Entropia pokusu

Stanislav Palúch

Fakulta riadenia a informatiky, Žilinská univerzita

2. marca 2020

Pokus

Ak chceme vedieť odpoveď na otázku „Čo si dostal z algebry", chceme vedieť ktorý jav z množiny javov $\{A,\ B,\ C,\ D,\ E,\ FX\}$ nastal.

Ak chceme vedieť, či vonku mrzne, chceme vedieť, ktorý jav z množiny javov $\{(-\infty, 0),\ \langle 0, \infty)\}$ nastal.

Odpoveď na otázku o odchode vlaku Tatran dá jeden z javov množiny $\{\langle 0, 1),\ \langle 1, 2),\ \dots,\ \langle 1439, 1440)\}$

Definícia

Nech $(\Omega, \mathcal{A}, P)$ je pravdepodobnostný priestor. **Konečný merateľný rozklad istého javu** *je konečná množina javov (t. j. podmnožín Ω) $\{A_1, A_2, \dots, A_n\}$ taká, že $A_i \in \mathcal{A}$ pre $i = 1, 2, \dots, n$, $\bigcup_{i=1}^{n} A_i = \Omega$ a $A_i \cap A_j = \emptyset$ pre $i \neq j$.*
Konečný merateľný rozklad $\mathbf{A} = \{A_1, A_2, \dots, A_n\}$ *istého javu Ω nazývame tiež* **pokusom**.

Pokus

Ak chceme vedieť odpoveď na otázku „Čo si dostal z algebry“, chceme vedieť ktorý jav z množiny javov $\{A,\ B,\ C,\ D,\ E,\ FX\}$ nastal.

Ak chceme vedieť, či vonku mrzne, chceme vedieť, ktorý jav z množiny javov $\{(-\infty, 0),\ \langle 0, \infty)\}$ nastal.

Odpoveď na otázku o odchode vlaku Tatran dá jeden z javov množiny $\{\langle 0, 1),\ \langle 1, 2),\ \dots,\ \langle 1439, 1440)\}$

Definícia

Nech $(\Omega, \mathcal{A}, P)$ je pravdepodobnostný priestor. **Konečný merateľný rozklad istého javu** *je konečná množina javov (t. j. podmnožín Ω) $\{A_1, A_2, \dots, A_n\}$ taká, že $A_i \in \mathcal{A}$ pre $i = 1, 2, \dots, n$, $\bigcup_{i=1}^{n} A_i = \Omega$ a $A_i \cap A_j = \emptyset$ pre $i \neq j$.*
Konečný merateľný rozklad $\mathbf{A} = \{A_1, A_2, \dots, A_n\}$ istého javu Ω nazývame tiež **pokusom**.

Pokus

Ak chceme vedieť odpoveď na otázku „Čo si dostal z algebry“, chceme vedieť ktorý jav z množiny javov $\{A,\ B,\ C,\ D,\ E,\ FX\}$ nastal.

Ak chceme vedieť, či vonku mrzne, chceme vedieť, ktorý jav z množiny javov $\{(-\infty, 0),\ \langle 0, \infty)\}$ nastal.

Odpoveď na otázku o odchode vlaku Tatran dá jeden z javov množiny $\{\langle 0, 1),\ \langle 1, 2),\ \dots,\ \langle 1439, 1440)\}$

Definícia

Nech $(\Omega, \mathcal{A}, P)$ je pravdepodobnostný priestor. **Konečný merateľný rozklad istého javu** *je konečná množina javov (t. j. podmnožín Ω) $\{A_1, A_2, \dots, A_n\}$ taká, že $A_i \in \mathcal{A}$ pre $i = 1, 2, \dots, n$, $\bigcup_{i=1}^{n} A_i = \Omega$ a $A_i \cap A_j = \emptyset$ pre $i \neq j$.*
Konečný merateľný rozklad $\mathbf{A} = \{A_1, A_2, \dots, A_n\}$ istého javu Ω nazývame tiež **pokusom**.

Pokus

Ak chceme vedieť odpoveď na otázku „Čo si dostal z algebry“, chceme vedieť ktorý jav z množiny javov $\{A,\ B,\ C,\ D,\ E,\ FX\}$ nastal.

Ak chceme vedieť, či vonku mrzne, chceme vedieť, ktorý jav z množiny javov $\{(-\infty, 0),\ \langle 0, \infty)\}$ nastal.

Odpoveď na otázku o odchode vlaku Tatran dá jeden z javov množiny $\{\langle 0, 1),\ \langle 1, 2),\ \dots,\ \langle 1439, 1440)\}$

Definícia

Nech $(\Omega, \mathcal{A}, P)$ *je pravdepodobnostný priestor.* **Konečný merateľný rozklad istého javu** *je konečná množina javov (t. j. podmnožín* Ω*)* $\{A_1, A_2, \dots, A_n\}$ *taká, že* $A_i \in \mathcal{A}$ *pre* $i = 1, 2, \dots, n$, $\bigcup_{i=1}^n A_i = \Omega$ *a* $A_i \cap A_j = \emptyset$ *pre* $i \neq j$.
Konečný merateľný rozklad $\mathbf{A} = \{A_1, A_2, \dots, A_n\}$ *istého javu* Ω *nazývame tiež* **pokusom**.

Pokus

V niektorej literatúre sa od množín $\mathbf{A} = \{A_1, A_2, \ldots, A_n\}$ pokusu $\mathbf{A}$ žiadajú oslabené predpoklady,
a to $P\left(\bigcup_{i=1}^{n} A_i\right) = 1$ a $P(A_i \cap A_j) = 0$ pre $i \neq j$.

Ak dostaneme správu, že nastal jav $A_i \in \mathbf{A}$ s pravdepodobnosťou $P(A_i)$, dostaneme s ňou informáciu $-\log_2 P(A_i)$ bitov.

Predstavme si teraz, že máme základnú množinu javov Ω rozdelenú na konečný počet disjunktných javov $A_1, A_2, \ldots, A_n$. Chceme uskutočniť pokus na určenie toho javu A_i, ktorý nastal.

Pred vykonaním pokusu máme neistotu o jeho výsledku. Po uskutočnení pokusu sa výsledok dozvieme a naša neistota zmizne.
Môžeme teda povedať, že veľkosť neistoty pred pokusom sa rovná množstvu informácie, ktorú nám dodá vykonanie pokusu.

Ak majú všetky množiny A_i rovnakú pravdepodobnosť, potom nezávisle na tom, ktorý z javov pokusu $\mathbf{A}$ nastal dostaneme rovnakú informáciu $I(A_i) = \log_2 n$.

Pokus

V niektorej literatúre sa od množín $\mathbf{A} = \{A_1, A_2, \dots, A_n\}$ pokusu $\mathbf{A}$ žiadajú oslabené predpoklady,
a to $P\left(\bigcup_{i=1}^{n} A_i\right) = 1$ a $P(A_i \cap A_j) = 0$ pre $i \neq j$.

