Teorie kódování - Kapitola 1 - Entropie

Jan Paseka

Ústav matematiky a statistiky
Masarykova univerzita

14. února 2023

Obsah

Obsah

Motivace

FI Uvažme následující tvrzení.

A Výsledek běhu mezi dvěma rovnocennými závodníky je méně nejistý než výsledek běhu mezi šesti rovnocennými závodníky.

Motivace

FI Uvažme následující tvrzení.

A Výsledek běhu mezi dvěma rovnocennými závodníky je méně nejistý než výsledek běhu mezi šesti rovnocennými závodníky.

B Výsledek rulety je více nejistý než vrh kostkou.

Motivace

FI Uvažme následující tvrzení.

A Výsledek běhu mezi dvěma rovnocennými závodníky je méně nejistý než výsledek běhu mezi šesti rovnocennými závodníky.

B Výsledek rulety je více nejistý než vrh kostkou.

C Výsledek vrhu ideální kostkou je více nejistý než výsledek vrhu falešnou kostkou.

Motivace

FI Uvažme následující tvrzení.

A Výsledek běhu mezi dvěma rovnocennými závodníky je méně nejistý než výsledek běhu mezi šesti rovnocennými závodníky.

B Výsledek rulety je více nejistý než vrh kostkou.

C Výsledek vrhu ideální kostkou je více nejistý než výsledek vrhu falešnou kostkou.

Motivace

FI Uvažme následující tvrzení.

A Výsledek běhu mezi dvěma rovnocennými závodníky je méně nejistý než výsledek běhu mezi šesti rovnocennými závodníky.

B Výsledek rulety je více nejistý než vrh kostkou.

C Výsledek vrhu ideální kostkou je více nejistý než výsledek vrhu falešnou kostkou.

Zřejmě s výše uvedenými tvrzeními lze okamžitě souhlasit. Obtížné však bude definovat, co je to vlastně nejistota.

Motivace

FI Uvažme následující tvrzení.

A Výsledek běhu mezi dvěma rovnocennými závodníky je méně nejistý než výsledek běhu mezi šesti rovnocennými závodníky.

B Výsledek rulety je více nejistý než vrh kostkou.

C Výsledek vrhu ideální kostkou je více nejistý než výsledek vrhu falešnou kostkou.

Zřejmě s výše uvedenými tvrzeními lze okamžitě souhlasit. Obtížné však bude definovat, co je to vlastně nejistota. Podívejme se na dvě různé náhodné veličiny X a Y. Nechť

$$P(X = 0) = p, \quad P(X = 1) = 1 - p,$$

a

$$P(Y = 100) = p, \quad P(Y = 200) = 1 - p,$$

přičmž $0 < p < 1$.

Motivace

Zřejmě by nám definice nejistoty měla zajistit, že X a Y jsou stejně nejisté. Tedy nejistota X a tedy i Y by měla být funkcí *pouze* pravděpodobnosti p. Tato vlastnost nejistoty musí být rozšiřitelná i na náhodné proměnné, které nabývají více než dvou hodnot. Tedy:

> Nejistota náhodné proměnné Z, která nabývá hodnoty a_i s pravděpodobností p_i, $(1 \leq i \leq n)$, je funkcí *pouze* pravděpodobností $p_1, \ldots, p_n$.

Obsah

Definice nejistoty

Proto značíme takovouto funkci jako $H(p_1, \ldots, p_n)$, přičemž předpokládáme splnění následujících přirozených podmínek:

(A1) $H(p_1, \ldots, p_n)$ je maximální, když $p_1 = p_2 = \ldots = p_n = 1/n$.

(A2) Pro každou permutaci π na $\{1, \ldots, n\}$ platí

$$H(p_1, \ldots, p_n) = H(p_{\pi(1)}, \ldots, p_{\pi(n)}).$$

Tedy H je symetrická funkce svých argumentů, tj. její výsledek nezávisí na pořadí.

(A3) $H(p_1, \ldots, p_n) \geq 0$ a rovnost nastává právě tehdy, když $p_i = 1$ pro nějaké i.
Nejistota má tedy vždy nezápornou hodnotu a je nulová právě tehdy, když je jakákoliv náhoda vyloučena.

Definice nejistoty

(A4)

$$H(p_1, \ldots, p_n, 0) = H(p_1, \ldots, p_n).$$

Nejistota vrhu šestibokou ideální kostkou je tatáž jako nejistota vrhu sedmibokou kostkou, u které je nemožné, aby padla 7, ale ostatní případy jsou si rovnocenné.

(A5)

$$H\left(\frac{1}{n}, \ldots, \frac{1}{n}\right) \leq H\left(\frac{1}{n+1}, \ldots, \frac{1}{n+1}\right).$$

Výsledek běhu mezi dvěma závodníky je méně nejistý než výsledek běhu mezi více závodníky.

(A6) $H(p_1, \ldots, p_n)$ je spojitá funkce svých parametrů. Malé změny na vstupu dají malé změny na výstupu.

Definice nejistoty

(A7) Jsou-li $m, n \in \mathbf{N}$, pak

$$H\left(\frac{1}{m \cdot n}, \ldots, \frac{1}{m \cdot n}\right) = H\left(\frac{1}{m}, \ldots, \frac{1}{m}\right) + H\left(\frac{1}{n}, \ldots, \frac{1}{n}\right).$$

Tato podmínka říká, že nejistota vrhu $m \cdot n-$stranné kostky je obsažena ve vrhu m-stranné kostky následovaná vrhem n-stranné kostky, a je rovna součtu individuálních nejistot.

(A8) Nechť $p = p_1 + \cdots + p_m$ a $q = q_1 + \cdots + q_n$, p_i, q_j jsou nezáporné. Jsou-li p a q kladná čísla, $p + q = 1$, pak platí

$$H(p_1, ..., p_m, q_1, \ldots, q_n) = H(p, q) + p \cdot H(p_1/p, \ldots, p_m/p) + q \cdot H(q_1/q, \ldots, q_n/q).$$

Představme si, že máme $n + m$ uchazečů na místo v konkurzu - z toho je m mužů a n žen, s pravděpodobnostmi p_i, q_j vítězství v konkurzu. Pak nejistota výsledku konkurzu je nejistota, že vyhraje muž nebo žena plus vážený součet, nejistot výhry mezi muži a ženami.

Obsah

Charakterizace nejistoty

Věta 1.1

Buď $H(p_1, \ldots, p_n)$ funkce definovaná pro každé přirozené číslo n a pro všechny hodnoty $p_1, \ldots, p_n$ tak, že $p_i \geq 0$ a

$$\sum_{i=1}^{n} p_i = 1.$$

Pokud H splňuje axiómy (A1)-(A8), pak platí

$$H(p_1, p_2, \ldots, p_n) = -\lambda \sum_{k} p_k \cdot \log p_k, \tag{1.1}$$

kde λ je libovolná kladná konstanta a sumuje se přes všechna k taková,že $p_k > 0$.

Důkaz Věty 1.1

MA Nechť H splňuje axiómy (A1)-(A8). Definujme

(1) $g(n) = H(1/n, \ldots, 1/n)$ pro $n \in \mathbf{N}$. Z (A7) pak

$$g(n^k) = g(n) + g(n^{k-1})$$

Důkaz Věty 1.1

MA Nechť H splňuje axiómy (A1)-(A8). Definujme

(1) $g(n) = H(1/n, \ldots, 1/n)$ pro $n \in \mathbf{N}$. Z (A7) pak

$$g(n^k) = g(n) + g(n^{k-1})$$

a tedy

(2) $g(n^k) = k \cdot g(n)$.
Buď dále $r, s \in \mathbf{N} - \{1\}, n \in \mathbf{N}$ a $m = m(n, r, s) \in \mathbf{N}$ tak, že

Důkaz Věty 1.1

MA Nechť H splňuje axiómy (A1)-(A8). Definujme

(1) $g(n) = H(1/n, \ldots, 1/n)$ pro $n \in \mathbf{N}$. Z (A7) pak

$$g(n^k) = g(n) + g(n^{k-1})$$

a tedy

(2) $g(n^k) = k \cdot g(n)$.
Buď dále $r, s \in \mathbf{N} - \{1\}, n \in \mathbf{N}$ a $m = m(n, r, s) \in \mathbf{N}$ tak, že

(3) $r^m \leq s^n \leq r^{m+1}$;

Důkaz Věty 1.1

MA Nechť H splňuje axiómy (A1)-(A8). Definujme

(1) $g(n) = H(1/n, \ldots, 1/n)$ pro $n \in \mathbf{N}$. Z (A7) pak

$$g(n^k) = g(n) + g(n^{k-1})$$

a tedy

(2) $g(n^k) = k \cdot g(n)$.
Buď dále $r, s \in \mathbf{N} - \{1\}, n \in \mathbf{N}$ a $m = m(n, r, s) \in \mathbf{N}$ tak, že

