U)$	1	1/2	1/8	1/3
$\mathfrak{I}(U)$ [bit]	0	1	3	$\log 3 \doteq 1,585$

Proč logaritmus?

Pro nezávislé jevy U, V kladné pravděpodobnosti platí

$$-\log \mathbb{P}(U \cap V) = -\log(\mathbb{P}(U)\mathbb{P}(V)) = \mathfrak{I}(U) + \mathfrak{I}(V).$$

Pro $\mathbb{P}(U) = 0$ informační obsah v této přednášce nedefinujeme.

Entropie rozdělení a náhodné veličiny

Definice 3.2

Pro rozdělení P na konečné množině A je

$$\mathcal{H}(P) = \sum_{a \in s(P)} P(a)(-\log P(a)).$$

Pro náhodnou veličinu X definujeme

$$\mathcal{H}(X) := \mathcal{H}(P_X) = \sum_{a \in s(P_X)} P_X(a)(-\log P_X(a)).$$

Entropie je průměrný informační obsah výsledku

Váha každého výsledku je jeho pravděpodobnost.

- $\mathcal{H}(X)$ závisí pouze na rozdělení, ne na názvech hodnot.
- $\mathcal{H}(X) \geq 0$: každý člen je nezáporný, protože $0 < P_X(a) \leq 1$.
- Pro konečné veličiny je tento součet vždy konečný.

Informační obsah jako náhodná veličina

Definice 3.3 a vztah $\mathcal{H}(X) = \mathbb{E}\mathfrak{I}_X$

Pro $\omega \in s(X)$ položíme

$$\mathfrak{I}_X(\omega) = -\log P_X(X(\omega)).$$

Jde o informační obsah jevu $\{\omega' : X(\omega') = X(\omega)\}$.

Důkaz vztahu se střední hodnotou

Na jevu $\{X = a\}$ je $\mathfrak{I}_X$ konstantní a rovna $-\log P_X(a)$. Proto

$$\mathbb{E}\mathfrak{I}_X = \sum_{a \in s(P_X)} \underbrace{P_X(a)}_{\text{váha}} \underbrace{(-\log P_X(a))}_{\text{hodnota}} = \mathcal{H}(X).$$

$\mathfrak{I}_X$ je náhodná veličina. $\mathcal{H}(X)$ je jedno číslo.



Základní příklady entropie

Jistá hodnota. Jestliže $\mathbb{P}(X = a) = 1$, pak $\mathcal{H}(X) = -\log 1 = 0$.
Obráceně: nulový součet nezáporných členů vyžaduje $P_X(a) = 1$ pro každé a v nosiči. Nosič tedy obsahuje jedinou hodnotu.

Rovnoměrné rozdělení na m hodnotách.

$$\mathcal{H}(X) = \sum_{j=1}^m \frac{1}{m} \log m = \log m.$$

Rovnoměrný binární řetězec délky n má 2^n možností a entropii n bitů.

Nerovnoměrný bit. Pro pravděpodobnosti $3/4$ a $1/4$ je

$$\mathcal{H}(X) = \frac{3}{4} \log \frac{4}{3} + \frac{1}{4} \log 4 \doteq 0,811 \text{ bitu.}$$

Počet binárních symbolů sám o sobě neurčuje jejich entropii.

Deterministická transformace

Tvrzení 3.4

Nechť $X : \Omega \rightarrow A$ a $f : A \rightarrow A'$. Potom

$$\mathfrak{I}_{f(X)} \leq \mathfrak{I}_X \text{ s.j.}, \quad \mathcal{H}(f(X)) \leq \mathcal{H}(X).$$

Je-li f injektivní, platí v obou vztazích rovnost.

Význam

Funkce f může sloučit několik hodnot X do jedné. Tím zapomeneme, která z původních hodnot nastala.

Příklad: hod spravedlivou kostkou a jeho parita

Pro $X \in \{1, \dots, 6\}$ rovnoměrné a $f(X) = X \bmod 2$ platí

$$\mathcal{H}(X) = \log 6, \quad \mathcal{H}(f(X)) = 1.$$

Prosté přejmenování stěn kostky entropii zachová.

Důkaz: transformace nezvětší informaci

Tvrzení 3.4, krok za krokem

Zvolme $\omega \in s(X)$ a označme $a = X(\omega)$. Pak

$$\{\omega' \in \Omega : X(\omega') = a\} \subseteq \{\omega' \in \Omega : f(X(\omega')) = f(a)\},$$

takže $P_{f(X)}(f(a)) \geq P_X(a) > 0$. Funkce $-\log$ je klesající, a proto

$$\mathfrak{I}_{f(X)}(\omega) = -\log P_{f(X)}(f(a)) \leq -\log P_X(a) = \mathfrak{I}_X(\omega).$$

Nerovnost platí na množině míry 1. Monotonie střední hodnoty dává

$$\mathcal{H}(f(X)) = \mathbb{E}\mathfrak{I}_{f(X)} \leq \mathbb{E}\mathfrak{I}_X = \mathcal{H}(X).$$

Je-li f injektivní, oba jevy jsou stejné a všude výše nastává rovnost.