Ak dostaneme správu, že nastal jav $A_i \in \mathbf{A}$ s pravdepodobnosťou $P(A_i)$, dostaneme s ňou informáciu $-\log_2 P(A_i)$ bitov.

Predstavme si teraz, že máme základnú množinu javov Ω rozdelenú na konečný počet disjunktných javov $A_1, A_2, \dots, A_n$. Chceme uskutočniť pokus na určenie toho javu A_i, ktorý nastal.

Pred vykonaním pokusu máme neistotu o jeho výsledku. Po uskutočnení pokusu sa výsledok dozvieme a naša neistota zmizne.
Môžeme teda povedať, že veľkosť neistoty pred pokusom sa rovná množstvu informácie, ktorú nám dodá vykonanie pokusu.

Ak majú všetky množiny A_i rovnakú pravdepodobnosť, potom nezávisle na tom, ktorý z javov pokusu $\mathbf{A}$ nastal dostaneme rovnakú informáciu $I(A_i) = \log_2 n$.

Pokus

V niektorých prípadoch môžeme pokus organizovať – môžeme určiť, aké budú jednotlivé množiny rozkladu, čo chceme urobiť tak, aby sme dostali po vykonaní pokusu čo najväčšiu informáciu.
Rozklad množiny Ω na javy, z ktorých každý zodpovedá jednému z výsledkov pokusu, volíme podľa vhodne zvolenej otázky, súboru otázok, možností meracieho postupu a podobne.

Správne zvolený experiment je v mnohých odboroch ľudskej činnosti jedným z rozhodujúcich predpokladov úspechu.

Príklad: Kedy odchádza Tatran zo Žiliny do Bratislavy?

$$\mathbf{P}_1 = \{\langle 0,\ 720), \langle 720,\ 1440)\}$$

$$\mathbf{P}_2 = \{\langle 0,1),\ \langle 1,2),\ \dots,\ \langle 1439,1440)\}$$

Odpoveď na výsledok pokusu $\mathbf{P}_1$ dá 1 bit informácie.
Odpoveď na výsledok pokusu $\mathbf{P}_2$ dá $10.49 = \log_2(1440)$ bitov informácie.

Pokus

Čo však v prípade, keď javy pokusu nemajú rovnakú pravdepodobnosť?

Nech $\Omega = A_1 \cup A_2$, $A_1 \cap A_2 = \emptyset$, $P(A_1) = 0.1$, $P(A_2) = 0.9$.

Ak vyjde A_1, dostaneme informáciu $I(A_1) = -\log_2(0.1) = 3.32$ bitov,

ale ak vyjde A_2, dostaneme informáciu $I(A_2) = -\log_2(0.9) = 0.15$ bitu.

Výsledná informácia teda závisí na výsledku pokusu.

Predstavme si teraz, že pokus vykonáme veľký počet krát – napr. 100 krát.

Približne v desiatich prípadoch dostaneme informáciu 3.32 bitov,

približne v 90 prípadoch dostaneme informáciu 0.15 bitu,

celkovú získanú informáciu možno vyčísliť ako
$10 \times 3.32 + 90 \times 0.15 = 33.2 + 13.5 = 46.7$ bitov.

Priemerná informácia na jeden pokus je $46.7/100 = 0.467$ bitov.

Pokus

Čo však v prípade, keď javy pokusu nemajú rovnakú pravdepodobnosť?

Nech $\Omega = A_1 \cup A_2$, $A_1 \cap A_2 = \emptyset$, $P(A_1) = 0.1$, $P(A_2) = 0.9$.

Ak vyjde A_1, dostaneme informáciu $I(A_1) = -\log_2(0.1) = 3.32$ bitov,

ale ak vyjde A_2, dostaneme informáciu $I(A_2) = -\log_2(0.9) = 0.15$ bitu.

Výsledná informácia teda závisí na výsledku pokusu.
Predstavme si teraz, že pokus vykonáme veľký počet krát – napr. 100 krát.

Približne v desiatich prípadoch dostaneme informáciu 3.32 bitov,

približne v 90 prípadoch dostaneme informáciu 0.15 bitu,

celkovú získanú informáciu možno vyčísliť ako
$10 \times 3.32 + 90 \times 0.15 = 33.2 + 13.5 = 46.7$ bitov.

Priemerná informácia na jeden pokus je $46.7/100 = 0.467$ bitov.

Definícia

Shannonova definícia entropie.
Nech $(\Omega, \mathcal{A}, P)$ *je pravdepodobnostný priestor, na ktorom je daná informácia* $I(A) = -\log_2 P(A)$. *Nech* $\mathbf{P} = \{A_1, A_2, \dots, A_n\}$ *je pokus.* **Entropia** $H(\mathbf{P})$ **pokusu P** *je stredná hodnota diskrétnej náhodnej veličiny* X, *ktorá nadobúda na podmnožine* A_i *hodnotu* $I(A_i)$, *t. j.*

$$H(\mathbf{P}) = \sum_{i=1}^{n} I(A_i)P(A_i) = -\sum_{i=1}^{n} P(A_i).\log_2 P(A_i) \tag{1}$$

Čo sa stane, keď sa v pokuse $\mathbf{P} = \{A_1, A_2, \dots, A_n\}$ vyskytne množina A_i s nulovou pravdepodobnosťou. Potom totiž výraz $-P(a_i).\log_2 P(A_i)$ nie je definovaný.

Pretože $\lim_{x\to 0+} x\log_2(x) = 0$, je prirodzené definovať funkciu $\eta(x)$ nasledovne

$$\eta(x) = \begin{cases} -x.\log_2(x) & \text{ak } x > 0 \\ 0 & \text{ak } x = 0. \end{cases}$$

Potom by Shannonova formula pre entropiu mala byť v tvare

$$H(\mathbf{P}) = \sum_{i=1}^{n} \eta(P(A_i)).$$

Definícia

Shannonova definícia entropie.
Nech $(\Omega, \mathcal{A}, P)$ *je pravdepodobnostný priestor, na ktorom je daná informácia* $I(A) = -\log_2 P(A)$. *Nech* $\mathbf{P} = \{A_1, A_2, \dots, A_n\}$ *je pokus.* **Entropia** $H(\mathbf{P})$ **pokusu P** *je stredná hodnota diskrétnej náhodnej veličiny* X*, ktorá nadobúda na podmnožine* A_i *hodnotu* $I(A_i)$*, t. j.*

$$H(\mathbf{P}) = \sum_{i=1}^{n} I(A_i)P(A_i) = -\sum_{i=1}^{n} P(A_i).\log_2 P(A_i) \qquad (1)$$

Čo sa stane, keď sa v pokuse $\mathbf{P} = \{A_1, A_2, \dots, A_n\}$ vyskytne množina A_i s nulovou pravdepodobnosťou. Potom totiž výraz $-P(a_i).\log_2 P(A_i)$ nie je definovaný.
Pretože $\lim_{x\to 0+} x \log_2(x) = 0$, je prirodzené definovať funkciu $\eta(x)$ nasledovne

$$\eta(x) = \begin{cases} -x.\log_2(x) & \text{ak } x > 0 \\ 0 & \text{ak } x = 0. \end{cases}$$

Potom by Shannonova formula pre entropiu mala byť v tvare

$$H(\mathbf{P}) = \sum_{i=1}^{n} \eta(P(A_i)).$$

Shannonova definícia entropie

Odteraz budeme predpokladať, že výraz $0.\log_2(0)$ je definovaný a že $0.\log_2(0) = 0$.