(3) $r^m \leq s^n \leq r^{m+1}$; pak dle (2) a monotonie g (dle (A5)) máme

$$g(r^m) \leq g(s^n) \leq g(r^{m+1}),$$

tedy

$$m \cdot g(r) \leq n \cdot g(s) \leq (m+1) \cdot g(r).$$

Důkaz Věty 1.1

Z (3) pak máme

$$m \cdot \ln(r) \leq n \cdot \ln(s) \leq (m+1) \cdot \ln(r)$$

a tedy

$$\left| \frac{g(s)}{g(r)} - \frac{\ln(s)}{\ln(r)} \right| \leq \frac{1}{n}.$$

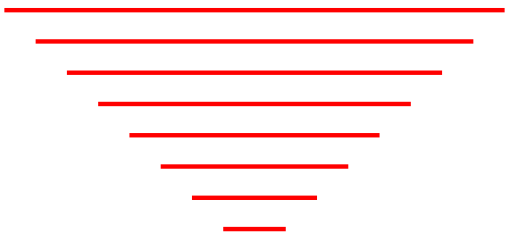
Důkaz Věty 1.1

Z (3) pak máme

$$m \cdot \ln(r) \leq n \cdot \ln(s) \leq (m+1) \cdot \ln(r)$$

a tedy

$$\left| \frac{g(s)}{g(r)} - \frac{\ln(s)}{\ln(r)} \right| \leq \frac{1}{n}.$$



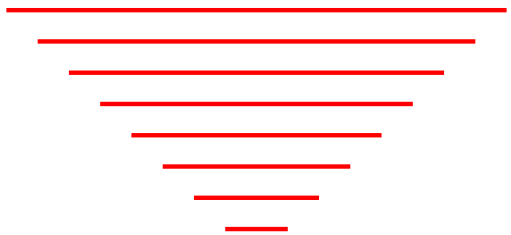
Důkaz Věty 1.1

Z (3) pak máme

$$m \cdot \ln(r) \leq n \cdot \ln(s) \leq (m+1) \cdot \ln(r)$$

a tedy

$$\left| \frac{g(s)}{g(r)} - \frac{\ln(s)}{\ln(r)} \right| \leq \frac{1}{n}.$$



Protože n bylo libovolné přirozené číslo, je

(4)

$$\frac{g(s)}{\ln(s)} = \frac{g(r)}{\ln(r)} = \lambda,$$

kde λ je nějaká (kladná) konstanta.

Důkaz Věty 1.1

Tedy

(5)

$$g(s) = \lambda \cdot \ln(s), \text{ tj. } H(\frac{1}{s}, \ldots, \frac{1}{s}) = -\lambda \cdot \ln(\frac{1}{s}).$$

Důkaz Věty 1.1

Tedy

(5)

$$g(s) = \lambda \cdot \ln(s), \text{ tj. } H(\frac{1}{s}, \ldots, \frac{1}{s}) = -\lambda \cdot \ln(\frac{1}{s}).$$

Buď $0 < p < 1$ racionální, $p = t/n$, $t, n \in \mathbf{N}$.

Důkaz Věty 1.1

Tedy

(5)

$$g(s) = \lambda \cdot \ln(s), \text{ tj. } H(\frac{1}{s}, \ldots, \frac{1}{s}) = -\lambda \cdot \ln(\frac{1}{s}).$$

Buď $0 < p < 1$ racionální, $p = t/n$, $t, n \in \mathbf{N}$.
Položme $q = (n - t)/n$. Z (A8) pak

$$g(n) = H(\frac{1}{n}, \ldots, \frac{1}{n}) = H(\frac{t}{n}, \frac{n-t}{n}) + \frac{t}{n} \cdot g(t) + \frac{n-t}{n} \cdot g(n-t).$$

Důkaz Věty 1.1

Tedy

(5)

$$g(s) = \lambda \cdot \ln(s), \text{ tj. } H(\frac{1}{s}, \ldots, \frac{1}{s}) = -\lambda \cdot \ln(\frac{1}{s}).$$

Buď $0 < p < 1$ racionální, $p = t/n$, $t, n \in \mathbf{N}$.
Položme $q = (n - t)/n$. Z (A8) pak

$$g(n) = H(\frac{1}{n}, \ldots, \frac{1}{n}) = H(\frac{t}{n}, \frac{n-t}{n}) + \frac{t}{n} \cdot g(t) + \frac{n-t}{n} \cdot g(n-t).$$

Z (5) pak jednoduchou úpravou

$$H(\frac{t}{n}, \frac{n-t}{n}) = -\lambda \cdot (\frac{t}{n}) \cdot \ln \frac{t}{n} - \lambda \cdot (\frac{n-t}{n}) \cdot \ln \frac{n-t}{n}.$$

Důkaz Věty 1.1

Zejména pak

(6)
$$H(p, 1-p) = -\lambda \cdot (p \cdot \ln p + (1-p) \cdot \ln(1-p)),$$

a to pro každé racionální číslo p mezi 0 a 1. Ze spojitosti H platí (6) pro všechna $0 < p < 1$.

Důkaz Věty 1.1

Zejména pak

(6)
$$H(p, 1-p) = -\lambda \cdot (p \cdot \ln p + (1-p) \cdot \ln(1-p)),$$

a to pro každé racionální číslo p mezi 0 a 1. Ze spojitosti H platí (6) pro všechna $0 < p < 1$.

Dokažme, že pro každé $N \in \mathbf{N}$ platí

(7)
$$H(p_1, \ldots, p_N) = -\lambda \cdot \sum_{i=1}^{N} p_i \cdot \ln p_i,$$

přičemž $p_i > 0$ a $p_1 + \cdots + p_N = 1$, a to indukcí podle N.

Důkaz Věty 1.1

Z (6) víme, že (7) platí pro $N = 2$ a triviálně platí pro $N = 1$ z (A3).

Důkaz Věty 1.1

Z (6) víme, že (7) platí pro $N = 2$ a triviálně platí pro $N = 1$ z (A3). Předpokládejme, že (7) platí pro N a uvažujme $H(p_1, \ldots, p_{N+1})$. Položme $p = p_1 + \cdots + p_N, q = p_{N+1}$. Případ $p = 0$ je jasný, předpokládejme tedy, že $p > 0$ a použijme (A8).

Důkaz Věty 1.1

Z (6) víme, že (7) platí pro $N = 2$ a triviálně platí pro $N = 1$ z (A3). Předpokládejme, že (7) platí pro N a uvažujme $H(p_1, \ldots, p_{N+1})$. Položme $p = p_1 + \cdots + p_N, q = p_{N+1}$. Případ $p = 0$ je jasný, předpokládejme tedy, že $p > 0$ a použijme (A8). Máme pak

$$\begin{aligned} H(p_1, \ldots, p_{N+1}) &= H(p, q) + p \cdot H\left(\frac{p_1}{p}, \ldots, \frac{p_N}{p}\right) + \overbrace{q \cdot H(1)}^{=0} \\ &= -\lambda \cdot p \cdot \ln p - \lambda \cdot q \cdot \ln q + p \cdot (-\lambda) \cdot \sum_{i=1}^{N} \frac{p_i}{p} \ln \frac{p_i}{p}, \end{aligned}$$

z indukčního předpokladu.

Důkaz Věty 1.

Upravíme-li poslední rovnost na tvar

$$H(p_1,\ldots,p_{N+1}) = -\lambda\cdot(p\cdot\ln p + p_{N+1}\cdot\ln p_{N+1} + \sum_{i=1}^{N} p_i\cdot(\ln p_i - \ln p)),$$

a vzpomeneme-li si, že $\sum_{i=1}^{N} p_i = p$, obdržíme hledanou rovnost

$$H(p_1,\ldots,p_{N+1}) = -\lambda\cdot\sum_{i=1}^{N+1} p_i\cdot\ln p_i.$$

Ekvivalentní definice nejistoty

FI Na základě výše uvedené věty pak definujeme

Definice 1

Buď X náhodná proměnná s konečným oborem hodnot s odpovídajícími pravděpodobnostmi $p_1, p_2, \ldots, p_n$. Pak definujeme nejistotu neboli entropii náhodné veličiny X jako

$$H(X) = -\sum_k p_k \cdot \log_2 p_k, \tag{1.2}$$

kde suma se bere pouze přes ta k, pro která je $p_k > 0$.

Ekvivalentní definice nejistoty

Poznámka 1.2

Nadále budeme vždy (bez újmy na obecnosti) předpokládat, že pro všechny členy pravé strany (1.2) *jsou pravděpodobnosti* p_k *nenulové.*

Poznámka 1.3

Podmínky (A1)-(A8) odpovídají axiomům pro entropii navrženým Shannonem.

Poznámka 1.4

Entropii náhodné veličiny X můžeme rovněž interpretovat jakožto střední hodnotu náhodné proměnné $\log_2 \frac{1}{p(X)}$. *Tedy*

$$H(X) = E(\log_2 \frac{1}{p(X)}) = -\sum_k p_k \cdot \log_2 p_k. \qquad (1.3)$$

Ekvivalentní definice nejistoty

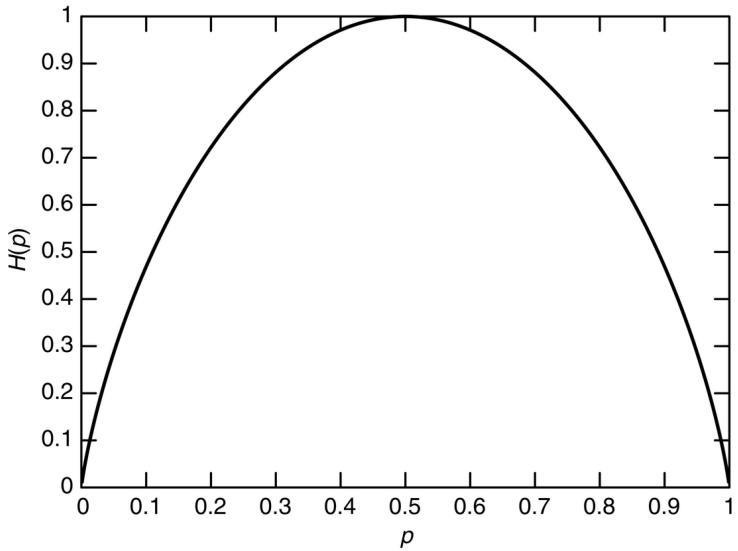
Příklad 1.5

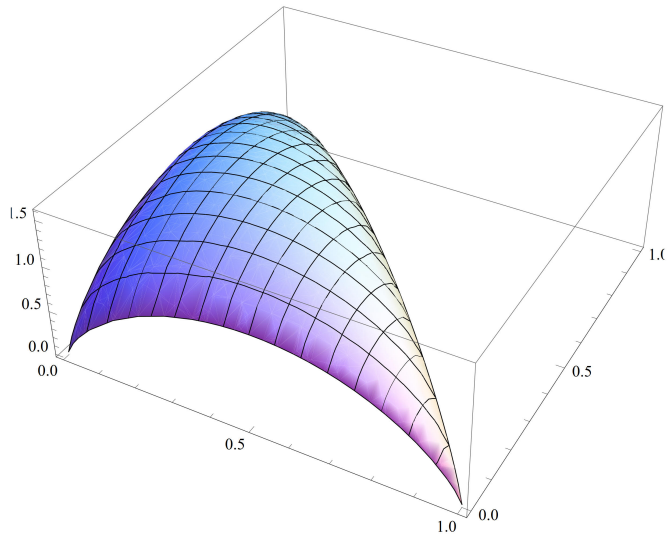
Nechť

$$X = \begin{cases} 1 & \text{s pravděpodobností } p \\ 0 & \text{s pravděpodobností } 1-p. \end{cases} \tag{1.4}$$