Sdružená entropie a projekce

Dvojice $(X, Y) : \Omega \rightarrow A \times B$ je opět náhodná veličina:

$$P_{X,Y}(a, b) = \mathbb{P}(X = a, Y = b), \quad \mathcal{H}(X, Y) := \mathcal{H}(P_{X,Y}).$$

Tvrzení 3.5

$$\boxed{\mathfrak{I}_X \leq \mathfrak{I}_{X,Y} \text{ s.j.}, \quad \mathcal{H}(X) \leq \mathcal{H}(X, Y).}$$

Důkaz. Použijeme tvrzení 3.4 na veličinu (X, Y) a projekci

$$\pi_X : A \times B \rightarrow A, \quad \pi_X(a, b) = a.$$

Protože $\pi_X(X, Y) = X$, dostáváme obě nerovnosti. □

Stejně platí $\mathcal{H}(Y) \leq \mathcal{H}(X, Y)$. K určení celé dvojice potřebujeme alespoň tolik informace jako k určení složky.

Nezávislé veličiny: informace se sčítá

Tvrzení 3.6

Pro **společně nezávislé** $X_0, \dots, X_{n-1}$ a $X_{[0..n)} = (X_0, \dots, X_{n-1})$ platí

$$\mathfrak{I}_{X_{[0..n)}} = \sum_{i=0}^{n-1} \mathfrak{I}_{X_i} \quad \text{s. j.}, \quad \mathcal{H}(X_{[0..n)}) = \sum_{i=0}^{n-1} \mathcal{H}(X_i).$$

Důkaz. Pro $\omega \in s(X_{[0..n)})$ položíme $a_i = X_i(\omega)$. Nezávislost dává $P_{X_{[0..n)}}(a_0, \dots, a_{n-1}) = \prod_i P_{X_i}(a_i)$, tedy

$$\mathfrak{I}_{X_{[0..n)}}(\omega) = -\log \prod_{i=0}^{n-1} P_{X_i}(a_i) = \sum_{i=0}^{n-1} -\log P_{X_i}(a_i) = \sum_{i=0}^{n-1} \mathfrak{I}_{X_i}(\omega).$$

Druhá rovnost plyne z linearitě $\mathbb{E}$. □

Nezávislost pouze po dvojicích pro $n \geq 3$ nestačí.

Proč potřebujeme podmíněnou entropii

Pro dvě nezávislé veličiny už víme, že

$$\mathcal{H}(X, Y) = \mathcal{H}(X) + \mathcal{H}(Y).$$

Pokud ale $Y = X$, druhá složka nepřináší žádnou novou informaci:

$$\mathcal{H}(X, X) = \mathcal{H}(X).$$

Poslední rovnost plyne z injektivitě zobrazení $a \mapsto (a, a)$.

Obecný rozklad bude mít tvar

$$\mathcal{H}(X, Y) = \mathcal{H}(X) + \mathcal{H}(Y \mid X).$$

Nejprve zjistíme X . Potom zjišťujeme, co ještě zbývá určit o Y .
Nyní druhý člen přesně definujeme a rovnost dokážeme.

Podmiňování jevem

Pro $U \in \mathcal{F}$ s $\mathbb{P}(U) > 0$ definujeme

$$\mathbb{P}_U(V) := \mathbb{P}(V \mid U) = \frac{\mathbb{P}(V \cap U)}{\mathbb{P}(U)}, \quad V \in \mathcal{F}.$$

Proč dělíme? Příklad se spravedlivou šestistěnnou kostkou

Víme, že padlo sudé číslo: $U = \{2, 4, 6\}$. Ptáme se na $V = \{2\}$.

$\mathbb{P}(V \mid U) = (1/6)/(3/6) = 1/3$: tři stejně možné výsledky v U .

Dělení $\mathbb{P}(U)$ změní celkovou váhu U z $1/2$ na 1 .

Měníme pravděpodobnostní míru na stejném měřitelném prostoru.

- $\mathbb{P}_U(V) \geq 0$ a $\mathbb{P}_U(\Omega) = 1$.
- Pro disjunktní V_i jsou i $V_i \cap U$ disjunktní, a proto $\mathbb{P}_U(\bigcup_i V_i) = \sum_i \mathbb{P}_U(V_i)$.

Podmíněné rozdělení veličiny Y je tedy

$$P_{Y|U}(b) = \mathbb{P}(Y = b \mid U) = \frac{\mathbb{P}(\{Y = b\} \cap U)}{\mathbb{P}(U)}.$$

Jeho hodnoty se přes $b \in B$ sčítají na 1 .

Podmiňování hodnotou veličiny

Pro $a \in s(P_X)$ použijeme jev $U = \{X = a\}$:

$$P_{Y|X=a}(b) = P_{Y|X}(a, b) = \frac{P_{X,Y}(a, b)}{P_X(a)}.$$

- $P_{Y|X=a}$ je pro **pevné** a rozdělení na B .
- $P_{Y|X}$ označuje **rodinu** těchto rozdělení.
- Pořadí argumentů ve skriptech: nejprve a (podmínka), potom b .