Entropia pokusu vyjadruje mieru nášho váhania pred jeho vykonaním.

Axiomatická definícia entropie

Majme pokus $\mathbf{P} = \{A_1, A_2, \ldots, A_n\}$,
nech $p_1 = P(A_1)$, $p_2 = P(A_2)$, $\ldots$, $p_n = P(A_n)$.

Predpokladáme, že funkcia H nezávisí od konkrétneho tvaru pravdepodobnostného priestoru $(\Omega, \mathcal{A}, P)$, ale závisí iba od čísel $p_1, p_2, \ldots, p_n$, teda

$$H(\mathbf{P}) = H(p_1, p_2, \ldots, p_n)$$

Funkcia $H(p_1, p_2, \ldots, p_n)$ by mala mať niektoré prirodzené vlastnosti vyplývajúce z jej významu.

Tieto vlastnosti možno formulovať ako axiómy, z ktorých potom možno odvodiť vlastnosti, resp. tvar funkcie H.
Existuje niekoľko sústav axióm, my uvedieme tzv. Fadejevovu sústavu z roku 1956:

Fadejevove axiómy

AF0: Funkcia $y = H(p_1, p_2, \dots, p_n)$ je definovaná pre všetky n a pre všetky $p_1 \geq 0,\ p_2 \geq 0, \dots,\ p_n \geq 0$ také, že $\sum_{i=1}^{n} p_i = 1$, a nadobúda reálne hodnoty.

AF1: $y = H(p, 1-p)$ je spojitá funkcia premennej $p \in \langle 0, 1 \rangle$.

AF2: $y = H(p_1, p_2, \dots, p_n)$ je symetrická funkcia, t. j. pre ľubovoľnú permutáciu π čísel $1, 2, \dots, n$ platí:

$$H(p_{\pi[1]}, p_{\pi[2]}, \dots, p_{\pi[n]}) = H(p_1, p_2, \dots, p_n) \qquad (2)$$

AF3: Ak $p_n = q_1 + q_2 > 0,\ q_1 \geq 0,\ q_2 \geq 0$, potom

$$H(p_1, p_2, \dots, p_{n-1}, q_1, q_2) = $$
$$= H(p_1, p_2, \dots, p_{n-1}, p_n) + p_n . H\left(\frac{q_1}{p_n}, \frac{q_2}{p_n}\right) \qquad (3)$$

Fadejevove axiómy

AF0: Funkcia $y = H(p_1, p_2, \dots, p_n)$ je definovaná pre všetky n a pre všetky $p_1 \geq 0,\ p_2 \geq 0, \dots,\ p_n \geq 0$ také, že $\sum_{i=1}^{n} p_i = 1$, a nadobúda reálne hodnoty.

AF1: $y = H(p, 1-p)$ je spojitá funkcia premennej $p \in \langle 0, 1 \rangle$.

AF2: $y = H(p_1, p_2, \dots, p_n)$ je symetrická funkcia, t. j. pre ľubovoľnú permutáciu π čísel $1, 2, \dots, n$ platí:

$$H(p_{\pi[1]}, p_{\pi[2]}, \dots, p_{\pi[n]}) = H(p_1, p_2, \dots, p_n) \qquad (2)$$

AF3: Ak $p_n = q_1 + q_2 > 0$, $q_1 \geq 0$, $q_2 \geq 0$, potom

$$H(p_1, p_2, \dots, p_{n-1}, q_1, q_2) =$$
$$= H(p_1, p_2, \dots, p_{n-1}, p_n) + p_n . H\left(\frac{q_1}{p_n}, \frac{q_2}{p_n}\right) \qquad (3)$$

Fadejevove axiómy

AF0: Funkcia $y = H(p_1, p_2, \dots, p_n)$ je definovaná pre všetky n a pre všetky $p_1 \geq 0,\ p_2 \geq 0, \dots,\ p_n \geq 0$ také, že $\sum_{i=1}^{n} p_i = 1$, a nadobúda reálne hodnoty.

AF1: $y = H(p, 1-p)$ je spojitá funkcia premennej $p \in \langle 0, 1 \rangle$.

AF2: $y = H(p_1, p_2, \dots, p_n)$ je symetrická funkcia, t. j. pre ľubovoľnú permutáciu π čísel $1, 2, \dots, n$ platí:

$$H(p_{\pi[1]}, p_{\pi[2]}, \dots, p_{\pi[n]}) = H(p_1, p_2, \dots, p_n) \qquad (2)$$

AF3: Ak $p_n = q_1 + q_2 > 0,\ q_1 \geq 0,\ q_2 \geq 0$, potom

$$H(p_1, p_2, \dots, p_{n-1}, q_1, q_2) =$$
$$= H(p_1, p_2, \dots, p_{n-1}, p_n) + p_n . H\left(\frac{q_1}{p_n}, \frac{q_2}{p_n}\right) \qquad (3)$$

Fadejevove axiómy

AF0: Funkcia $y = H(p_1, p_2, \dots, p_n)$ je definovaná pre všetky n a pre všetky $p_1 \geq 0,\ p_2 \geq 0, \dots,\ p_n \geq 0$ také, že $\sum_{i=1}^{n} p_i = 1$, a nadobúda reálne hodnoty.

AF1: $y = H(p, 1-p)$ je spojitá funkcia premennej $p \in \langle 0, 1 \rangle$.

AF2: $y = H(p_1, p_2, \dots, p_n)$ je symetrická funkcia, t. j. pre ľubovoľnú permutáciu π čísel $1, 2, \dots, n$ platí:

$$H(p_{\pi[1]}, p_{\pi[2]}, \dots, p_{\pi[n]}) = H(p_1, p_2, \dots, p_n) \qquad (2)$$

AF3: Ak $p_n = q_1 + q_2 > 0,\ q_1 \geq 0,\ q_2 \geq 0$, potom

$$H(p_1, p_2, \dots, p_{n-1}, q_1, q_2) = \\ = H(p_1, p_2, \dots, p_{n-1}, p_n) + p_n . H\left(\frac{q_1}{p_n}, \frac{q_2}{p_n}\right) \qquad (3)$$

Shannonova axióma

K týmto axiómam pridáme ešte Shannonovu axiómu. Označme

$$F(n) = H\left(\underbrace{\frac{1}{n}, \frac{1}{n}, \dots, \frac{1}{n}}_{n\text{–krát}}\right) \tag{4}$$

Shannonova axióma znie:

AS4: Ak $m < n$, potom $F(m) < F(n)$.

Fadejevove axiómy AF0, až AF3 sú dostatočné na odvodenie tvaru funkcie H a dá sa z nich dokázať i platnosť Shannonovej axiómy AS4.