Pak

$$H(X) = -p\log_2 p - (1-p)\log_2(1-p). \tag{1.5}$$





Příklady

Cvičení 1.6

1. *Který dostih má větší entropii: ten, ve kterém je sedm žokejů, tři z nich vyhrají s pravděpodobností $\frac{1}{6}$ a čtyři z nich s pravděpodobností $\frac{1}{8}$ nebo dostih, ve kterém je 8 žokejů s dvěma koni s pravděpodobností výhry $\frac{1}{4}$ a šest koní s pravděpodobností výhry $\frac{1}{12}$?*

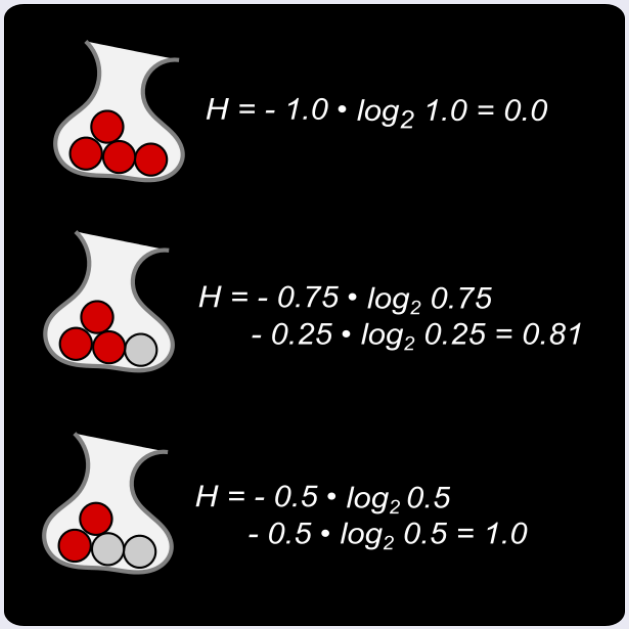
Příklady

Cvičení 1.6

1. *Který dostih má větší entropii: ten, ve kterém je sedm žokejů, tři z nich vyhrají s pravděpodobností $\frac{1}{6}$ a čtyři z nich s pravděpodobností $\frac{1}{8}$ nebo dostih, ve kterém je 8 žokejů s dvěma koni s pravděpodobností výhry $\frac{1}{4}$ a šest koní s pravděpodobností výhry $\frac{1}{12}$?*
2. *Ověřte, že výše definovaná funkce entropie splňuje podmínky (A1)-(A8).*

Příklady

Příklad 1.7



$p_r = 0.5$
$p_g = 0.25$
$p_y = 0.125$
$p_b = 0.125$

$H = 1.75$

Obsah

Entropie náhodného vektoru

FI Řekli jsme, že pro náhodnou proměnnou X s konečným oborem hodnot a s pravděpodobnostmi $p_1, \ldots, p_n$ tak, že

$$\sum p_i = 1 \text{ a } p_i > 0 \ (1 \leq i \leq n),$$

definujeme entropii X jako

$$H(X) = -\sum_{k=1}^{n} p_k \cdot \log_2 p_k.$$

Entropie náhodného vektoru

Analogicky pak pro náhodný vektor $\mathbf{X}$, který nabývá pouze konečně mnoha hodnot $\mathbf{u}_1, \ldots, \mathbf{u}_m$, definujeme jeho *entropii* náhodného vektoru jako

$$H(\mathbf{X}) = -\sum_{k=1}^{m} p(\mathbf{u}_k) \cdot \log_2 p(\mathbf{u}_k). \tag{2.1}$$

Entropie náhodného vektoru

Analogicky pak pro náhodný vektor $\mathbf{X}$, který nabývá pouze konečně mnoha hodnot $\mathbf{u}_1, \ldots, \mathbf{u}_m$, definujeme jeho *entropii* náhodného vektoru jako

$$H(\mathbf{X}) = -\sum_{k=1}^{m} p(\mathbf{u}_k) \cdot \log_2 p(\mathbf{u}_k). \tag{2.1}$$

Je-li například $\mathbf{X}$ 2-dimenzionální náhodný vektor, $\mathbf{X} = (U, V)$ s

$$p_{ij} = P(U = a_i, V = b_j),$$

budeme často psát

$$H(\mathbf{X}) = H(U, V) = -\sum p_{ij} \cdot \log_2 p_{ij}.$$

Entropie náhodného vektoru

Zcela obecně, jsou-li $X_1, \ldots, X_m$ náhodné proměnné tak, že každá z nich nabývá pouze konečně mnoha hodnot, lze pak považovat $\mathbf{X} = (X_1, \ldots, X_m)$ za náhodný vektor a definovat souhrnou entropii $X_1, \ldots, X_m$ jako

$$H(X_1, \ldots, X_m) = H(\mathbf{X}) = -\sum_{(x_1, \ldots, x_m)} p(x_1, \ldots, x_m) \cdot \log_2 p(x_1, \ldots, x_m), \tag{2.2}$$

kde $p(x_1, \ldots, x_m) = P(X_1 = x_1, X_2 = x_2, \ldots, X_m = x_m)$.

Entropie náhodného vektoru

Zcela obecně, jsou-li $X_1, \ldots, X_m$ náhodné proměnné tak, že každá z nich nabývá pouze konečně mnoha hodnot, lze pak považovat $\mathbf{X} = (X_1, \ldots, X_m)$ za náhodný vektor a definovat souhrnou entropii $X_1, \ldots, X_m$ jako

$$H(X_1, \ldots, X_m) = H(\mathbf{X}) = -\sum_{(x_1, \ldots, x_m)} p(x_1, \ldots, x_m) \cdot \log_2 p(x_1, \ldots, x_m), \tag{2.2}$$

kde $p(x_1, \ldots, x_m) = P(X_1 = x_1, X_2 = x_2, \ldots, X_m = x_m)$.

Snadno se ověří, že:

$$H(\mathbf{X}) = 0 \text{ právě tehdy, když } \mathbf{X} \text{ je konstantní.} \tag{2.3}$$

Obsah

Horní hranice entropie

Horní hranice pro H je určena následující větou:

Věta 2.1

Pro každé přirozené číslo n máme

$$H(p_1, \ldots, p_n) \leq \log_2 n,$$

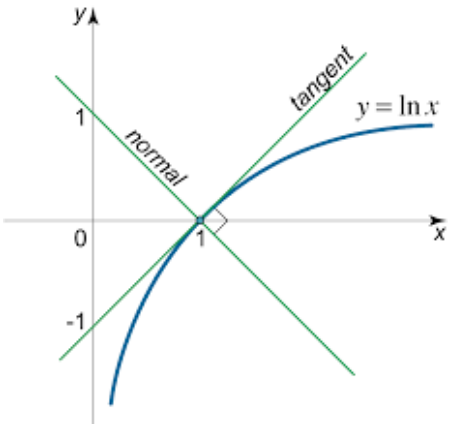
přičemž rovnost nastává právě tehdy, když $p_1 = p_2 = \cdots = p_n = n^{-1}$.

Důkaz Věty 2.1

Zřejmě

$$\log_2 x = \log_2 \mathrm{e} \log_{\mathrm{e}} x.$$

Protože logaritmus je konkávní funkce, tj. leží celá pod tečnou, máme (bereme-li derivaci v bodě 1)



$$\log_{\mathrm{e}} x \leq x - 1,$$

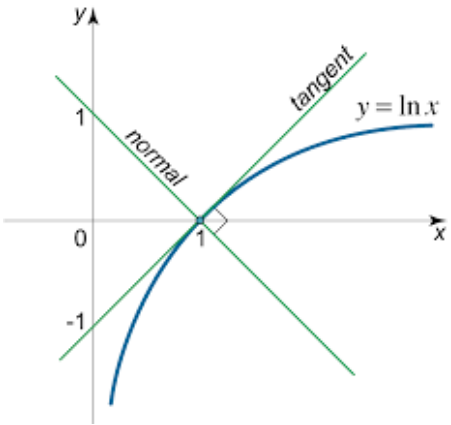
přičemž rovnost nastává právě tehdy, když $x = 1$.