Kontrola normalizace

$$\sum_{b \in B} P_{Y|X}(a, b) = \frac{\sum_{b \in B} P_{X,Y}(a, b)}{P_X(a)} = \frac{P_X(a)}{P_X(a)} = 1.$$

Při $P_X(a) = 0$ nepodmiňujeme. Při $P_{X,Y}(a, b) = 0$ vyjde hodnota 0.

Průběžný příklad: sdružené rozdělení

Příklad 3.13, první část

$P_{X,Y}$	b_0	b_1	P_X
a_0	1/2	0	1/2
a_1	1/4	1/4	1/2
P_Y	3/4	1/4	1

Marginální rozdělení

$$P_X(a) = \sum_b P_{X,Y}(a, b),$$

$$P_Y(b) = \sum_a P_{X,Y}(a, b).$$

Součty přes zbývající složku dávají marginální rozdělení.

Entropie spočítáme přímo z definice:

$$\mathcal{H}(X) = \frac{1}{2} \cdot 1 + \frac{1}{2} \cdot 1 = 1,$$

$$\mathcal{H}(Y) = \frac{3}{4} \log \frac{4}{3} + \frac{1}{4} \cdot 2 \doteq 0,811,$$

$$\mathcal{H}(X, Y) = \frac{1}{2} \cdot 1 + \frac{1}{4} \cdot 2 + \frac{1}{4} \cdot 2 = \frac{3}{2}.$$

Průběžný příklad: podmíněná rozdělení

Příklad 3.13, normalizace řádků a sloupců

Známe X

$$P_{Y|X}(a, b) = \frac{P_{X,Y}(a, b)}{P_X(a)}.$$

$X \backslash Y$	b_0	b_1
a_0	1	0
a_1	1/2	1/2

Pro $X = a_0$ je Y jisté.

Pro $X = a_1$ je Y rovnoměrný
bit.

Známe Y

$$P_{X|Y}(b, a) = \frac{P_{X,Y}(a, b)}{P_Y(b)}.$$

$X \backslash Y$	b_0	b_1
a_0	2/3	0
a_1	1/3	1

Například

$$\mathbb{P}(X = a_0 \mid Y = b_0) = \frac{1/2}{3/4} = \frac{2}{3}.$$

Vlevo má součet 1 každý řádek. Vpravo každý sloupec.

Dělíme řádkovým, resp. sloupcovým součtem původní tabulky.

Podmíněný informační obsah

Definice 3.7

Pro $\omega \in s(X, Y)$ označme $a = X(\omega)$, $b = Y(\omega)$. Definujeme

$$\begin{aligned}\mathfrak{I}_{Y|X}(\omega) &= -\log P_{Y|X}(a, b) \\ &= -\log \frac{P_{X,Y}(a, b)}{P_X(a)} = \log \frac{P_X(a)}{P_{X,Y}(a, b)}.\end{aligned}$$

- Na tomto nosiči jsou $P_{X,Y}(a, b)$ i $P_X(a)$ kladné.
- Podmíněná pravděpodobnost je nejvýše 1, takže $\mathfrak{I}_{Y|X}(\omega) \geq 0$.

V průběžném příkladu

$(X(\omega), Y(\omega))$	(a_0, b_0)	(a_1, b_0)	(a_1, b_1)
$\mathfrak{I}_{Y X}(\omega)$	$-\log 1 = 0$	$-\log(1/2) = 1$	$-\log(1/2) = 1$

Buňku (a_0, b_1) vynecháváme, protože má nulovou pravděpodobnost.

Podmínění jevem a podmínění veličinou

Definice 3.8 a 3.9

Pevný jev U , $\mathbb{P}(U) > 0$: entropie jednoho rozdělení

$$\mathcal{H}(Y \mid U) := \mathcal{H}(P_{Y|U}).$$

Zvláště $\mathcal{H}(Y \mid X = a) = \mathcal{H}(P_{Y|X=a})$ pro $a \in s(P_X)$. Počítáme s pravděpodobnostmi **za dané podmínky**.

Veličina X : průměr podmíněného informačního obsahu

$$\mathcal{H}(Y \mid X) := \mathbb{E} \mathfrak{I}_{Y|X}.$$

Střední hodnotu zde počítáme vzhledem k **původní míře $\mathbb{P}$** . Výsledek je opět jedno nezáporné číslo.

► Příklad: podmínka a průměr

Zápis $Y \mid X$ neoznačuje nově zavedenou náhodnou veličinu. Celý výraz $\mathcal{H}(Y \mid X)$ je nové značení.

► Příklad rozdílných měř

Výpočet podmíněné entropie

Tvrzení 3.10

Pro konečné veličiny X, Y platí

$$\mathcal{H}(Y | X) = \sum_{(a,b) \in s(P_{X,Y})} P_{X,Y}(a,b) \log \frac{P_X(a)}{P_{X,Y}(a,b)}.$$

Ekvivalentně:

$$\mathcal{H}(Y | X) = \sum_{a \in s(P_X)} P_X(a) \mathcal{H}(Y | X = a).$$

Význam druhé rovnosti

Pro každou možnou hodnotu a spočítáme zbývající nejistotu o Y .
Potom tyto nejistoty zprůměrujeme vahami $P_X(a)$.