Veta

Shannonova entropia

$$H(\mathbf{P}) = \sum_{i=1}^{n} I(A_i)P(A_i) = -\sum_{i=1}^{n} P(A_i)\log_2 P(A_i)$$

spĺňa axiómy AF1 až AF3 a Shannonovu axiómu AS4.

Veta

Funkcia $y = H(p_1, p_2, \dots, p_n)$ je spojitá funkcia na množine

$$\mathcal{Q}_n = \left\{(x_1, x_2, \dots, x_n) \mid x_i \geq 0 \text{ pre } i = 1, 2 \dots, n, \ \sum_{i=1}^{n} x_i = 1\right\}.$$

Dôkaz matematickou indukciou podľa m.
Pre $m = 2$ je tvrdenie axiómou AF1.
Nech funkcia $y = H(x_1, x_2, \dots x_m)$ je spojitá na $\mathcal{Q}_m$.
Nech $(p_1, p_2, \dots, p_m, p_{m+1}) \in \mathcal{Q}_{m+1}$.
Predpokladajme, že aspoň jedno z čísel p_m, p_{m+1} je nenulové (inak zmeníme poradie čísel p_i).
Použitím axiómy AF3 máme:

$$H(p_1, p_2, \dots, p_m, p_{m+1}) = H\big(p_1, p_2, \dots, p_{m-1}, (p_m + p_{m+1})\big) + \\ + (p_m + p_{m+1}).H\left(\frac{p_m}{(p_m + p_{m+1})}, \frac{p_{m+1}}{(p_m + p_{m+1})}\right) \quad (5)$$

Spojitosť prvého sčítanca pravej strany (5) vyplýva z indukčného predpokladu, spojitosť druhého sčítanca vyplýva z axiómy AF1. $\square$

Axiomatická definícia entropie

Veta

$$H(1,0) = 0.$$

Dôkaz.

$$H\left(\frac{1}{2}, \underbrace{\frac{1}{2}, 0}\right) = H\left(\frac{1}{2}, \frac{1}{2}\right) + \frac{1}{2}.H(1,0)$$

$$H\left(\frac{1}{2}, \frac{1}{2}, 0\right) = H\left(0, \underbrace{\frac{1}{2}, \frac{1}{2}}\right) = H(0,1) + H\left(\frac{1}{2}, \frac{1}{2}\right) = H\left(\frac{1}{2}, \frac{1}{2}\right) + H(1,0)$$

□

Veta

$$H(p_1, p_2, \dots, p_n, 0) = H(p_1, p_2, \dots, p_n).$$

Dôkaz.
Aspoň jedno z čísel $p_1, p_2, \dots, p_n$ je kladné. Nech je to p_n (inak zmeníme poradie). Potom

$$H(p_1, p_2, \dots, p_n, 0) = H(p_1, p_2, \dots, p_n) + p_n.\underbrace{H(1,0)}_{0} \qquad (6)$$

□

Axiomatická definícia entropie

Veta *Nech $p_n = q_1 + q_2 + \dots + q_m > 0$. Potom*

$$H(p_1, p_2, \dots, p_{n-1}, q_1, q_2, \dots, q_m) =$$

$$= H(p_1, p_2, \dots, p_{n-1}, p_n) + p_n . H\left(\frac{q_1}{p_n}, \frac{q_2}{p_n}, \dots, \frac{q_m}{p_n}\right) \qquad (7)$$

Dôkaz matematickou indukciou podľa m.
Pre $m = 2$ je tvrdenie axiómou AF3.
Nech tvrdenie platí pre nejaké $m \geq 2$.
Položme $p' = q_2 + q_3 + \dots + q_{m+1}$, predpokladajme, že $p' > 0$ (inak zameníme poradie $q_1, q_2, \dots, q_{m+1}$). Podľa indukčného predpokladu

$$H(p_1, p_2, \dots, p_{n-1}, q_1, \underbrace{q_2, \dots, q_{m+1}}_{p' = \sum_{k=2}^{m} q_k}) =$$

$$= H(p_1, p_2, \dots, p_{n-1}, \underbrace{q_1, p'}_{p_n}) + p' . H\left(\frac{q_2}{p'}, \dots, \frac{q_{m+1}}{p'}\right) =$$

$$= H(p_1, p_2, \dots, p_n) + p_n . \left[H\left(\frac{q_1}{p_n}, \frac{p'}{p_n}\right) + \frac{p'}{p_n} H\left(\frac{q_2}{p'}, \dots, \frac{q_{m+1}}{p'}\right)\right]. \qquad (8)$$

Ďalej podľa indukčného predpokladu platí

$$H\left(\frac{q_1}{p_n}, \underbrace{\frac{q_2}{p_n}, \dots, \frac{q_{m+1}}{p_n}}_{\frac{p'}{p_n}}\right) = H\left(\frac{q_1}{p_n}, \frac{p'}{p_n}\right) + \frac{p'}{p_n} H\left(\frac{q_2}{p'}, \dots, \frac{q_{m+1}}{p'}\right). \qquad (9)$$

Vidíme, že pravá strana (9) je totožná s obsahom veľkej hranatej zátvorky na pravej strane vzťahu (8).
Dosadením ľavej strany vzťahu (9) do (8) dostávame (7). □

Veta

Nech pre $i = 1, 2, \dots, n$ máme $p_i = q_{i1} + q_{i2} + \dots + q_{im_i} > 0$. Potom

$$H(\underbrace{q_{11}, q_{12} \dots q_{1m_1}}_{p_1}, \underbrace{q_{21}, q_{22}, \dots, q_{2m_2}}_{p_2}, \dots, \underbrace{q_{n1}, q_{n2}, \dots, q_{nm_n}}_{p_n}) =$$

$$= H(p_1, p_2, \dots, p_n) + \sum_{i=1}^{n} p_i . H\left(\frac{q_{i1}}{p_i}, \frac{q_{i2}}{p_i}, \dots, \frac{q_{im_i}}{p_i}\right) \qquad (10)$$

Dôkaz opakovaným použitím predchádzajúcej vety. □

Veta *Označme $F(n) = H\left(\frac{1}{n}, \frac{1}{n}, \ldots, \frac{1}{n}\right)$. Potom $F(mn) = F(m) + F(n)$.*

Dôkaz.
Použitím (10) máme

$$\begin{aligned}
F(mn) &= H\Bigg(\underbrace{\underbrace{\frac{1}{mn},\ldots,\frac{1}{mn}}_{m\text{-krát}},\ldots\underbrace{\frac{1}{mn},\ldots,\frac{1}{mn}}_{m\text{-krát}}}_{n\text{-krát}}\Bigg) = \\
&= H\left(\frac{1}{n}, \frac{1}{n}, \ldots, \frac{1}{n}\right) + \sum_{i-1}^{n} \frac{1}{n} H\left(\frac{1}{m}, \frac{1}{m}, \ldots, \frac{1}{m}\right) = \\
&= H\left(\frac{1}{n}, \frac{1}{n}, \ldots, \frac{1}{n}\right) + H\left(\frac{1}{m}, \frac{1}{m}, \ldots, \frac{1}{m}\right) = F(n) + F(m)
\end{aligned}$$

□

Axiomatická definícia entropie

Veta

$$F(n^k) = k.F(n)$$

Dôkaz.

$$\begin{aligned} F(n^k) = F(n^{k-1}) + F(n) = F(n^{k-2}) + F(n) + F(n) = \\ = F(n^{k-3}) + F(n) + F(n) + F(n) = \cdots = \\ = \underbrace{F(n) + F(n) + \cdots + F(n)}_{k \text{ krát}} = k.F(n) \end{aligned}$$

$\square$

Dôsledky:

1. $F(1) = F(1^2) = 2.F(1)$, čoho vyplýva, že $F(1) = 0$, a teda $F(1) = c.\log_2(1)$ pre každé c.
2. Pretože podľa axiómy AS4 je funkcia F na množine prirodzených čísel rastúca, je pre každé $m > 1$ $F(m) > F(1) = 0$.