Důkaz Věty 2.1

Zřejmě

$$\log_2 x = \log_2 \mathrm{e} \log_{\mathrm{e}} x.$$

Protože logaritmus je konkávní funkce, tj. leží celá pod tečnou, máme (bereme-li derivaci v bodě 1)



$$\log_{\mathrm{e}} x \leq x - 1,$$

přičemž rovnost nastává právě tehdy, když $x = 1$. Tedy, je-li $(q_1, \ldots, q_n)$ libovolné pravděpodobnostní rozdělení, pak máme

$$\log_{\mathrm{e}}(q_k/p_k) \leq (q_k/p_k) - 1,$$

s rovností právě tehdy, když $q_k = p_k$.

Důkaz Věty 2.1

Tudíž,

$$\sum p_i \cdot \log_e(q_i/p_i) \leq \sum q_k - \sum p_k = 0,$$

a z toho pak

$$\sum p_i \cdot \log_2 q_i \leq \sum p_i \cdot \log_2 p_i.$$

Důkaz Věty 2.1

Tudíž,

$$\sum p_i \cdot \log_e(q_i/p_i) \leq \sum q_k - \sum p_k = 0,$$

a z toho pak

$$\sum p_i \cdot \log_2 q_i \leq \sum p_i \cdot \log_2 p_i.$$

Položíme-li $q_i = 1/n$, obdržíme dosazením

$$H(p_1, \ldots, p_n) = -\sum p_i \cdot \log_2 p_i \leq \log_2 n,$$

což se mělo dokázat. Zbytek tvrzení o rovnosti plyne bezprostředně z důkazu.

Klíčové lemma

Poznamenejme, že jsme dokázali velmi užitečnou nerovnost:

Lemma 2.2

Je-li $\mathbf{p} = (p_i : 1 \leq i \leq n)$ *dané pravděpodobnostní rozdělení, pak minimum funkce*

$$G(q_1, \ldots, q_n) = -\sum p_i \cdot \log_2 q_i$$

přes všechna pravděpodobnostní rozdělení $\mathbf{q} = (q_1, \ldots, q_n)$ *nastává, pokud* $q_k = p_k \quad (1 \leq k \leq n)$.

Někdy též mluvíme o **relativní entropii**

$$D(\mathbf{p}||\mathbf{q}) = \sum p_i \cdot \log_2 \frac{p_i}{q_i}.$$

Tedy $D(\mathbf{p}||\mathbf{q}) \geq 0$ a $D(\mathbf{p}||\mathbf{q}) = 0$, pokud $\mathbf{p} = \mathbf{q}$.

Obsah

Sčítání entropií

Věta 2.3

Jsou-li X a Y náhodné proměnné s konečným oborem hodnot, platí pak

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

přičemž rovnost nastává tehdy a jen tehdy, když X a Y jsou nezávislé.

Důkaz Věty 2.3

Předpokládejme, že

$$r_i = P(X = a_i)\ (1 \leq i \leq m), \quad s_j = P(Y = b_j)\ (1 \leq j \leq n),$$

$$t_{ij} = P(X = a_i, Y = b_j)\ (1 \leq i \leq m, 1 \leq j \leq n).$$

Důkaz Věty 2.3

Předpokládejme, že

$$r_i = P(X = a_i)\ (1 \le i \le m), \quad s_j = P(Y = b_j)\ (1 \le j \le n),$$

$$t_{ij} = P(X = a_i, Y = b_j)\ (1 \le i \le m, 1 \le j \le n).$$

Pak

$$\begin{aligned} H(X) + H(Y) &= -\left(\sum_i r_i \cdot \log_2 r_i + \sum_j s_j \cdot \log_2 s_j \right) \\ &= -\left(\sum_{i,j} t_{ij} \cdot \log_2 r_i + \sum_{j,i} t_{ij} \cdot \log_2 s_j \right), \end{aligned}$$

protože

$$\sum_j t_{ij} = r_i, \sum_i t_{ij} = s_j.$$

Důkaz Věty 2.3

Odtud pak

$$\begin{aligned} H(X) + H(Y) &= -\sum_{i,j} t_{ij} \cdot \log_2(r_i \cdot s_j) \\ &\geq -\sum_{i,j} t_{ij} \cdot \log_2(t_{ij}) = H(X, Y) \end{aligned}$$

dle předchozího lemmatu.

Důkaz Věty 2.3

Odtud pak

$$\begin{aligned} H(X) + H(Y) &= -\sum_{i,j} t_{ij} \cdot \log_2(r_i \cdot s_j) \\ &\geq -\sum_{i,j} t_{ij} \cdot \log_2(t_{ij}) = H(X, Y) \end{aligned}$$

dle předchozího lemmatu. Rovnost nastane právě tehdy, když

$$r_i \cdot s_j = t_{ij}.$$

Ale to je právě podmínka nezávislosti X a Y.

Sčítání entropií

Jednoduchým rozšířením této metody lze dokázat:

$$H(X_1, \ldots, X_n) \leq H(X_1) + \cdots + H(X_n), \tag{2.4}$$

přičemž rovnost nastává právě tehdy, když $X_1, \ldots, X_n$ jsou navzájem nezávislé;

Sčítání entropií

Jednoduchým rozšířením této metody lze dokázat:

$$H(X_1, \ldots, X_n) \leq H(X_1) + \cdots + H(X_n), \tag{2.4}$$

přičemž rovnost nastává právě tehdy, když $X_1, \ldots, X_n$ jsou navzájem nezávislé; případně

$$H(\mathbf{U}, \mathbf{V}) \leq H(\mathbf{U}) + H(\mathbf{V}) \tag{2.5}$$

pro každou dvojici náhodných vektorů $\mathbf{U}, \mathbf{V}$, přičemž rovnost nastává právě tehdy, když $\mathbf{U}$ a $\mathbf{V}$ jsou nezávislé náhodné vektory.

Sčítání entropií

Jednoduchým rozšířením této metody lze dokázat:

$$H(X_1, \ldots, X_n) \leq H(X_1) + \cdots + H(X_n), \tag{2.4}$$

přičemž rovnost nastává právě tehdy, když $X_1, \ldots, X_n$ jsou navzájem nezávislé; případně

$$H(\mathbf{U}, \mathbf{V}) \leq H(\mathbf{U}) + H(\mathbf{V}) \tag{2.5}$$

pro každou dvojici náhodných vektorů $\mathbf{U}, \mathbf{V}$, přičemž rovnost nastává právě tehdy, když $\mathbf{U}$ a $\mathbf{V}$ jsou nezávislé náhodné vektory.
Důkazy (jež lze dokázat přesně stejným způsobem jako větu 2.3 z předchozího lemmatu 2.2) jsou ponechány čtenáři.

Sčítání entropií

Cvičení 2.4

1. *Dvě ideální kostky jsou vrženy; X označuje hodnotu získanou první kostkou, Y hodnotu získanou druhou kostkou. Dokažte, že $H(X, Y) = H(X) + H(Y)$. Dokažte, že je-li $Z = X + Y$, pak*

$$H(Z) < H(X, Y).$$

2. *Dokažte, že pro každou náhodnou proměnnou X,*

$$H(X, X^2) = H(X).$$

3. *Dokažte, že pro každou posloupnost náhodných proměnných $(X_i : 1 \leq i < \infty)$,*

$$H(X_1, \ldots, X_n) \leq H(X_1, \ldots, X_{n+1}).$$

Obsah

Podmíněná entropie

FI Předpokládejme, že X je náhodná proměnná na pravděpodobnostním prostoru Ω a A je událost z Ω. Nabývá-li X konečné množiny hodnot $\{a_i : 1 \leq i \leq m\}$, je přirozené definovat ***podmíněnou entropii*** náhodné proměnné X určenou událostí A jako

$$H(X|A) = -\sum_{k=1}^{m} P(X = a_k|A) \log P(X = a_k|A).$$

Podmíněná entropie

Úplně stejně, je-li Y jiná náhodná proměnná nabývající hodnot b_k $(1 \leq k \leq m)$, definujeme ***podmíněnou entropii*** náhodné proměnné X určenou náhodnou proměnnou Y jako

$$H(X|Y) = \sum_j H(X|Y = b_j)P(Y = b_j).$$

Podmíněná entropie

Úplně stejně, je-li Y jiná náhodná proměnná nabývající hodnot b_k $(1 \leq k \leq m)$, definujeme ***podmíněnou entropii*** náhodné proměnné X určenou náhodnou proměnnou Y jako

$$H(X|Y) = \sum_j H(X|Y = b_j)P(Y = b_j).$$

Považujeme $H(X|Y)$ za entropii náhodné proměnné X určenou jistou hodnotou Y zprůměrovanou přes všechny hodnoty, jichž může Y nabývat.
Zcela triviální důsledky definic jsou:

$$H(X|X) = 0, \tag{3.1}$$

$$H(X|Y) = H(X) \text{ jsou-li } X \text{ a } Y \text{ nezávislé.} \tag{3.2}$$

Podmíněná entropie

Příklad 3.1

Buď X náhodná proměnná získaná vrháním ideální kostky. Buď dále Y jiná náhodná proměnná určená tímtéž experimentem, přičemž Y se rovná 1, je-li vržená hodnota lichá a 0 v ostatních případech. Protože kostka je ideální, platí

$$H(X) = \log 6, H(Y) = \log 2 \text{ a } H(X|Y) = \log 3.$$

Podmíněná entropie

Příklad 3.1

Buď X náhodná proměnná získaná vrháním ideální kostky. Buď dále Y jiná náhodná proměnná určená tímtéž experimentem, přičemž Y se rovná 1, je-li vržená hodnota lichá a 0 v ostatních případech. Protože kostka je ideální, platí

$$H(X) = \log 6, H(Y) = \log 2 \text{ a } H(X|Y) = \log 3.$$

Jsou-li $\mathbf{U}$ a $\mathbf{V}$ náhodné vektory, přirozeně rozšíříme definici podmíněné entropie následovně

$$H(\mathbf{U}|\mathbf{V}) = \sum_j H(\mathbf{U}|\mathbf{V} = \mathbf{v}_j) P(\mathbf{V} = \mathbf{v}_j), \tag{3.3}$$

přičemž se sčítá, jako obvykle, přes (konečný) obor hodnot $\mathbf{v}_j$ tak, že odpovídající pravděpodobnost je kladná.