Příští dva slidy dokazují obě rovnosti včetně toho, přes které nosiče sčítáme.

Důkaz 3.10: střední hodnota po dvojicích

Na celém jevu $\{X = a, Y = b\}$ je $\mathfrak{I}_{Y|X}$ konstantní:

$$\mathfrak{I}_{Y|X} = \log \frac{P_X(a)}{P_{X,Y}(a,b)}.$$

Tento jev má pravděpodobnost $P_{X,Y}(a,b)$. Tedy

$$\begin{aligned}\mathcal{H}(Y | X) &= \mathbb{E}\mathfrak{I}_{Y|X} \\ &= \sum_{(a,b) \in s(P_{X,Y})} P_{X,Y}(a,b) \log \frac{P_X(a)}{P_{X,Y}(a,b)}.\end{aligned}$$

Proč je každý člen dobře definován?

Z $P_{X,Y}(a,b) > 0$ plyne $P_X(a) \geq P_{X,Y}(a,b) > 0$. Nikde nedělíme nulou ani nepočítáme $\log 0$. □

Důkaz 3.10: vážený průměr po hodnotách X

Pro $a \in s(P_X)$ položíme $B_a = s(P_{Y|X=a})$. Z definice:

$$\mathcal{H}(Y \mid X = a) = \sum_{b \in B_a} \frac{P_{X,Y}(a, b)}{P_X(a)} \log \frac{P_X(a)}{P_{X,Y}(a, b)}.$$

Vynásobením $P_X(a)$ a sečtením přes a dostaneme

$$\begin{aligned} & \sum_{a \in s(P_X)} P_X(a) \mathcal{H}(Y \mid X = a) \\ &= \sum_{a \in s(P_X)} \sum_{b \in B_a} P_{X,Y}(a, b) \log \frac{P_X(a)}{P_{X,Y}(a, b)}. \end{aligned}$$

Sčítáme přes stejné dvojice jako v předchozím vzorci, protože

$$P_{X,Y}(a, b) > 0 \iff P_X(a) > 0 \text{ a } P_{Y|X=a}(b) > 0.$$

Výsledný součet je tedy $\mathcal{H}(Y \mid X)$.

□

Průběžný příklad: podmíněné entropie

Příklad 3.13, průměrování

Z podmíněných rozdělení $(1, 0)$ a $(1/2, 1/2)$ vychází

$$\mathcal{H}(Y \mid X = a_0) = 0, \quad \mathcal{H}(Y \mid X = a_1) = 1.$$

Proto

$$\boxed{\mathcal{H}(Y \mid X) = \frac{1}{2} \cdot 0 + \frac{1}{2} \cdot 1 = \frac{1}{2}.}$$

V opačném směru máme rozdělení $(2/3, 1/3)$ a $(0, 1)$, tedy

$$\mathcal{H}(X \mid Y = b_0) = \frac{2}{3} \log \frac{3}{2} + \frac{1}{3} \log 3 \doteq 0,918,$$

$$\mathcal{H}(X \mid Y = b_1) = 0,$$

$$\mathcal{H}(X \mid Y) = \frac{3}{4} \cdot \mathcal{H}(X \mid Y = b_0) + \frac{1}{4} \cdot 0 \doteq 0,689.$$

Obecně $\mathcal{H}(Y \mid X) \neq \mathcal{H}(X \mid Y)$.

Konkrétní podmínka může nejistotu zvýšit

V našem příkladu platí

$$\underbrace{\mathcal{H}(Y \mid X = a_0)}_0 < \underbrace{\mathcal{H}(Y)}_{\doteq 0,811} < \underbrace{\mathcal{H}(Y \mid X = a_1)}_1.$$

- Při $X = a_0$ už známe i Y : zbývající nejistota je nulová.
- Při $X = a_1$ jsou obě hodnoty Y stejně pravděpodobné: nejistota je větší než před zjištěním této podmínky.

Průměr přes podmínky je jiný údaj

$$\mathcal{H}(Y \mid X) = \frac{1}{2} \cdot 0 + \frac{1}{2} \cdot 1 = 0,5 < \mathcal{H}(Y).$$

Obecná nerovnost $\mathcal{H}(Y \mid X) \leq \mathcal{H}(Y)$ bude ve skriptech dokázána až v oddílu 3.2. Pro jednotlivou hodnotu a neplatí.

Vzájemný informační obsah

Definice 3.11

Pro $\omega \in s(X, Y)$, $a = X(\omega)$ a $b = Y(\omega)$ definujeme

$$\mathfrak{I}_{X:Y}(\omega) = \log \frac{P_{X,Y}(a, b)}{P_X(a)P_Y(b)}.$$

Porovnáváme skutečnou pravděpodobnost dvojice s tím, co by odpovídalo nezávislosti.

$$P_{X,Y}(a, b) > P_X(a)P_Y(b) \implies \mathfrak{I}_{X:Y}(\omega) > 0,$$

$$P_{X,Y}(a, b) = P_X(a)P_Y(b) \implies \mathfrak{I}_{X:Y}(\omega) = 0,$$

$$P_{X,Y}(a, b) < P_X(a)P_Y(b) \implies \mathfrak{I}_{X:Y}(\omega) < 0.$$

Tento informační obsah může být záporný

Nulová hodnota pro jednu dvojici (a, b) ještě neznamena nezávislost celých veličin.

