Cvičení - Entropie náhodných veličin, sdružená, podmíněná a vzájemná entropie

Michal Kupsa, Štěpán Holub

Zadání:

1. Mějme hod kostkou na které padá každé číslo stejně pravděpodobně a náhodnou veličinu X a Y, kde X nabývá hodnot S a L podle toho, zda padla sudá, nebo lichá, Y nabývá hodnot M nebo V podle toho, zda padla malá, nebo velká hodnota (1-3 vs 4-6).

   - Zapište tabulky pro $P_{X,Y}$, P_X, P_Y, $P_{X,Y}$, $P_{X|Y}$, $P_{Y|X}$.
   - Spočtěte všechny možné entropie, t.j. entropii X a Y, sdruženou, vzájemnou a obě podmíněné. Pro výslednou prezentaci můžete volit Vennův diagram (výsledek zjednodušte na logaritmy prvočísel).
   - Spočtěte entropii $\mathcal{H}(X|Y = M)$.

2. (Konkrétní Ω) V předchozím příkladu zvolte konkrétní $\Omega = \{1, \dots, 6\}$ s pravděpodobností $\mathbb{P}$ definovanou předpisem $\mathbb{P}(A) = \frac{\#A}{6}$, pro všechny podmnožiny $A \subseteq \Omega$. Definujte X a Y jako funkce na Ω a vyneste $\mathfrak{I}_X$, $\mathfrak{I}_{X|Y}$, $\mathfrak{I}_{X,Y}$ a $\mathfrak{I}_{X:Y}$ do grafu, kde Ω je nyní přirozenou součástí reálné osy x.

3. Dokažte $s(X) \subseteq s(f(X))$, $s(P_{X,Y}) \subseteq s(P_X) \times s(P_Y)$, $s(X,Y) \subseteq s(X) \cap s(Y)$.

4. Spočtete $\mathcal{H}(X|X)$ a $\mathcal{I}(X : X)$.

5. Spočtete $\mathcal{H}(X|(X,Y))$ a $\mathcal{I}((X,Y) : X)$.

6. Platí rovnost $\mathcal{H}((X,Z) \mid (X,Y)) = \mathcal{H}(Z \mid Y)$?

7. Platí rovnost $\mathcal{H}((X,Z) \mid (X,Y)) = \mathcal{H}(Z \mid (X,Y))$?

8. Platí nerovnost $\mathcal{I}((X,Z) : (X,Y)) \geq \mathcal{H}(X)$?

9. Pokud jsou Z a Y nezávislé, platí rovnost $\mathcal{I}((X,Z) : (X,Y)) = \mathcal{H}(X)$?

10. Pro obecné konečné náhodné veličiny s hodnotami v reálných číslech rozhodněte, zda platí:

(a) $\mathcal{H}(X^2 \mid X) = 0$, kde X^2 značí veličinu $\omega \mapsto (X(\omega))^2$, nikoli tedy vektor (X, X).

(b) $\mathcal{H}(X \mid X^2) = 0$

(c) $\mathcal{H}(X+Y \mid X) \geq \mathcal{H}(Y|X)$, zvažte platnost i opačné nerovnosti a rovnosti.

(d) $\mathcal{H}(X + Y \mid Z + Y) \geq \mathcal{H}(X|Z)$, zvažte platnost i opačné nerovnosti a rovnosti.

Řešení a komentář:

1. Tabulky pravděpodobností:

$P_{X,Y}$:

$X\backslash Y$	M	V
S	$1/6$	$1/3$
L	$1/3$	$1/6$

$P_{Y|X}$, $P_{X|Y}$:

$X\backslash Y$	M	V
S	$1/3$	$2/3$
L	$2/3$	$1/3$

P_X:

X	
L	$1/2$
S	$1/2$

P_Y:

X	
M	$1/2$
V	$1/2$

Stačí spočíst dvě entropie a sdruženou a pak použít součtové vzorce.