Obsah

Vlastnosti podmíněné entropie

Jako první příklad, jakým způsobem entropie $H(\mathbf{U}|\mathbf{V})$ měří nejistotu o $\mathbf{U}$ obsaženou ve $\mathbf{V}$, dokážeme:

$H(\mathbf{U}|\mathbf{V}) = 0$ právě tehdy, když $\mathbf{U} = g(\mathbf{V})$ pro nějakou funkci g. (3.4)

Vlastnosti podmíněné entropie

Jako první příklad, jakým způsobem entropie $H(\mathbf{U}|\mathbf{V})$ měří nejistotu o $\mathbf{U}$ obsaženou ve $\mathbf{V}$, dokážeme:

$H(\mathbf{U}|\mathbf{V}) = 0$ právě tehdy, když $\mathbf{U} = g(\mathbf{V})$ pro nějakou funkci g. (3.4)

Důkaz. Pravá strana z definice podmíněné entropie je součet konečného počtu nezáporných veličin.

Vlastnosti podmíněné entropie

Jako první příklad, jakým způsobem entropie $H(\mathbf{U}|\mathbf{V})$ měří nejistotu o $\mathbf{U}$ obsaženou ve $\mathbf{V}$, dokážeme:

$$H(\mathbf{U}|\mathbf{V}) = 0 \text{ právě tehdy, když } \mathbf{U} = g(\mathbf{V}) \text{ pro nějakou funkci } g. \tag{3.4}$$

Důkaz. Pravá strana z definice podmíněné entropie je součet konečného počtu nezáporných veličin.

Tudíž, aby byla nulová, potřebujeme $H(\mathbf{U}|\mathbf{V} = \mathbf{v}_j) = 0$ pro všechna j.

Vlastnosti podmíněné entropie

Jako první příklad, jakým způsobem entropie $H(\mathbf{U}|\mathbf{V})$ měří nejistotu o $\mathbf{U}$ obsaženou ve $\mathbf{V}$, dokážeme:

$$H(\mathbf{U}|\mathbf{V}) = 0 \text{ právě tehdy, když } \mathbf{U} = g(\mathbf{V}) \text{ pro nějakou funkci } g. \tag{3.4}$$

Důkaz. Pravá strana z definice podmíněné entropie je součet konečného počtu nezáporných veličin.

Tudíž, aby byla nulová, potřebujeme $H(\mathbf{U}|\mathbf{V} = \mathbf{v}_j) = 0$ pro všechna j.

Ale opět každá z těchto nezáporných veličin je nulová právě tehdy, když $\mathbf{U}$ je jednoznačně určená $\mathbf{V}$.

Vlastnosti podmíněné entropie - řetězcové pravidlo

Poněkud více nám dává následující výsledek, který matematicky vyjadřuje ideu, že naše definice podmíněné entropie X při daném Y korektně měří zbývající nejistotu.

Vlastnosti podmíněné entropie - řetězcové pravidlo

Poněkud více nám dává následující výsledek, který matematicky vyjadřuje ideu, že naše definice podmíněné entropie X při daném Y korektně měří zbývající nejistotu.

Věta 3.2

Pro každou dvojici náhodných proměnných X a Y, které nabývají pouze konečné množiny hodnot, platí

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

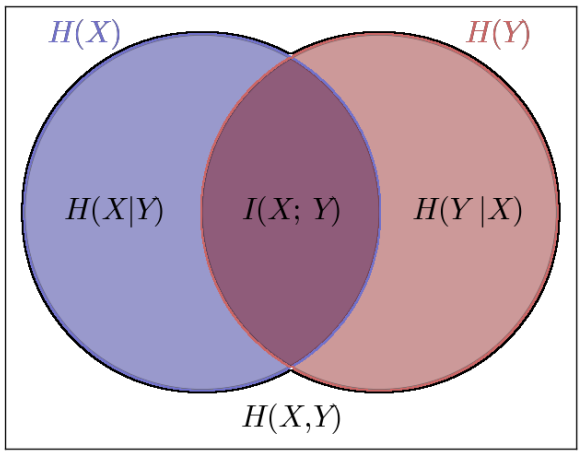
Vlastnosti podmíněné entropie - řetězcové pravidlo

Poněkud více nám dává následující výsledek, který matematicky vyjadřuje ideu, že naše definice podmíněné entropie X při daném Y korektně měří zbývající nejistotu.

Věta 3.2

Pro každou dvojici náhodných proměnných X a Y, které nabývají pouze konečné množiny hodnot, platí

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



Důkaz Věty 3.2

Bez ztráty na obecnosti lze předpokládat, že X a Y nabývají pouze celočíselných hodnot a, kde to bude nutné, že $p_{ij} = P(X = i, Y = j)$.

Důkaz Věty 3.2

Bez ztráty na obecnosti lze předpokládat, že X a Y nabývají pouze celočíselných hodnot a, kde to bude nutné, že $p_{ij} = P(X = i, Y = j)$.

Nyní

$$\begin{aligned} H(X,Y) &= -\sum_i \sum_j P(X=i, Y=j) \log P(X=i, Y=j) \\ &= -\sum_i \sum_j P(X=i, Y=j) \log P(X=i|Y=j) P(Y=j) \\ &= -\sum_i \sum_j p_{ij} \log P(X=i|Y=j) - \sum_i \sum_j p_{ij} \log P(Y=j). \end{aligned}$$

Důkaz Věty 3.2

Tedy

$$
\begin{aligned}
H(X,Y) &= -\sum_i \sum_j p_{ij} \log P(X=i|Y=j) - \sum_i \sum_j p_{ij} \log P(Y=j) \\
&= -\sum_i \sum_j P(X=i|Y=j) P(Y=j) \log P(X=i|Y=j) + H(Y) \\
&= -\sum_j P(Y=j) \sum_i P(X=i|Y=j) \log P(X=i|Y=j) + H(Y) \\
&= \sum_j P(Y=j) H(X|Y=j) + H(Y) \\
&= H(X|Y) + H(Y), \text{ což bylo třeba dokázat.}
\end{aligned}
$$

Vlastnosti podmíněné entropie - řetězcové pravidlo

Věta 3.3

Jsou-li $\mathbf{U}$ *a* $\mathbf{V}$ *náhodné vektory, které nabývají pouze konečné množiny hodnot, platí*

$$H(\mathbf{U}, \mathbf{V}) = H(\mathbf{V}) + H(\mathbf{U}|\mathbf{V}).$$

Vlastnosti podmíněné entropie - řetězcové pravidlo

Věta 3.3

Jsou-li $\mathbf{U}$ *a* $\mathbf{V}$ *náhodné vektory, které nabývají pouze konečné množiny hodnot, platí*

$$H(\mathbf{U}, \mathbf{V}) = H(\mathbf{V}) + H(\mathbf{U}|\mathbf{V}).$$

Důkaz Věty 3.3: Procházíme důkazem předchozí věty, ale namísto X a Y nabývajích pouze celočíselných hodnot i a j máme $\mathbf{U}$ a $\mathbf{V}$ nabývajích hodnot $\mathbf{u}_i$ a $\mathbf{v}_j$, kde $\mathbf{u}_i$ a $\mathbf{v}_j$ jsou zadané vektory.

Vlastnosti podmíněné entropie

Důsledek 3.4

Pro každou dvojici $\mathbf{X}$ *a* $\mathbf{Y}$ *náhodných vektorů je* $H(\mathbf{X}|\mathbf{Y}) \leq H(\mathbf{X})$ *a rovnost nastává právě tehdy, když* $\mathbf{X}$ *a* $\mathbf{Y}$ *jsou nezávislé.*

Vlastnosti podmíněné entropie

Důsledek 3.4

Pro každou dvojici $\mathbf{X}$ *a* $\mathbf{Y}$ *náhodných vektorů je* $H(\mathbf{X}|\mathbf{Y}) \leq H(\mathbf{X})$ *a rovnost nastává právě tehdy, když* $\mathbf{X}$ *a* $\mathbf{Y}$ *jsou nezávislé.*

Důkaz Důsledku 3.4: Z Věty 3.3 máme

$$H(\mathbf{X}|\mathbf{Y}) = H(\mathbf{X}, \mathbf{Y}) - H(\mathbf{Y}).$$

Vlastnosti podmíněné entropie

Důsledek 3.4

Pro každou dvojici $\mathbf{X}$ *a* $\mathbf{Y}$ *náhodných vektorů je* $H(\mathbf{X}|\mathbf{Y}) \leq H(\mathbf{X})$ *a rovnost nastává právě tehdy, když* $\mathbf{X}$ *a* $\mathbf{Y}$ *jsou nezávislé.*

Důkaz Důsledku 3.4: Z Věty 3.3 máme

$$H(\mathbf{X}|\mathbf{Y}) = H(\mathbf{X},\mathbf{Y}) - H(\mathbf{Y}).$$

Ale $H(\mathbf{X}|\mathbf{Y}) + H(\mathbf{Y}) = H(\mathbf{X},\mathbf{Y}) \leq H(\mathbf{X}) + H(\mathbf{Y})$, s rovností právě tehdy, když $\mathbf{X}$ a $\mathbf{Y}$ jsou nezávislé.