Vzájemná entropie neboli vzájemná informace

Definice 3.12

Průměrováním vzhledem k původní míře dostaneme

$$\mathcal{I}(X : Y) := \mathbb{E}\mathfrak{I}_{X:Y} = \sum_{(a,b) \in s(P_{X,Y})} P_{X,Y}(a,b) \log \frac{P_{X,Y}(a,b)}{P_X(a)P_Y(b)}.$$

- Skripta používají název **vzájemná entropie** a symbol $\mathcal{I}$.
- Symetrie je přímo v definici: $\mathcal{I}(X : Y) = \mathcal{I}(Y : X)$.
- Jsou-li X, Y nezávislé, poměr v logaritmu je 1, tedy $\mathfrak{I}_{X:Y} = 0$ skoro jistě a $\mathcal{I}(X : Y) = 0$.

Obecná nezápornost $\mathcal{I}(X : Y)$ a obrácená implikace k nezávislosti patří až do oddílu 3.2. Znaménko průměru nelze odvodit z nezápornosti jednotlivých členů, protože ty mohou být záporné.

Průběžný příklad: kladný i záporný obsah

Příklad 3.13, přímý výpočet vzájemné informace

(a, b)	$P_{X,Y}(a, b)$	$P_X(a)P_Y(b)$	$\mathcal{I}_{X:Y}$
(a_0, b_0)	$1/2$	$3/8$	$\log(4/3) \doteq 0,415$
(a_1, b_0)	$1/4$	$3/8$	$\log(2/3) \doteq -0,585$
(a_1, b_1)	$1/4$	$1/8$	$\log 2 = 1$

Vážíme **sdruženým rozdělením** $P_{X,Y}$:

$$\mathcal{I}(X : Y) = \frac{1}{2} \log \frac{4}{3} + \frac{1}{4} \log \frac{2}{3} + \frac{1}{4} \doteq 0,311 \text{ bitu.}$$

Dvojice (a_1, b_0) nastává méně často, než by odpovídalo nezávislosti. Její záporný obsah je slučitelný s kladným průměrem.

Součtové vztahy pro dvě veličiny

Tvrzení 3.14, bez předpokladu nezávislosti

Na nosiči $s(X, Y)$ platí

$$\mathfrak{I}_{X,Y} = \mathfrak{I}_X + \mathfrak{I}_{Y|X}, \quad \mathfrak{I}_X = \mathfrak{I}_{X:Y} + \mathfrak{I}_{X|Y}.$$

Po přechodu ke středním hodnotám:

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

$$\mathcal{H}(X) = \mathcal{I}(X : Y) + \mathcal{H}(X | Y).$$

- Celou dvojici určíme tak, že zjistíme X a poté zbytek o Y .
- Informace o X se rozkládá na část společnou s Y a na část, která při znalosti Y stále chybí.

V každém vztahu můžeme zaměnit role X a Y .

Důkaz 3.14: rozklad sdruženého obsahu

Pro $\omega \in s(X, Y)$ pišme $a = X(\omega)$ a $b = Y(\omega)$. Z definice podmíněné pravděpodobnosti platí

$$P_{X,Y}(a, b) = P_X(a)P_{Y|X}(a, b).$$

Obě pravděpodobnosti v součinu jsou kladné. Použijeme $-\log$:

$$\begin{aligned}\mathfrak{I}_{X,Y}(\omega) &= -\log P_{X,Y}(a, b) \\ &= -\log P_X(a) - \log P_{Y|X}(a, b) \\ &= \mathfrak{I}_X(\omega) + \mathfrak{I}_{Y|X}(\omega).\end{aligned}$$

Linearita střední hodnoty dává

$$\mathcal{H}(X, Y) = \mathbb{E}(\mathfrak{I}_X + \mathfrak{I}_{Y|X}) = \mathcal{H}(X) + \mathcal{H}(Y | X).$$

Všechny veličiny jsou konečné, takže jsou všechny tyto střední hodnoty konečné. □

Důkaz 3.14: rozklad pomocí vzájemné informace

Pro stejná ω, a, b využijeme $P_{X|Y}(b, a) = P_{X,Y}(a, b)/P_Y(b)$:

$$\begin{aligned}\mathfrak{I}_{X:Y}(\omega) + \mathfrak{I}_{X|Y}(\omega) &= \log \frac{P_{X,Y}(a, b)}{P_X(a)P_Y(b)} + \log \frac{P_Y(b)}{P_{X,Y}(a, b)} \\ &= \log \frac{1}{P_X(a)} \\ &= \mathfrak{I}_X(\omega).\end{aligned}$$

Všechny zkrácené pravděpodobnosti jsou na nosiči kladné.
Přechodem ke středním hodnotám dostáváme

$$\boxed{\mathcal{H}(X) = \mathcal{I}(X : Y) + \mathcal{H}(X | Y).}$$

Po záměně X, Y také $\mathcal{H}(Y) = \mathcal{I}(X : Y) + \mathcal{H}(Y | X)$. □

Další podoby vzájemné informace

Odečtením v právě dokázaných vztazích získáme

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

Poslední rovnost plyne dosazením $\mathcal{H}(X | Y) = \mathcal{H}(X, Y) - \mathcal{H}(Y)$.