$$\mathcal{H}(X) = \mathcal{H}(Y) = \mathcal{H}\left(\frac{1}{2}, \frac{1}{2}\right) = 1,$$

$$\mathcal{H}(X, Y) = \mathcal{H}\left(\frac{1}{3}, \frac{1}{6}, \frac{1}{3}, \frac{1}{6}\right) = \frac{1}{3} + \log 3,$$

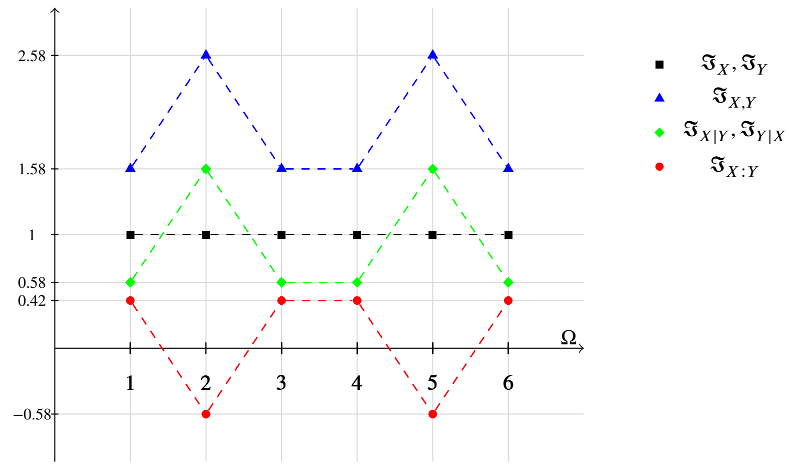
$$\mathcal{H}(X|Y) = \mathcal{H}(X, Y) - \mathcal{H}(X) = \mathcal{H}(X, Y) - \mathcal{H}(Y) = \mathcal{H}(Y|X) = \log 3 - \frac{2}{3},$$

$$\mathcal{I}(X : Y) = \mathcal{H}(X) + \mathcal{H}(Y) - \mathcal{H}(X, Y) = \frac{5}{3} - \log 3.$$

Dále

$$\mathcal{H}(X \mid M) = \mathcal{H}(X \mid V) = \mathcal{H}\left(\frac{1}{3}, \frac{2}{3}\right) = \log 3 - \frac{2}{3}.$$

2. Všechny informační obsahy, včetně podmíněných musí být nezáporné, vyjma vzájemného obsahu $\mathfrak{I}_{X:Y}$. Ten zde nabývá kladné i záporné hodnoty. Přesto jeho střední hodnota, v tomto případě aritmetický průměr, musí být nezáporná.



3. Pro pevné $\omega' \in s(X)$ z definice platí, že

$$X^{-1}(X(\omega')) := \{\omega \mid X(\omega) = X(\omega')\} \ .$$

   má nenulovou pravděpodobnost.

   Ptáme se, zda platí $\omega' \in s(f \circ X)$, tedy zda množina

$$(f \circ X)^{-1}((f \circ X)(\omega')) := \{\omega \mid (f \circ X)(\omega) = (f \circ X)(\omega')\}$$

   má také nenulovou pravděpodobnost. Protože $X(\omega) = X(\omega')$ implikuje $(f \circ X)(\omega) = (f \circ X)(\omega')$, platí inkluze

$$X^{-1}(X(\omega')) \subseteq (f \circ X)^{-1}((f \circ X)(\omega')),$$

   čímž je důkaz hotov, protože pravděpodobnost nadmnožiny je nejméně pravděpodobnost podmnožiny (z aditivity míry).