Vlastnosti podmíněné entropie

Důsledek 3.4

Pro každou dvojici $\mathbf{X}$ *a* $\mathbf{Y}$ *náhodných vektorů je* $H(\mathbf{X}|\mathbf{Y}) \leq H(\mathbf{X})$ *a rovnost nastává právě tehdy, když* $\mathbf{X}$ *a* $\mathbf{Y}$ *jsou nezávislé.*

Důkaz Důsledku 3.4: Z Věty 3.3 máme

$$H(\mathbf{X}|\mathbf{Y}) = H(\mathbf{X}, \mathbf{Y}) - H(\mathbf{Y}).$$

Ale $H(\mathbf{X}|\mathbf{Y}) + H(\mathbf{Y}) = H(\mathbf{X}, \mathbf{Y}) \leq H(\mathbf{X}) + H(\mathbf{Y})$, s rovností právě tehdy, když $\mathbf{X}$ a $\mathbf{Y}$ jsou nezávislé.

Tedy $H(\mathbf{X}|\mathbf{Y}) \leq H(\mathbf{X})$ a rovnost nastává právě tehdy, když $\mathbf{X}$ a $\mathbf{Y}$ jsou nezávislé.

Vlastnosti podmíněné entropie

Cvičení 3.5

1. *Ukažte, že pro každou náhodnou proměnnou X platí*

$$H(X^2|X) = 0,$$

ale uveďte příklad, že $H(X|X^2)$ není vždy nulová.

Vlastnosti podmíněné entropie

Cvičení 3.5

1. *Ukažte, že pro každou náhodnou proměnnou X platí*

$$H(X^2|X) = 0,$$

ale uveďte příklad, že $H(X|X^2)$ není vždy nulová.

2. *Náhodná proměnná X nabývá celočíselných hodnot $1, \ldots, 2N$ se stejnou pravděpodobností. Náhodná proměnná Y je definovaná $Y = 0$, je-li X sudá, ale $Y = 1$, je-li X lichá. Ukažte, že*

$$H(X|Y) = H(X) - 1,$$

ale že $H(Y|X) = 0$.

Obsah

Informace

FI Zdá se, že R.V.L. Hartley byl v r. 1928 první, kdo se pokusil přiřadit kvantitativní míru k pojmu informace.

Informace

FI Zdá se, že R.V.L. Hartley byl v r. 1928 první, kdo se pokusil přiřadit kvantitativní míru k pojmu informace.

Jeho myšlenky byly pak pomocí matematické statistiky a teorie pravděpodobnosti zevšeobecněny Shannonem, Wienerem a Fischerem koncem 40. let téhož století a daly základ kvantitativní teorii informace, která se zaměřovala na problémy související s množstvím informace, ale ne na problém, co to vlastně informace je.

Informace

FI Zdá se, že R.V.L. Hartley byl v r. 1928 první, kdo se pokusil přiřadit kvantitativní míru k pojmu informace.

Jeho myšlenky byly pak pomocí matematické statistiky a teorie pravděpodobnosti zevšeobecněny Shannonem, Wienerem a Fischerem koncem 40. let téhož století a daly základ kvantitativní teorii informace, která se zaměřovala na problémy související s množstvím informace, ale ne na problém, co to vlastně informace je.

Naznačily však, že je třeba hledat odpověď v obsahových a kvalitativních projevech informace.

Informace

Racionální příčinu tohoto pokusu můžeme částečně popsat následovně.

Informace

Racionální příčinu tohoto pokusu můžeme částečně popsat následovně.

Předpokládejme, že E_1 a E_2 jsou dvě události v pravděpodobnostním prostoru Ω spojené jistým experimentem a předpokládejme, že funkce I je naše míra informace.

Informace

Racionální příčinu tohoto pokusu můžeme částečně popsat následovně.

Předpokládejme, že E_1 a E_2 jsou dvě události v pravděpodobnostním prostoru Ω spojené jistým experimentem a předpokládejme, že funkce I je naše míra informace. Mají-li E_1 a E_2 pravděpodobnosti p_1 a p_2, pak můžeme argumentovat tím, že každá přirozená míra obsahu informace by měla splňovat

$$I(p_1 p_2) = I(p_1) + I(p_2)$$

Informace

Racionální příčinu tohoto pokusu můžeme částečně popsat následovně.

Předpokládejme, že E_1 a E_2 jsou dvě události v pravděpodobnostním prostoru Ω spojené jistým experimentem a předpokládejme, že funkce I je naše míra informace. Mají-li E_1 a E_2 pravděpodobnosti p_1 a p_2, pak můžeme argumentovat tím, že každá přirozená míra obsahu informace by měla splňovat

$$I(p_1 p_2) = I(p_1) + I(p_2)$$

na základě toho, že, pro dvě nezávislé realizace experimentu, informace, pro kterou výsledky těchto experimentů dopadnou jako E_1 následováno E_2, by měla být součtem informací získaných provedením těchto experimentů zvlášť.

Informace

Připustíme-li, že výše uvedená rovnost má jistou platnost, a přejeme-li si mít naši míru nezápornou a spojitou v p, což jsou oba přirozené předpoklady, zbývá nám s malou alternativou *definovat informaci* I události E kladné pravděpodobnosti jako

$$I(E) = -\log_2 P(E),$$

přičemž jsme vybrali 2 jako základ našich logaritmů, abychom zachovali soulad s moderní konvencí.

Informace

Připustíme-li, že výše uvedená rovnost má jistou platnost, a přejeme-li si mít naši míru nezápornou a spojitou v p, což jsou oba přirozené předpoklady, zbývá nám s malou alternativou *definovat informaci* I události E kladné pravděpodobnosti jako

$$I(E) = -\log_2 P(E),$$

přičemž jsme vybrali 2 jako základ našich logaritmů, abychom zachovali soulad s moderní konvencí.

Hartley původně použil logaritmy o základu 10. Někdy též mluvíme o **Hartleyově míře informace**.

Obsah

Informace

Platí totiž následující tvrzení

Věta 4.1

Funkce $I(p)$, definovaná pro všechna $0 < p \leq 1$, splňuje podmínky

- $I(p) \geq 0$, *pro všechna* $0 < p \leq 1$,
- $I(p \cdot q) = I(p) + I(q)$ *pro všechny* $0 < p, q \leq 1$ *takové, že* p *a* q *jsou pravděpodobnosti navzájem nezávislých jevů, a*
- *podmínku spojitosti vzhledem k* p

právě tehdy, když je tvaru

$$I(p) = -\lambda log_2 p,$$

kde λ *je kladná konstanta.*

Důkaz Věty 4.1

Ponecháme za cvičení ukázat, že každá funkce výše uvedeného tvaru splňuje všechny tři podmínky.

Důkaz Věty 4.1

Ponecháme za cvičení ukázat, že každá funkce výše uvedeného tvaru splňuje všechny tři podmínky.

Abychom dokázali obrácené tvrzení, připomeňme, že z vlastnosti $I(p \cdot q) = I(p) + I(q)$ máme $I(p^n) = nI(p)$ pro všechna kladná přirozená čísla n. Speciálně tedy platí

$$I(p^{\frac{1}{n}}) = \frac{1}{n} I(p).$$

Důkaz Věty 4.1

Ponecháme za cvičení ukázat, že každá funkce výše uvedeného tvaru splňuje všechny tři podmínky.

Abychom dokázali obrácené tvrzení, připomeňme, že z vlastnosti $I(p \cdot q) = I(p) + I(q)$ máme $I(p^n) = nI(p)$ pro všechna kladná přirozená čísla n. Speciálně tedy platí

$$I(p^{\frac{1}{n}}) = \frac{1}{n} I(p).$$

Odtud pak máme

$$I(p^{\frac{n}{m}}) = nI(p^{\frac{1}{m}}) = \frac{n}{m} I(p),$$

tj. platí $I(p^q) = qI(p)$ pro všechna kladná racionální čísla q.

Důkaz Věty 4.1

Ze spojitosti umocňování a funkce I máme, že pro všechna kladná reálná čísla r musí platit

$$I(p^r) = rI(p).$$

Důkaz Věty 4.1

Ze spojitosti umocňování a funkce I máme, že pro všechna kladná reálná čísla r musí platit

$$I(p^r) = rI(p).$$

Buď nyní p libovolné, pevně zvolené číslo, $0 < p < 1$. Protože každé číslo q, $0 < q < 1$ lze psát ve tvaru $q = p^{\log_p q}$, máme pak

$$I(q) = I(p^{\log_p q}) = I(p)\log_p q = I(p)\frac{\log_2 q}{\log_2 p} = -\lambda \log_2 q,$$

pro vhodnou kladnou konstantu $\lambda = -\frac{I(p)}{\log_2 p}$.

Důkaz Věty 4.1

Ze spojitosti umocňování a funkce I máme, že pro všechna kladná reálná čísla r musí platit

$$I(p^r) = rI(p).$$

Buď nyní p libovolné, pevně zvolené číslo, $0 < p < 1$. Protože každé číslo q, $0 < q < 1$ lze psát ve tvaru $q = p^{\log_p q}$, máme pak

$$I(q) = I(p^{\log_p q}) = I(p)\log_p q = I(p)\frac{\log_2 q}{\log_2 p} = -\lambda \log_2 q,$$

pro vhodnou kladnou konstantu $\lambda = -\frac{I(p)}{\log_2 p}$.

Ze spojitosti funkce I plyne $I(1) = 0$.