Úplný rozklad sdružené entropie

$$\boxed{\mathcal{H}(X, Y) = \underbrace{\mathcal{H}(X | Y)}_{\text{zbývá o } X} + \underbrace{\mathcal{I}(X : Y)}_{\text{společná část}} + \underbrace{\mathcal{H}(Y | X)}_{\text{zbývá o } Y} .}$$

Analogicky na nosiči: $\mathfrak{I}_{X:Y} = \mathfrak{I}_X + \mathfrak{I}_Y - \mathfrak{I}_{X,Y}$. U bodových obsahů může být společný člen záporný.

Průběžný příklad: kontrola součtových vztahů

Příklad 3.13, všechny hodnoty v bitech

$\mathcal{H}(X)$	1	$\mathcal{H}(Y)$	$\doteq 0,811$
$\mathcal{H}(Y \mid X)$	0,5	$\mathcal{H}(X \mid Y)$	$\doteq 0,689$
$\mathcal{H}(X, Y)$	1,5	$\mathcal{I}(X : Y)$	$\doteq 0,311$

$$\mathcal{H}(X, Y) = \mathcal{H}(X) + \mathcal{H}(Y \mid X) = 1 + 0,5 = 1,5,$$

$$\mathcal{H}(X) = \mathcal{I}(X : Y) + \mathcal{H}(X \mid Y) \doteq 0,311 + 0,689 = 1,$$

$$\mathcal{I}(X : Y) = \mathcal{H}(X) + \mathcal{H}(Y) - \mathcal{H}(X, Y) \doteq 1 + 0,811 - 1,5.$$

Stejný výsledek třemi cestami

Prímý součet logaritmů, průměr podmíněných entropií a součtové vztahy se navzájem kontrolují.

Řetězové pravidlo pro tři veličiny

Dvojici (X_0, X_1) můžeme chápat jako jednu náhodnou veličinu.

Dvakrát použijeme vztah $\mathcal{H}(U, V) = \mathcal{H}(U) + \mathcal{H}(V \mid U)$:

$$\begin{aligned}\mathcal{H}(X_0, X_1, X_2) &= \mathcal{H}(X_0, X_1) + \mathcal{H}(X_2 \mid X_0, X_1) \\ &= \mathcal{H}(X_0) + \mathcal{H}(X_1 \mid X_0) + \mathcal{H}(X_2 \mid X_0, X_1).\end{aligned}$$

Každý další přírůstek podmiňujeme všemi předchozími hodnotami

$$X_0, \quad X_1 \mid X_0, \quad X_2 \mid (X_0, X_1).$$

- $\mathcal{H}(X_2 \mid X_0, X_1)$ znamená podmínění **celou dvojicí**.
- Bez dalšího předpokladu nelze X_0 z podmínky vynechat.
- Nezávislost veličin zde nepotřebujeme.

Řetězové pravidlo pro n veličin

Tvrzení 3.15

Nechť $X_{[0..i)} = (X_0, \dots, X_{i-1})$. Pak

$$\mathfrak{I}_{X_{[0..n)}} = \sum_{i=0}^{n-1} \mathfrak{I}_{X_i | X_{[0..i)}} \quad \text{na } s(X_{[0..n)}),$$

$$\mathcal{H}(X_{[0..n)}) = \sum_{i=0}^{n-1} \mathcal{H}(X_i | X_{[0..i)}).$$

Pro $i = 0$ je předchozí vektor prázdný. Používáme konvenci

$$\mathfrak{I}_{X_0 | X_{[0..0)}} := \mathfrak{I}_{X_0}, \quad \mathcal{H}(X_0 | X_{[0..0)}) := \mathcal{H}(X_0).$$

První člen je tedy nepodmíněný. Každý další člen měří informaci nové složky při znalosti všech předchozích.

Důkaz 3.15: indukce podle délky vektoru

Základ $n = 1$. Součet obsahuje jediný člen $\mathfrak{I}_{X_0}$.

Indukční krok. Předpokládejme tvrzení pro délku n . Na nosiči $s(X_{[0..n+1)})$ použijeme tvrzení 3.14 na dvojici $(X_{[0..n)}, X_n)$:

$$\begin{aligned}\mathfrak{I}_{X_{[0..n+1)}} &= \mathfrak{I}_{X_{[0..n)}} + \mathfrak{I}_{X_n|X_{[0..n)}} \\ &= \sum_{i=0}^{n-1} \mathfrak{I}_{X_i|X_{[0..i)}} + \mathfrak{I}_{X_n|X_{[0..n)}} \\ &= \sum_{i=0}^n \mathfrak{I}_{X_i|X_{[0..i)}}.\end{aligned}$$

Indukční předpoklad smíme použít, protože nosič celého vektoru je podmnožinou nosiče jeho prefixu.