Veta

$$\text{Nech } F(n) = H\left(\frac{1}{n}, \frac{1}{n}, \dots, \frac{1}{n}\right). \text{ Potom } F(n) = c.\log_2(n).$$

Dôkaz.
Vezmime dve prirodzené čísla $m > 1$, $n > 1$ a ľubovoľne veľké prirodzené číslo K. Potom existuje prirodzené číslo k také, že

$$m^k \leq n^K < m^{k+1}. \tag{11}$$

Pretože F je rastúca funkcia, je aj

$$F(m^k) \leq F(n^K) < F(m^{k+1}).$$

Použitím $F(n^k) = k.F(n)$ dostávame

$$k.F(m) \leq K.F(n) < (k+1).F(m).$$

Z posledného výrazu máme ($F(m) > 0$, takže ním možno deliť bez zmeny nerovností)

$$\frac{k}{K} \leq \frac{F(n)}{F(m)} < \frac{k+1}{K}. \tag{12}$$

Axiomatická definícia entropie

Pretože platí (11) môžeme podobnou úvahou postupne písať

$$\log_2(m^k) \le \log_2(n^K) < \log_2(m^{k+1})$$

$$k.\log_2(m) \le K.\log_2(n) < (k+1).\log_2(m),$$

a teda (spomeňme si, že $m > 1$ a teda $\log_2(m) > 0$)

$$\frac{k}{K} \le \frac{\log_2(n)}{\log_2(m)} < \frac{k+1}{K}. \tag{13}$$

Ak porovnáme výrazy (12) a (13) vidíme, že oba zlomky $\dfrac{F(n)}{F(m)}$, $\dfrac{\log_2(n)}{\log_2(m)}$ ležia v intervale $\left\langle \dfrac{k}{K}, \dfrac{k+1}{K} \right)$ dĺžky $\dfrac{1}{K}$ a teda

$$\left| \frac{F(n)}{F(m)} - \frac{\log_2(n)}{\log_2(m)} \right| < \frac{1}{K}. \tag{14}$$

Axiomatická definícia entropie

Celý postup môžeme zopakovať pre ľubovoľne veľké číslo K, a preto (14) musí platiť pre ľubovoľné K, čo je možné len tak, že

$$\frac{F(n)}{F(m)} = \frac{\log_2(n)}{\log_2(m)},$$

a teda

$$F(n) = F(m).\frac{\log_2(n)}{\log_2(m)} = \left(\frac{F(m)}{\log_2(m)}\right)\log_2(n). \tag{15}$$

Ak v (15) fixujeme m a položíme $c = \dfrac{F(m)}{\log_2(m)}$, dostaneme
$F(n) = c.\log_2(n)$. $\square$

Axiomatická definícia entropie

Veta

Nech $p_1 \geq 0,\ p_2 \geq 0, \dots,\ p_n \geq 0$, $\sum_{i=1}^{n} p_i = 1$. *Potom existuje* $c > 0$ *také, že*

$$H(p_1, p_2, \dots, p_n) = -c. \sum_{i=1}^{n} p_i . \log_2(p_i). \tag{16}$$

Axiomatická definícia entropie

Dôkaz.

Dokážeme najprv (16) pre $p_1, p_2, \dots, p_n$ racionálne – t. j. v tvare zlomkov dvoch celých nezáporných čísel. Nech s je spoločný menovateľ zlomkov $p_1, p_2, \dots, p_n$, nech $p_i = \dfrac{q_i}{s}$ pre $i = 1, 2, \dots, n$. Podľa (10) vety 6 môžeme písať

$$c \log_2(s) = F(s) = H\Bigg(\underbrace{\frac{1}{s}, \dots, \frac{1}{s},}_{q_1\text{-krát}}\ \underbrace{\frac{1}{s}, \dots, \frac{1}{s},}_{q_2\text{-krát}}\ \dots\ \underbrace{\frac{1}{s}, \dots, \frac{1}{s},}_{q_n\text{-krát}}\Bigg) =$$

$$= H(p_1, p_2, \dots, p_n) + \sum_{i=1}^{n} p_i . H\left(\frac{1}{q_i}, \frac{1}{q_i} \dots, \frac{1}{q_i}\right) =$$

$$= H(p_1, p_2, \dots, p_n) + \sum_{i=1}^{n} p_i . F(q_i) =$$

$$= H(p_1, p_2, \dots, p_n) + c . \sum_{i=1}^{n} p_i . \log_2(q_i). \quad (17)$$

Axiomatická definícia entropie

Máme teda:

$$c.\log_2(s) = H(p_1, p_2, \dots, p_n) + c\sum_{i=1}^{n} p_i \log_2(q_i).$$

Preto môžeme písať

$$H(p_1, p_2, \dots, p_n) = c\log_2(s) - c.\sum_{i=1}^{n} p_i \log_2(q_i) =$$

$$= c\log_2(s)\sum_{i=1}^{n} p_i - c\sum_{i=1}^{n} p_i \log_2(q_i) = c\sum_{i=1}^{n} p_i \log_2(s) - c\sum_{i=1}^{n} p_i \log_2(q_i) =$$

$$= -c\sum_{i=1}^{n} p_i[\log_2(q_i) - \log_2(s)] = -c\sum_{i=1}^{n} p_i \log_2\left(\frac{q_i}{s}\right) =$$

$$= -c\sum_{i=1}^{n} p_i \log_2(p_i). \qquad (18)$$

Pretože funkcia H je spojitá a (18) platí pre všetky racionálne $p_1 \geq 0,\ p_2 \geq 0, \dots,\ p_n \geq 0$ také, že $\sum_{i=1}^{n} p_i = 1$, musí (18) platiť aj pre všetky reálne argumenty p_i spĺňajúce tie isté podmienky. $\square$

Axiomatická definícia entropie – stanovenie konštanty c

Aby entropia pokusu s dvoma rovnako pravdepodobnými javmi bola jednotková, musí byť $H(1/2, 1/2) = 1$, z čoho vyplýva

$$1 = H\left(\frac{1}{2}, \frac{1}{2}\right) = -c.\left[\frac{1}{2}.\log_2\left(\frac{1}{2}\right) + \frac{1}{2}.\log_2\left(\frac{1}{2}\right)\right] = -c.\left(-\frac{1}{2} - \frac{1}{2}\right) = c$$

$$\boxed{H(p_1, p_2, \ldots, p_n) = -\sum_{i=1}^{n} p_i.\log_2(p_i)}$$

$$\boxed{I(A) = -\log_2 P(A)}$$

Lemma

$$\ln(1+y) \le y \quad pre \quad y > -1 \tag{19}$$

Dôkaz.
Položme $g(y) = \ln(1+y) - y$ a hľadajme jej extrémy.
Je

$$g'(y) = \frac{1}{1+y} - 1, \quad g''(y) = -\frac{1}{(1+y)^2} \le 0.$$

Rovnica $g'(y) = 0$ má jediné riešenie $y = 0$ a $g''(0) = -1 < 0$.