Příklad

Příklad 4.2

Předpokládejme, že máme zdroj, který emituje řetězec binárních číslic 0 *a* 1*, každou se stejnou pravděpodobností a nezávisle pro po sobě jdoucích číslicích. Buď E událost, že prvních n číslic jsou střídavě nuly a jedničky. Pak evidentně*

$$I(E) = -\log_2 \frac{1}{2^n} = n$$

a totéž platí pro každou předepsanou posloupnost číslic délky n.

Příklad

Příklad 4.2

Předpokládejme, že máme zdroj, který emituje řetězec binárních číslic 0 *a* 1*, každou se stejnou pravděpodobností a nezávisle pro po sobě jdoucích číslicích. Buď E událost, že prvních n číslic jsou střídavě nuly a jedničky. Pak evidentně*

$$I(E) = -\log_2 \frac{1}{2^n} = n$$

a totéž platí pro každou předepsanou posloupnost číslic délky n.

Tedy "informačně-teoretická"jednotka informace, totiž *bit*, odpovídá přirozeně využití slova "bit", které znamená binární číslo v současné počítačové terminologii.

Informace pro náhodné vektory

Rozšíříme pojem informace na to, abychom pokryli náhodné proměnné a vektory následovně.

Informace pro náhodné vektory

Rozšíříme pojem informace na to, abychom pokryli náhodné proměnné a vektory následovně.

Předpokládejme, že $\mathbf{X}$ je náhodný vektor, který nabývá hodnoty $\mathbf{x}_1, \ldots, \mathbf{x}_m$, s pravděpodobnostmi $p_1, \ldots, p_m$.

Informace pro náhodné vektory

Rozšíříme pojem informace na to, abychom pokryli náhodné proměnné a vektory následovně.

Předpokládejme, že $\mathbf{X}$ je náhodný vektor, který nabývá hodnoty $\mathbf{x}_1, \ldots, \mathbf{x}_m$, s pravděpodobnostmi $p_1, \ldots, p_m$.

Pak každá z elementárních událostí $\mathbf{X} = \mathbf{x}_k$ $(1 \leq k \leq m)$ obsahuje sdruženou informaci rovnou $-\log_2 p_k$ a poznamenejme, že entropie vektoru $\mathbf{X}$ je určena vztahem

$$H(\mathbf{X}) = -\sum p_k \log_2 p_k = \sum p_k I(\mathbf{X} = \mathbf{x}_k),$$

Informace pro náhodné vektory

Rozšíříme pojem informace na to, abychom pokryli náhodné proměnné a vektory následovně.

Předpokládejme, že $\mathbf{X}$ je náhodný vektor, který nabývá hodnoty $\mathbf{x}_1, \ldots, \mathbf{x}_m$, s pravděpodobnostmi $p_1, \ldots, p_m$.

Pak každá z elementárních událostí $\mathbf{X} = \mathbf{x}_k$ $(1 \leq k \leq m)$ obsahuje sdruženou informaci rovnou $-\log_2 p_k$ a poznamenejme, že entropie vektoru $\mathbf{X}$ je určena vztahem

$$H(\mathbf{X}) = -\sum p_k \log_2 p_k = \sum p_k I(\mathbf{X} = \mathbf{x}_k),$$

tedy $H(\mathbf{X})$ má přirozenou interpretaci jako střední hodnota informace sdružené s elementárními událostmi určenými $\mathbf{X}$.

Obsah

Vzájemná informace pro náhodné vektory

Obecněji, jsou-li $\mathbf{X}$ a $\mathbf{Y}$ dva náhodné vektory, definujeme *vzájemnou informaci o* $\mathbf{X}$ *poskytnutou* $\mathbf{Y}$ jako číslo

$$I(\mathbf{X}|\mathbf{Y}) = H(\mathbf{X}) - H(\mathbf{X}|\mathbf{Y}).$$

Vzájemná informace pro náhodné vektory

Obecněji, jsou-li $\mathbf{X}$ a $\mathbf{Y}$ dva náhodné vektory, definujeme *vzájemnou informaci o* $\mathbf{X}$ *poskytnutou* $\mathbf{Y}$ jako číslo

$$I(\mathbf{X}|\mathbf{Y}) = H(\mathbf{X}) - H(\mathbf{X}|\mathbf{Y}).$$

Jinak řečeno, $I(\mathbf{X}|\mathbf{Y})$ vyjadřuje množství nejistoty o $\mathbf{X}$ odstraněné $\mathbf{Y}$.

Vzájemná informace pro náhodné vektory

Obecněji, jsou-li $\mathbf{X}$ a $\mathbf{Y}$ dva náhodné vektory, definujeme *vzájemnou informaci o* $\mathbf{X}$ *poskytnutou* $\mathbf{Y}$ jako číslo

$$I(\mathbf{X}|\mathbf{Y}) = H(\mathbf{X}) - H(\mathbf{X}|\mathbf{Y}).$$

Jinak řečeno, $I(\mathbf{X}|\mathbf{Y})$ vyjadřuje množství nejistoty o $\mathbf{X}$ odstraněné $\mathbf{Y}$. Zřejmě platí, že

$$I(\mathbf{X}|\mathbf{X}) = H(\mathbf{X}),$$

$I(\mathbf{X}|\mathbf{Y}) = 0$ právě tehdy, když $\mathbf{X}$ a $\mathbf{Y}$ jsou nezávislé.

Vzájemná informace pro náhodné vektory

Obecněji, jsou-li $\mathbf{X}$ a $\mathbf{Y}$ dva náhodné vektory, definujeme *vzájemnou informaci o* $\mathbf{X}$ *poskytnutou* $\mathbf{Y}$ jako číslo

$$I(\mathbf{X}|\mathbf{Y}) = H(\mathbf{X}) - H(\mathbf{X}|\mathbf{Y}).$$

Jinak řečeno, $I(\mathbf{X}|\mathbf{Y})$ vyjadřuje množství nejistoty o $\mathbf{X}$ odstraněné $\mathbf{Y}$. Zřejmě platí, že

$$I(\mathbf{X}|\mathbf{X}) = H(\mathbf{X}),$$

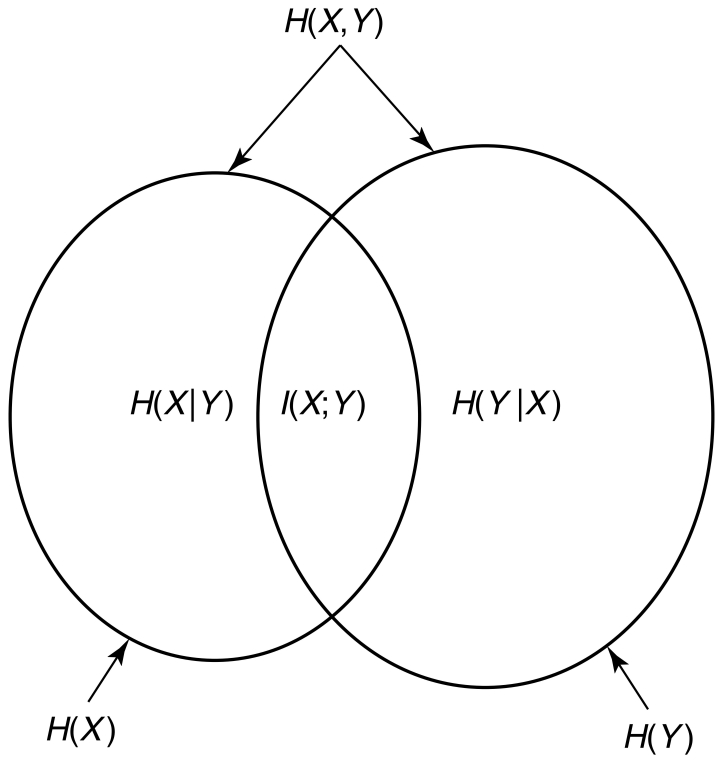
$I(\mathbf{X}|\mathbf{Y}) = 0$ právě tehdy, když $\mathbf{X}$ a $\mathbf{Y}$ jsou nezávislé.

Důkaz. Výsledek plyne bezprostředně z dřívější poznámky, že $H(\mathbf{X}) = H(\mathbf{X}|\mathbf{Y})$ právě tehdy, když $\mathbf{X}$ a $\mathbf{Y}$ jsou nezávislé.

Vzájemná informace pro náhodné vektory

Poněkud udivující symetrie v I je následující výsledek, který zřejmě nemá intuitivní vysvětlení.

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



Vzájemná informace pro náhodné vektory

Cvičení 4.3

1. *Co má větší informační obsah: posloupnost deseti písmen nad abecedou o 26 písmenech nebo posloupnost 26 číslic z množiny* $\{0, 1, \ldots, 9\}$*? [Předpokládejte, že všechny posloupnosti mají stejnou pravděpodobnost.]*
2. *Ideální kostka je vržena. Ukažte, že informace o hodnotě kostky daná znalostí, že se jedná o nesložené číslo, má velikost* $\log_2 \frac{3}{2}$.

Vztah vzájemné informace a relativní entropie

Věta 4.4

Jsou-li $\mathbf{X}$ *a* $\mathbf{Y}$ *dva náhodné vektory s pravděpodobnostními rozděleními* $\mathbf{p} = (p_i : 1 \leq i \leq m)$ *a* $\mathbf{q} = (q_j : 1 \leq j \leq n)$ *a náhodný vektor* $(\mathbf{X}, \mathbf{Y})$ *má pravděpodobnostní rozdělení* $\mathbf{p}_{(\mathbf{X},\mathbf{Y})} = (r_{i,j} : 1 \leq i \leq m, 1 \leq j \leq n)$*, pak*

$$I(\mathbf{X}|\mathbf{Y}) = \sum_{i,j} r_{i,j} \log_2 \frac{r_{i,j}}{p_i \cdot q_j} = D(\mathbf{p}_{(\mathbf{X},\mathbf{Y})} || \mathbf{pq}),$$

kde $\mathbf{pq}$ *je součinové rozdělení marginálů náhodného vektoru* $(\mathbf{X}, \mathbf{Y})$.