Linearita $\mathbb{E}$ dává příslušnou rovnost pro entropie.

□

Nezávislost jako zvláštní případ řetězového pravidla

Jsou-li $X_0, \dots, X_{n-1}$ společně nezávislé, pak pro každý prefix $a_{[0..i]}$ kladné pravděpodobnosti

$$P_{X_i | X_{[0..i)} = a_{[0..i)}(a_i) = \frac{\prod_{j=0}^i P_{X_j}(a_j)}{\prod_{j=0}^{i-1} P_{X_j}(a_j)} = P_{X_i}(a_i).$$

Podmíněné rozdělení se tedy neliší od nepodmíněného a

$$\mathcal{H}(X_i \mid X_{[0..i)}) = \mathcal{H}(X_i).$$

Řetězové pravidlo se zjednoduší na původní tvrzení 3.6:

$$\mathcal{H}(X_0, \dots, X_{n-1}) = \sum_{i=0}^{n-1} \mathcal{H}(X_i).$$

Bez nezávislosti podmínky v jednotlivých členech ponecháváme.

Přehled hlavních vztahů

$$\mathcal{H}(X) = \mathbb{E}\mathfrak{I}_X, \quad \mathfrak{I}_X(\omega) = -\log P_X(X(\omega)),$$

$$\mathcal{H}(Y | X) = \sum_{a \in s(P_X)} P_X(a) \mathcal{H}(Y | X = a),$$

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

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

$$\mathcal{H}(X_{[0..n)}) = \sum_{i=0}^{n-1} \mathcal{H}(X_i | X_{[0..i)}).$$

Informační obsah závisí na stavu. Entropie je jeho průměr.

U podmínění rozlišujeme pevnou hodnotu $X = a$ a průměr přes X .

Kontrolní otázky ze závěru oddílu 3.1

Skripta, strana 22. Odkazy vedou k odpovědím a důkazům.

1. Proč $\mathbb{P}(s(X)) = 1$? Kde používáme diskrétnost?

► Odpověď

2. Proč $\mathcal{H}(X) = \mathbb{E}\mathcal{I}_X$?

► Důkaz

3. Proč je v tvrzení 3.4 uvedeno „skoro jistě“?

► Odpověď

4. Proč $s(X, Y) \subseteq s(X)$?

► Důkaz

5. Proč $s(P_{X,Y}) = \{(a, b) : a \in s(P_X), b \in s(P_{Y|X=a})\}$?

► Důkaz

6. Co označují $\mathbb{P}_U$, $P_{Y|X}$, $\mathcal{H}(Y | X = a)$ a $\mathcal{H}(Y | X)$?

► Odpověď

Proč má nosič diskrétní veličiny míru jedna

Kontrolní otázka 1

Nechť je obor hodnot A konečný nebo spočetný. Pak

$$\Omega \setminus s(X) = \bigcup_{a \in A: P_X(a)=0} \{X = a\}.$$

Jde o nejvýše spočetné sjednocení disjunktních jevů míry nula. Proto

$$\mathbb{P}(\Omega \setminus s(X)) = \sum_{a \in A: P_X(a)=0} P_X(a) = 0, \quad \mathbb{P}(s(X)) = 1.$$

Role diskrétnosti: míra je spočetně aditivní. Nespočetné sjednocení nulových jevů už nemusí mít míru nula.

Například pro rovnoměrné X na $[0, 1]$ má každý jev $\{X = a\}$ pravděpodobnost 0, ale jejich sjednocení má pravděpodobnost 1. Naše definice diskrétního nosiče by zde dala prázdnou množinu.

Proč říkáme „skoro jistě“

Kontrolní otázka 3

Informační obsah $\mathfrak{I}_X(\omega) = -\log P_X(X(\omega))$ jsme definovali jen pro $\omega \in s(X)$.

Příklad. Necht $\Omega = \{u, v, w\}$ s pravděpodobnostmi $(1/2, 1/2, 0)$ a $X(u) = 0$, $X(v) = 1$, $X(w) = 2$. Potom

$$s(X) = \{u, v\}, \quad \mathfrak{I}_X(u) = \mathfrak{I}_X(v) = 1.$$

Ve stavu w bychom potřebovali $-\log 0$, což nepoužíváme.

Nerovnost $\mathfrak{I}_{f(X)} \leq \mathfrak{I}_X$ tedy formulujeme na nosiči.

Pro střední hodnotu to stačí

Doplňek nosiče má míru nula. Případné libovolné dodefinování na něm nezmění entropii ani tvrzení platící skoro jistě.

► Kontrolní otázky

Nosič dvojice a nosiče jejích složek

Kontrolní otázka 4

Nechť $\omega \in s(X, Y)$, $a = X(\omega)$ a $b = Y(\omega)$. Potom $P_{X,Y}(a, b) > 0$ a

$$P_X(a) = \sum_{b' \in B} P_{X,Y}(a, b') \geq P_{X,Y}(a, b) > 0.$$

Tedy $\omega \in s(X)$. Stejný argument platí pro Y , takže

$$s(X, Y) \subseteq s(X) \cap s(Y) \subseteq \Omega.$$

Odpovídající vztah v prostoru hodnot

$$s(P_{X,Y}) \subseteq s(P_X) \times s(P_Y) \subseteq A \times B.$$

Nosič veličiny leží v Ω . Nosič jejího rozdělení leží v oboru hodnot.