Funkcia $g(y)$ nadobúda svoje lokálne maximum v bode $y = 0$.
Keďže však bod $y = 0$ je jediný, v ktorom môže nastať extrém, funkcia $g(y)$ nadobúda v bode $y = 0$ aj svoje globálne maximum.

Je preto $g(y) \le 0$, t. j. $\ln(1+y) - y \le 0$ a teda $\ln(1+y) \le y$, pričom rovnosť nastáva práve vtedy, keď $y = 0$.

Ak v (21) píšeme $x - 1$ namiesto y dostaneme vzťah

$$\boxed{\ln(x) \le x - 1 \quad \text{pre} \quad x > 0,} \tag{20}$$

pričom rovnosť nastáva práve vtedy, keď $x = 1$.

Ďalšie vlastnosti entropie

Lemma

$$\ln(1+y) \le y \quad \textit{pre} \quad y > -1 \tag{19}$$

Dôkaz.

Položme $g(y) = \ln(1+y) - y$ a hľadajme jej extrémy.

Je

$$g'(y) = \frac{1}{1+y} - 1, \quad g''(y) = -\frac{1}{(1+y)^2} \le 0.$$

Rovnica $g'(y) = 0$ má jediné riešenie $y = 0$ a $g''(0) = -1 < 0$.

Funkcia $g(y)$ nadobúda svoje lokálne maximum v bode $y = 0$.
Keďže však bod $y = 0$ je jediný, v ktorom môže nastať extrém, funkcia $g(y)$ nadobúda v bode $y = 0$ aj svoje globálne maximum.

Je preto $g(y) \le 0$, t. j. $\ln(1+y) - y \le 0$ a teda $\ln(1+y) \le y$, pričom rovnosť nastáva práve vtedy, keď $y = 0$.

Ak v (21) píšeme $x - 1$ namiesto y dostaneme vzťah

$$\boxed{\ln(x) \le x - 1 \quad \text{pre} \quad x > 0,} \tag{20}$$

pričom rovnosť nastáva práve vtedy, keď $x = 1$.

Lemma

$$\ln(1+y) \le y \quad pre \quad y > -1 \tag{19}$$

Dôkaz.
Položme $g(y) = \ln(1+y) - y$ a hľadajme jej extrémy.
Je

$$g'(y) = \frac{1}{1+y} - 1, \quad g''(y) = -\frac{1}{(1+y)^2} \le 0.$$

Rovnica $g'(y) = 0$ má jediné riešenie $y = 0$ a $g''(0) = -1 < 0$.

Funkcia $g(y)$ nadobúda svoje lokálne maximum v bode $y = 0$.
Keďže však bod $y = 0$ je jediný, v ktorom môže nastať extrém, funkcia $g(y)$ nadobúda v bode $y = 0$ aj svoje globálne maximum.

Je preto $g(y) \le 0$, t. j. $\ln(1+y) - y \le 0$ a teda $\ln(1+y) \le y$, pričom rovnosť nastáva práve vtedy, keď $y = 0$.

Ak v (21) píšeme $x - 1$ namiesto y dostaneme vzťah

$$\boxed{\ln(x) \le x - 1 \quad \text{pre} \quad x > 0,} \tag{20}$$

pričom rovnosť nastáva práve vtedy, keď $x = 1$.

Ďalšie vlastnosti entropie

Lemma

$$\ln(1+y) \leq y \quad \textit{pre} \quad y > -1 \tag{19}$$

Dôkaz.
Položme $g(y) = \ln(1+y) - y$ a hľadajme jej extrémy.
Je

$$g'(y) = \frac{1}{1+y} - 1, \quad g''(y) = -\frac{1}{(1+y)^2} \leq 0.$$

Rovnica $g'(y) = 0$ má jediné riešenie $y = 0$ a $g''(0) = -1 < 0$.

Funkcia $g(y)$ nadobúda svoje lokálne maximum v bode $y = 0$.
Keďže však bod $y = 0$ je jediný, v ktorom môže nastať extrém, funkcia $g(y)$ nadobúda v bode $y = 0$ aj svoje globálne maximum.

Je preto $g(y) \leq 0$, t. j. $\ln(1+y) - y \leq 0$ a teda $\ln(1+y) \leq y$, pričom rovnosť nastáva práve vtedy, keď $y = 0$.

Ak v (21) píšeme $x - 1$ namiesto y dostaneme vzťah

$$\boxed{\ln(x) \leq x - 1 \quad \text{pre} \quad x > 0,} \tag{20}$$

pričom rovnosť nastáva práve vtedy, keď $x = 1$.

Ďalšie vlastnosti entropie

Lemma

$$\ln(1+y) \le y \quad \textit{pre} \quad y > -1 \tag{19}$$

Dôkaz.
Položme $g(y) = \ln(1+y) - y$ a hľadajme jej extrémy.
Je

$$g'(y) = \frac{1}{1+y} - 1, \quad g''(y) = -\frac{1}{(1+y)^2} \le 0.$$

Rovnica $g'(y) = 0$ má jediné riešenie $y = 0$ a $g''(0) = -1 < 0$.

Funkcia $g(y)$ nadobúda svoje lokálne maximum v bode $y = 0$.
Keďže však bod $y = 0$ je jediný, v ktorom môže nastať extrém, funkcia $g(y)$ nadobúda v bode $y = 0$ aj svoje globálne maximum.

Je preto $g(y) \le 0$, t. j. $\ln(1+y) - y \le 0$ a teda $\ln(1+y) \le y$, pričom rovnosť nastáva práve vtedy, keď $y = 0$.

Ak v (21) píšeme $x - 1$ namiesto y dostaneme vzťah

$$\boxed{\ln(x) \le x - 1 \quad \text{pre} \quad x > 0,} \tag{20}$$

pričom rovnosť nastáva práve vtedy, keď $x = 1$.