Vztah vzájemné informace a relativní entropie

Věta 4.4

Jsou-li $\mathbf{X}$ *a* $\mathbf{Y}$ *dva náhodné vektory s pravděpodobnostními rozděleními* $\mathbf{p} = (p_i : 1 \leq i \leq m)$ *a* $\mathbf{q} = (q_j : 1 \leq j \leq n)$ *a náhodný vektor* $(\mathbf{X}, \mathbf{Y})$ *má pravděpodobnostní rozdělení* $\mathbf{p}_{(\mathbf{X},\mathbf{Y})} = (r_{i,j} : 1 \leq i \leq m, 1 \leq j \leq n)$*, pak*

$$I(\mathbf{X}|\mathbf{Y}) = \sum_{i,j} r_{i,j} \log_2 \frac{r_{i,j}}{p_i \cdot q_j} = D(\mathbf{p}_{(\mathbf{X},\mathbf{Y})} || \mathbf{pq}),$$

kde $\mathbf{pq}$ *je součinové rozdělení marginálů náhodného vektoru* $(\mathbf{X}, \mathbf{Y})$.

$I(\mathbf{X}|\mathbf{Y})$ měří odlišnost sdruženého rozdělení náhodného vektoru $(\mathbf{X}, \mathbf{Y})$ od součinového rozdělení, kterým by se vektor $(\mathbf{X}, \mathbf{Y})$ řídil, pokud by $\mathbf{X}$ a $\mathbf{Y}$ byly nezávislé.

Důkaz Věty 4.4

Uvědomme si nejprve, že

$$\mathbf{p}_{(\mathbf{X},\mathbf{Y})} = \mathbf{p}_{\mathbf{X}|\mathbf{Y}}\mathbf{q},$$

kde $\mathbf{p}_{\mathbf{X}|\mathbf{Y}}$ je pravděpodobnostní rozdělení náhodného vektoru $\mathbf{X}|\mathbf{Y}$.

Důkaz Věty 4.4

Uvědomme si nejprve, že

$$\mathbf{p}_{(\mathbf{X},\mathbf{Y})} = \mathbf{p}_{\mathbf{X}|\mathbf{Y}}\mathbf{q},$$

kde $\mathbf{p}_{\mathbf{X}|\mathbf{Y}}$ je pravděpodobnostní rozdělení náhodného vektoru $\mathbf{X}|\mathbf{Y}$. Platí pak

$$\begin{aligned}
I(\mathbf{X}|\mathbf{Y}) &= H(\mathbf{X}) - H(\mathbf{X}|\mathbf{Y}) \\
&= -\sum_i p_i \log_2 p_i - \sum_j q_j H(\mathbf{X}|\mathbf{Y}=\mathbf{y}_j) \\
&= -\sum_{i,j} \left(r_{i,j} \log_2 p_i - q_j \mathbf{P}(\mathbf{X}=\mathbf{x}_i|\mathbf{Y}=\mathbf{y}_j) \log_2 \mathbf{P}(\mathbf{X}=\mathbf{x}_i|\mathbf{Y}=\mathbf{y}_j) \right) \\
&= \sum_{i,j} \left(r_{i,j} \log_2 \frac{r_{i,j}}{q_j} - r_{i,j} \log_2 p_i \right) \\
&= \sum_{i,j} r_{i,j} \log_2 \frac{r_{i,j}}{p_i \cdot q_j} = D(\mathbf{p}_{(\mathbf{X},\mathbf{Y})} || \mathbf{pq})
\end{aligned}$$

Obsah

Shrnutí

Shrneme-li předchozí odstavce, ukázali jsme, že nejistota a informace jsou tytéž veličiny a odstranění nejistoty je rovno podání informace. Obě veličiny jsou měřitelné matematickým pojmem entropie, který je jednoznačně definován (až na multiplikativní konstantu) veličinou

$$H = -\lambda \sum p_i \cdot \log p_i.$$

Konvence si žádá, aby logaritmy byly brány o základu 2, v kterémžto případě je jednotka entropie *bit*.

Obsah

Problémy k řešení

Problémy 1

1. *Diskžokej má slovník o kapacitě 10 000 slov a pronáší 1000 slov náhodně (opakování je dovoleno). Ukažte, že informační obsah jeho 1000 slov je mnohonásobně menší než obrazovky televizního přijímače o 500 řádcích a 600 sloupcích, přičemž každý pixel nabývá jednu z 16 úrovní jasu.*

Problémy k řešení

Problémy 1

1. *Diskžokej má slovník o kapacitě 10 000 slov a pronáší 1000 slov náhodně (opakování je dovoleno). Ukažte, že informační obsah jeho 1000 slov je mnohonásobně menší než obrazovky televizního přijímače o 500 řádcích a 600 sloupcích, přičemž každý pixel nabývá jednu z 16 úrovní jasu.*
2. *Jsou-li X a Y diskrétní náhodné proměnné, které nabývají pouze konečného počtu hodnot, ukažte, že*

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

Ukažte, že

$$H(g(X, Y)|X) = H(Y|X)$$

neplatí obecně pro $g : \mathbf{R}^2 \to \mathbf{R}$.

Problémy k řešení

Problémy 1

3. *Je-li* $(X_i : 1 \leq i < \infty)$ *posloupnost náhodných proměnných a* Y *je nějaká jiná náhodná proměnná, dokažte, že*

$$H(X_1, \ldots, X_n | Y) \leq H(X_1, \ldots, X_{n+1} | Y)$$

pro každé přirozené číslo n.

4. *Statistický přehled ženatých dvojic ukazuje, že 70% mužů mělo tmavé vlasy, 25% žen bylo blondýnek a že 80% blondýnek si bere tmavovlasé muže. Kolik informace o barvě mužových vlasů je sděleno barvou vlasů jeho ženy?*

5. *Jsou-li* $\mathbf{X}, \mathbf{Y}, \mathbf{Z}$ *náhodné vektory, přičemž každý z nich nabývá pouze konečně mnoha hodnot, dokažte, že*

$$H(\mathbf{Y}|\mathbf{X}) + H(\mathbf{Z}|\mathbf{X}) \geq H(\mathbf{Y}, \mathbf{Z}|\mathbf{X}).$$

Problémy k řešení

Problémy 1

6. *Dokažte, že pro každý náhodný vektor* $\mathbf{Y}$ *a pro každou množinu náhodných proměnných* $X_1, \ldots, X_{n+1}$ *platí*

$$H(\mathbf{Y}|X_1, \ldots, X_n) \geq H(\mathbf{Y}|X_1, \ldots, X_{n+1}).$$

7. *Jsou-li* X *a* Y *dvě náhodné proměnné a* f *a* g *jsou libovolné dvě funkce, dokažte, že*

$$H(f(X), g(Y)) \leq H(X, Y).$$

8. *Náhodná proměnná* X *má binomiální rozdělení s parametry* n *a* p *a platí, pro* $0 \leq k \leq n$,

$$P(X = k) = \binom{n}{k} p^k q^{n-k},$$

kde $0 < p < 1$ *a* $q = 1 - p$.
Dokažte, že $H(X) \leq -n(p\log p + q\log q)$.

Problémy k řešení

Problémy 1

9. *Náhodná veličina X má geometrické rozložení a nabývá celočíselných hodnot $k = 0, 1, 2, \ldots$ s pravděpodobnostmi*

$$p_k = P(X = k) = pq^k,$$

kde $0 < p$ a $p + q = 1$. Ukažte, že rozšíříme-li pojem entropie a definujeme-li

$$H(X) = -\sum_{k=0}^{\infty} p_k \cdot \log p_k,$$

kdykoliv pravá strana konverguje, pak zejména

$$H(X) = -(p\log p + q\log q)/p.$$

Problémy k řešení

Problémy 1

10. *Nazvěme dvě náhodné proměnné X a Y* ekvivalentní, *jestliže $H(X|Y) = 0$ a $H(Y|X) = 0$. Dokažte, že jsou-li X a Y ekvivalentní a Z a Y jsou ekvivalentní, jsou i Z a X jsou ekvivalentní.*
11. *Definujme* vzdálenost *mezi dvěma náhodnými proměnnými X a Y jako*

$$d(X, Y) = H(X|Y) + H(Y|X).$$

Dokažte, že pro všechny tři náhodné proměnné X, Y, Z platí

$$d(X, Y) + d(Y, Z) \geq d(X, Z).$$

Problémy k řešení

Problémy 1

12. *Předpokládejte, že X je náhodná proměnná nabývající hodnot $v_1, \ldots, v_n$. Ukažte, že je-li $E(X) = \mu$ a X je náhodná proměnná s maximální entropií vzhledem k těmto omezením, pak*

$$p_j = P(X = v_j) = Ae^{-\alpha v_j},$$

kde A a α jsou konstanty určené vztahy $E(X) = \mu$ a $\sum p_j = 1$.
Poznámka: hořejší příklad je ilustrací *principu maximální entropie*: jedná se o rozšíření Laplaceova principu nedostatečné příčiny; tento je často užíván ve statistické mechanice, vytváření obrazů a podobně jako je princip pro vybrání a priori distribuce vzhledem k různým omezením. Viz např. Guiasu a Shenitzer (1985).