   Z definic množin $s\left(P_{X,Y}\right)$, $s\left(P_X\right)$ a $s\left(P_Y\right)$ plyne, že je třeba dokázat, že z $P_{X,Y}(a,b) > 0$ plyne, že jsou nenulové i pravděpodobnosti $P_X(a)$ a $P_Y(b)$. Jinými slovy, otázka je, zda z nenulovosti $\mathbb{P}(X = a, Y = b)$ plyne nenulovost $\mathbb{P}(X = a)$ a $\mathbb{P}(Y = b)$. To platí, neboť první podmínka je přísnější než ty druhé. Ještě přesněji, uvědomme si, že zápis $(X = a, Y = b)$ označuje množinu takových ω, pro která $X(\omega) = a \wedge Y(\omega) = b$. Podobně $(X = a)$ a $(Y = b)$ označují množinu takových ω, pro která $X(\omega) = a$, respektive $Y(\omega) = b$. Proto dokonce platí:

$$(X = a, Y = b) = (X = a) \cap (Y = b) \,.$$

   Předchozí diskuse ukázala, že třetí úloha je reformulací té druhé.

4. Tvrzení nejlépe dokážeme ze součtových vzorců za použití tvrzení $\mathcal{H}(X, X) = \mathcal{H}(X)$. Pak bude platit

$$\begin{aligned} \mathcal{H}(X \mid X) &= \mathcal{H}(X, X) - \mathcal{H}(X) = 0, \\ \mathcal{I}(X : X) &= \mathcal{H}(X) + \mathcal{H}(X) - \mathcal{H}(X, X) = \mathcal{H}(X) \,. \end{aligned}$$

Tvrzení $\mathcal{H}(X,X) = \mathcal{H}(X)$ je speciálním případem věty, podle které

$$\mathcal{H}(f \circ Y) \leq \mathcal{H}(Y),$$

a to právě proto, že rozklad daný $f \circ Y$ je hrubší než rozklad daný Y (viz také bod 3).

Pokud je f prosté, platí $(f \circ X)(\omega) = (f \circ X)(\omega')$, právě když $X(\omega) = X(\omega')$. Obě veličiny tedy definují stejný rozklad Ω a mají tudíž stejnou entropii.

Nyní si stačí uvědomit, že $(X,X) = f \circ X$ pro prosté zobrazení $f : a \mapsto (a,a)$.

Výše uvedenou úvahu zformulujeme jako následující pozorování.
Pozorování: Pokud $X = f(Y)$ a $Y = f'(X)$, pak $\mathcal{H}(X) = \mathcal{H}(Y)$

5. Použitím součtových vzorců pro X a (X,Y) v kombinaci s výše uvedeným pozorováním dostáváme

$$\mathcal{H}(X \mid (X,Y)) = \mathcal{H}(X,X,Y) - \mathcal{H}(X,Y) = 0,$$
$$\mathcal{I}(X : (X,Y)) = \mathcal{H}(X) + \mathcal{H}(X,Y) - \mathcal{H}(X,X,Y) = \mathcal{H}(X).$$

U následujících otázek se vyplatí vyzkoušet jako možný protipříklad nějakou variantu trojice veličin X, Y a $X+Y$, kde X a Y jsou nezávislé.

6. Pokud $Z = X + Y$, platí ze součtových vzorců a podle výše uvedeného pozorování

$$\mathcal{H}((X,Z) \mid (X,Y)) = \mathcal{H}(X, X+Y, X, Y) - \mathcal{H}(X,Y) = 0$$
$$\mathcal{H}(Z \mid Y) = \mathcal{H}(X+Y, Y) - \mathcal{H}(Y) = \mathcal{H}(X,Y) - \mathcal{H}(Y) = \mathcal{H}(X \mid Y).$$

Rovnost tedy obecně neplatí, stačí aby X bylo nezávislé na Y a mělo nenulovou entropii.