Ďalšie vlastnosti entropie

Lemma

$$\ln(1+y) \leq y \quad pre \quad y > -1 \tag{19}$$

Dôkaz.
Položme $g(y) = \ln(1+y) - y$ a hľadajme jej extrémy.
Je

$$g'(y) = \frac{1}{1+y} - 1, \quad g''(y) = -\frac{1}{(1+y)^2} \leq 0.$$

Rovnica $g'(y) = 0$ má jediné riešenie $y = 0$ a $g''(0) = -1 < 0$.

Funkcia $g(y)$ nadobúda svoje lokálne maximum v bode $y = 0$.
Keďže však bod $y = 0$ je jediný, v ktorom môže nastať extrém, funkcia $g(y)$ nadobúda v bode $y = 0$ aj svoje globálne maximum.

Je preto $g(y) \leq 0$, t. j. $\ln(1+y) - y \leq 0$ a teda $\ln(1+y) \leq y$, pričom rovnosť nastáva práve vtedy, keď $y = 0$.

Ak v (21) píšeme $x - 1$ namiesto y dostaneme vzťah

$$\boxed{\ln(x) \leq x - 1 \quad \text{pre} \quad x > 0,} \tag{20}$$

pričom rovnosť nastáva práve vtedy, keď $x = 1$.

Ďalšie vlastnosti entropie

Lemma

Nech pre všetky $i = 1, 2, \ldots, n$ platí $p_i > 0$, $q_i > 0$, $\sum_{i=1}^{n} p_i = 1$, $\sum_{i=1}^{n} q_i = 1$. Potom

$$-\sum_{i=1}^{n} p_i \log_2(p_i) \leq -\sum_{i=1}^{n} p_i \log_2(q_i), \tag{21}$$

pričom rovnosť nastáva práve vtedy, keď $p_i = q_i$ pre všetky $i = 1, 2, \ldots, n$.

Dôkaz. Dosaďme teraz do nerovnosti $\ln(x) \leq x - 1$ za $x = \dfrac{q_i}{p_i}$. Postupnými úpravami dostávame

$$\ln\left(\frac{q_i}{p_i}\right) \leq \frac{q_i}{p_i} - 1$$

$$\ln(q_i) - \ln(p_i) \leq \frac{q_i}{p_i} - 1$$

$$p_i \ln(q_i) - p_i \ln(p_i) \leq q_i - p_i$$

$$-p_i \ln(p_i) \leq -p_i \ln(q_i) + q_i - p_i$$

$$-\sum_{i=1}^{n} p_i \ln(p_i) \leq -\sum_{i=1}^{n} p_i \ln(q_i) + \underbrace{\sum_{i=1}^{n} q_i}_{=1} - \underbrace{\sum_{i=1}^{n} p_i}_{=1}$$

Ďalšie vlastnosti entropie

Lemma

Nech pre všetky $i = 1, 2, \ldots, n$ platí $p_i > 0,\ q_i > 0$, $\sum_{i=1}^{n} p_i = 1$, $\sum_{i=1}^{n} q_i = 1$. Potom

$$-\sum_{i=1}^{n} p_i \log_2(p_i) \leq -\sum_{i=1}^{n} p_i \log_2(q_i), \qquad (21)$$

pričom rovnosť nastáva práve vtedy, keď $p_i = q_i$ pre všetky $i = 1, 2, \ldots, n$.

Dôkaz. Dosaďme teraz do nerovnosti $\ln(x) \leq x - 1$ za $x = \dfrac{q_i}{p_i}$. Postupnými úpravami dostávame

$$\ln\left(\frac{q_i}{p_i}\right) \leq \frac{q_i}{p_i} - 1$$

$$\ln(q_i) - \ln(p_i) \leq \frac{q_i}{p_i} - 1$$

$$p_i \ln(q_i) - p_i \ln(p_i) \leq q_i - p_i$$

$$-p_i \ln(p_i) \leq -p_i \ln(q_i) + q_i - p_i$$

$$-\sum_{i=1}^{n} p_i \ln(p_i) \leq -\sum_{i=1}^{n} p_i \ln(q_i) + \underbrace{\sum_{i=1}^{n} q_i}_{=1} - \underbrace{\sum_{i=1}^{n} p_i}_{=1}$$

Ďalšie vlastnosti entropie

Lemma

Nech pre všetky $i = 1, 2, \ldots, n$ *platí* $p_i > 0,\ q_i > 0$, $\sum_{i=1}^n p_i = 1$, $\sum_{i=1}^n q_i = 1$. *Potom*

$$-\sum_{i=1}^{n} p_i \log_2(p_i) \leq -\sum_{i=1}^{n} p_i \log_2(q_i), \tag{21}$$

pričom rovnosť nastáva práve vtedy, keď $p_i = q_i$ *pre všetky* $i = 1, 2, \ldots, n$.

Dôkaz. Dosaďme teraz do nerovnosti $\ln(x) \leq x - 1$ za $x = \dfrac{q_i}{p_i}$. Postupnými úpravami dostávame

$$
\begin{aligned}
\ln\left(\frac{q_i}{p_i}\right) &\leq \frac{q_i}{p_i} - 1 \\
\ln(q_i) - \ln(p_i) &\leq \frac{q_i}{p_i} - 1 \\
p_i \ln(q_i) - p_i \ln(p_i) &\leq q_i - p_i \\
-p_i \ln(p_i) &\leq -p_i \ln(q_i) + q_i - p_i \\
-\sum_{i=1}^{n} p_i \ln(p_i) &\leq -\sum_{i=1}^{n} p_i \ln(q_i) + \underbrace{\sum_{i=1}^{n} q_i}_{=1} - \underbrace{\sum_{i=1}^{n} p_i}_{=1}
\end{aligned}
$$

Ďalšie vlastnosti entropie

Lemma

Nech pre všetky $i = 1, 2, \ldots, n$ *platí* $p_i > 0,\ q_i > 0,\ \sum_{i=1}^n p_i = 1,\ \sum_{i=1}^n q_i = 1.$ *Potom*

$$-\sum_{i=1}^{n} p_i \log_2(p_i) \leq -\sum_{i=1}^{n} p_i \log_2(q_i), \tag{21}$$

pričom rovnosť nastáva práve vtedy, keď $p_i = q_i$ *pre všetky* $i = 1, 2, \ldots, n$.

Dôkaz. Dosaďme teraz do nerovnosti $\ln(x) \leq x - 1$ za $x = \dfrac{q_i}{p_i}$. Postupnými úpravami dostávame

$$\begin{aligned}
\ln\left(\frac{q_i}{p_i}\right) &\leq \frac{q_i}{p_i} - 1 \\
\ln(q_i) - \ln(p_i) &\leq \frac{q_i}{p_i} - 1 \\
p_i \ln(q_i) - p_i \ln(p_i) &\leq q_i - p_i \\
-p_i \ln(p_i) &\leq -p_i \ln(q_i) + q_i - p_i
\end{aligned}$$

$$-\sum_{i=1}^{n} p_i \ln(p_i) \leq -\sum_{i=1}^{n} p_i \ln(q_i) + \underbrace{\sum_{i=1}^{n} q_i}_{=1} - \underbrace{\sum_{i=1}^{n} p_i}_{=1}$$

Ďalšie vlastnosti entropie

Lemma

Nech pre všetky $i = 1, 2, \dots, n$ *platí* $p_i > 0,\ q_i > 0,\ \sum_{i=1}^{n} p_i = 1,\ \sum_{i=1}^{n} q_i = 1$. *Potom*

$$-\sum_{i=1}^{n} p_i \log_2(p_i) \le -\sum_{i=1}^{n} p_i \log_2(q_i), \tag{21}$$

pričom rovnosť nastáva práve vtedy, keď $p_i = q_i$ *pre všetky* $i = 1, 2, \dots, n$.