Tyto dva druhy nosičů nelze v inkluzích zaměňovat.

► Kontrolní otázky

Pevný jev mění i váhy ve střední hodnotě

Poznámka ve skriptech na straně 18

Házíme spravedlivou kostkou. Y je parita a $U = \{1, 2, 3\}$. Pod podmínkou U má sudý výsledek pravděpodobnost $1/3$ a lichý $2/3$.

$$\mathcal{H}(Y \mid U) = \frac{1}{3} \log 3 + \frac{2}{3} \log \frac{3}{2} \doteq 0,918.$$

Položme $g(\omega) = -\log P_{Y|U}(Y(\omega))$. V tomto příkladu je g definována všude, ale

$$\mathbb{E}_{\mathbb{P}} g = \frac{1}{2} \log 3 + \frac{1}{2} \log \frac{3}{2} \doteq 1,085 \neq \mathcal{H}(Y \mid U).$$

Správnou entropii dostaneme průměrováním **podmíněnou mírou**:

$$\mathcal{H}(Y \mid U) = \mathbb{E}_{\mathbb{P}_U} g.$$

Obecně navíc g nemusí být definována $\mathbb{P}$ -skoro jistě: podmíněné rozdělení může některým původně možným hodnotám dát nulu.

[► Zpět k definicím](#)

Přehled značení: pravděpodobnosti

Kontrolní otázka 6, první část

- $\mathbb{P}_U : \mathcal{F} \rightarrow [0, 1]$ je pro pevné U , $\mathbb{P}(U) > 0$, míra $V \mapsto \mathbb{P}(V \cap U)/\mathbb{P}(U)$.
- $\mathbb{P}(V \mid U)$ označuje hodnotu $\mathbb{P}_U(V)$. Výraz $V \mid U$ není jevem, na který bychom aplikovali $\mathbb{P}$.
- $\mathbb{P}(U \mid V)$ naopak vyžaduje $\mathbb{P}(V) > 0$ a rovná se $\mathbb{P}_V(U)$.
- $\mathbb{P}(Y = b \mid U) = \mathbb{P}_U(\{Y = b\}) = P_{Y|U}(b)$. Pro pevné U jde o rozdělení na B .
- $\mathbb{P}(Y = b \mid X = a) = P_{Y|X=a}(b)$ pro $P_X(a) > 0$. Pro pevné a jde o rozdělení na B .

Celá tabulka $P_{Y|X} : s(P_X) \times B \rightarrow [0, 1]$ je rodinou rozdělení. Každý řádek má součet 1.

► Kontrolní otázky

Přehled značení: entropie a střední hodnoty

Kontrolní otázka 6, druhá část

Pro konečné A je $\mathcal{H}$ zobrazení z množiny rozdělení na A do $[0, \infty)$.

Zápis	Přesný význam
$\mathcal{H}(X)$	$\mathcal{H}(P_X)$
$\mathcal{H}(X, Y)$	$\mathcal{H}(P_{X,Y})$, rozdělení na $A \times B$
$\mathcal{H}(Y \mid U)$	$\mathcal{H}(P_{Y U})$, $\mathbb{P}(U) > 0$
$\mathcal{H}(Y \mid X = a)$	$\mathcal{H}(P_{Y X=a})$, $P_X(a) > 0$
$\mathcal{H}(Y \mid X)$	$\mathbb{E}_{\mathbb{P}} \mathfrak{I}_{Y X}$
$\mathcal{I}(X : Y)$	$\mathbb{E}_{\mathbb{P}} \mathfrak{I}_{X:Y}$

Každý řádek označuje **jedno číslo**. Naproti tomu $\mathfrak{I}_X$, $\mathfrak{I}_{Y|X}$ a $\mathfrak{I}_{X:Y}$ jsou funkce stavu ω , definované skoro jistě.

► Kontrolní otázky

Jedna podmínka versus průměr

V průběžném příkladu je $P_X(a_0) = P_X(a_1) = 1/2$. Každý řádek udává rozdělení Y za uvedené podmínky:

Podmínka	$Y = b_0$	$Y = b_1$	$\mathcal{H}(Y \mid X = a)$
$X = a_0$	1	0	0 bitů
$X = a_1$	1/2	1/2	1 bit

Pevná podmínka: $\mathcal{H}(Y \mid X = a)$ vybírá jeden řádek.

Při $X = a_0$ známe Y jistě; při $X = a_1$ zbývá jeden bit nejistoty.

Podmínění veličinou: vážený průměr přes řádky

$$\mathcal{H}(Y \mid X) = \frac{1}{2} \cdot 0 + \frac{1}{2} \cdot 1 = \boxed{\frac{1}{2} \text{ bitu.}}$$

V polovině případů nezbyvá žádná nejistota, v druhé polovině jeden bit. **V průměru tedy zbývá půl bitu.**

► Zpět na slid 18