7. Tato rovnost platí díky součtovému vzorci a díky.

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

8. I tato nerovnost by intuitivně měla platit: veličiny (X,Z) a (X,Y) určitě sdílejí veškerou informaci obsaženou v X. Ze součtových vzorců (pro dvě a tři veličiny) platí:

$$\mathcal{I}((X,Z) : (X,Y)) = \mathcal{H}(X,Z) + \mathcal{H}(X,Y) - \mathcal{H}(X,Y,Z),$$
$$\mathcal{I}(Z : Y \mid X) = \mathcal{H}(X,Z) + \mathcal{H}(X,Y) - \mathcal{H}(X,Y,Z) - \mathcal{H}(X).$$

Vidíme, že

$$\mathcal{I}((X,Z) : (X,Y)) = \mathcal{H}(X) + \mathcal{I}(Z : Y \mid X)$$

a nerovnost plyne z nezápornosti podmíněné vzájemné entropie.

9. Z předchozího bodu vidíme, že rovnost platí právě tehdy, když jsou Z a Y nezávislé podmíněně vzhledem k veličině X. To je ale jiná podmínka než nepodmíněná nezávislost. V našem případě můžeme volit případ $Z = Y + X$, kdy jsou sice Y a Z nezávislé, ale za podmínky X jsou „zcela závislé“ a jejich vzájemná informace za podmínky X je celé $\mathcal{H}(Y \mid X) = \mathcal{H}(Z \mid X) = \mathcal{H}(Y) = \mathcal{H}(Z)$.

   Upozorněme, že neplatí ani opačná implikace. Veličiny mohou být podmíněně nezávislé, a přitom být závislé. Uvažme např. nezávislé veličiny X Y, Z, kde X a Y jsou uniformně rozdělené na $\{0, 1\}$ a Z je uniformně rozdělené na $\{0, 2\}$. Pak jsou veličiny $X + Z$ a $Y + Z$ uniformně rozdělené na $\{0, 1, 2, 3\}$ a za podmínky Z jsou nezávislé uniformně rozdělené (na $\{0, 1\}$ nebo $\{2, 3\}$). Nejsou ale nezávislé: hodnota jedné z nich omezuje definiční obor té druhé na polovinu.

10. (a) Platí $\mathcal{H}\left(X^2 \mid X\right) = \mathcal{H}\left(X^2, X\right) - \mathcal{H}(X) = \mathcal{H}(X) - \mathcal{H}(X) = 0$.

    (b) Platí $\mathcal{H}\left(X \mid X^2\right) = \mathcal{H}\left(X^2, X\right) - \mathcal{H}\left(X^2\right) = \mathcal{H}(X) - \mathcal{H}\left(X^2\right)$. Pro jakoukoli veličinu X s nenulovou entropií a s hodnotami ± 1 je entropie X^2 nulová a rovnost neplatí.

    (c) Podle pozorování výše platí rovnost.

    (d) Mezi $\mathcal{H}(X + Y \mid Z + Y)$ a $\mathcal{H}(X \mid Z)$ může být libovolný vztah. Důvodem je, že přičtení veličiny Y může mít na veličiny X a Z velmi různý vliv. Pokud je např. $Y = -X$, platí $\mathcal{H}(X + Y \mid Z + Y) = 0$, zatímco $\mathcal{H}(X \mid Z)$ může být nenulová.

    Pokud je naopak například $Z = -X$ platí $\mathcal{H}(X \mid Z) = 0$ a zbývá ověřit, že hodnota $\mathcal{H}(X + Y \mid -X + Y)$ může být nenulová. Uvažme např. veličinu X s uniformním rozdělením na $\{0, 1\}$ a na ní nezávislou veličinu Y s uniformním rozdělením na $\{1, 2\}$. Pak je $\mathcal{H}(X + Y \mid -X + Y = c) = 0$ pro $c = 0$ a $c = 2$, ale $\mathcal{H}(X + Y \mid -X + Y = 1) = 1$, protože $X + Y$ je v tomto případě se stejnou pravděpodobností 1, nebo 3. Vzhledem k tomu, že entropie za podmínky je váženým průměrem entropií pro jednotlivé hodnoty podmínky s nenulovou pravděpodobností, a vzhledem k tomu, že $-X + Y = 1$ v polovině případů, dostáváme

$$\mathcal{H}(X + Y \mid -X + Y) = 1/2.$$