Dôkaz. Dosaďme teraz do nerovnosti $\ln(x) \le x - 1$ za $x = \dfrac{q_i}{p_i}$. Postupnými úpravami dostávame

$$\begin{aligned}
\ln\left(\frac{q_i}{p_i}\right) &\le \frac{q_i}{p_i} - 1 \\
\ln(q_i) - \ln(p_i) &\le \frac{q_i}{p_i} - 1 \\
p_i \ln(q_i) - p_i \ln(p_i) &\le q_i - p_i \\
-p_i \ln(p_i) &\le -p_i \ln(q_i) + q_i - p_i \\
-\sum_{i=1}^{n} p_i \ln(p_i) &\le -\sum_{i=1}^{n} p_i \ln(q_i) + \underbrace{\sum_{i=1}^{n} q_i}_{=1} - \underbrace{\sum_{i=1}^{n} p_i}_{=1}
\end{aligned}$$

Ďalšie vlastnosti entropie

$$-\sum_{i=1}^{n} p_i \frac{\ln(p_i)}{\ln(2)} \leq -\sum_{i=1}^{n} p_i \frac{\ln(q_i)}{\ln(2)}$$

$$-\sum_{i=1}^{n} p_i \log_2(p_i) \leq -\sum_{i=1}^{n} p_i \log_2(q_i),$$

pričom rovnosť v prvých troch riadkoch nastáva práve vtedy, keď $p_i = q_i$, rovnosť v posledných troch riadkoch nastáva práve vtedy, keď $p_i = q_i$ pre všetky $i = 1, 2, \ldots, n$. $\square$

Ďalšie vlastnosti entropie

$$-\sum_{i=1}^{n} p_i \frac{\ln(p_i)}{\ln(2)} \leq -\sum_{i=1}^{n} p_i \frac{\ln(q_i)}{\ln(2)}$$

$$-\sum_{i=1}^{n} p_i \log_2(p_i) \leq -\sum_{i=1}^{n} p_i \log_2(q_i),$$

pričom rovnosť v prvých troch riadkoch nastáva práve vtedy, keď $p_i = q_i$, rovnosť v posledných troch riadkoch nastáva práve vtedy, keď $p_i = q_i$ pre všetky $i = 1, 2, \ldots, n$. $\square$

Ďalšie vlastnosti entropie

Veta

Pre dané n funkcia

$$H(p_1, p_2, \dots, p_n) = -\sum_{i=1}^{n} p_i \log_2(p_i)$$

nadobúda maximum pre $p_1 = p_2 = \cdots = p_n = 1/n$.

Dôkaz. Vezmime $p_1, p_2, \dots, p_n$ ľubovoľné také, že splňujú predpoklady vety, a položme v (21) $q_1 = q_2 = \cdots = q_n = \dfrac{1}{n}$.

Potom

$$H(p_1, p_2, \dots, p_n) = -\sum_{i=1}^{n} p_i \log_2(p_i) \le -\sum_{i=1}^{n} p_i \log_2\left(\frac{1}{n}\right) =$$

$$= -\log_2\left(\frac{1}{n}\right) \cdot \underbrace{\sum_{i=1}^{n} p_i}_{=1} = -\log_2(\frac{1}{n}) = \log_2 n = H\left(\frac{1}{n}, \frac{1}{n}, \dots, \frac{1}{n}\right)$$

□

Veta

Pre dané n funkcia

$$H(p_1, p_2, \dots, p_n) = -\sum_{i=1}^{n} p_i \log_2(p_i)$$

nadobúda maximum pre $p_1 = p_2 = \cdots = p_n = 1/n$.

Dôkaz. Vezmime $p_1, p_2, \dots, p_n$ ľubovoľné také, že splňujú predpoklady vety, a položme v (21) $q_1 = q_2 = \cdots = q_n = \dfrac{1}{n}$.

Potom

$$H(p_1, p_2, \dots, p_n) = -\sum_{i=1}^{n} p_i \log_2(p_i) \leq -\sum_{i=1}^{n} p_i \log_2\left(\frac{1}{n}\right) =$$

$$= -\log_2\left(\frac{1}{n}\right) \cdot \underbrace{\sum_{i=1}^{n} p_i}_{=1} = -\log_2\left(\frac{1}{n}\right) = \log_2 n = H\left(\frac{1}{n}, \frac{1}{n}, \dots, \frac{1}{n}\right)$$

$\square$

Ďalšie vlastnosti entropie

Veta

Pre dané n funkcia

$$H(p_1, p_2, \dots, p_n) = -\sum_{i=1}^{n} p_i \log_2(p_i)$$

nadobúda maximum pre $p_1 = p_2 = \cdots = p_n = 1/n$.

Dôkaz. Vezmime $p_1, p_2, \dots, p_n$ ľubovoľné také, že splňujú predpoklady vety, a položme v (21) $q_1 = q_2 = \cdots = q_n = \dfrac{1}{n}$.
Potom

$$H(p_1, p_2, \dots, p_n) = -\sum_{i=1}^{n} p_i \log_2(p_i) \le -\sum_{i=1}^{n} p_i \log_2\left(\frac{1}{n}\right) =$$

$$= -\log_2\left(\frac{1}{n}\right) \cdot \underbrace{\sum_{i=1}^{n} p_i}_{=1} = -\log_2(\frac{1}{n}) = \log_2 n = H\left(\frac{1}{n}, \frac{1}{n}, \dots, \frac{1}{n}\right)$$

□

Podmienená entropia

$\mathbf{B} = \{B_1, B_2, \dots, B_m\}$ pokus na pravdepodobnostnom priestore $(\Omega, \mathcal{A}, P)$.

Predpokladajme, že nastal elementárny jav $\omega \in \Omega$.

Chceme vedieť, ktorý z javov B_j nastal, t. j. pre ktoré $j = 1, 2, \dots m$ je $\omega \in B_j$.

Pre nejaké ohraničenia nemôžeme vykonať pokus $\mathbf{B}$ (tým skôr sa nemôžeme dozvedieť, ktorý jav $\omega \in \Omega$ nastal), ale dozvieme sa výsledok pokusu $\mathbf{A} = \{A_1, A_2, \dots, A_n\}$.

Predpokladajme, že jeho výsledkom je jav A_i. Ak už vieme, že nastal jav A_i, javy B_1, B_2, $\dots$, B_m nastanú s pravdepodobnosťami

$$P(B_1|A_i), P(B_2|A_i), \dots, P(B_m|A_i).$$

Neurčitosť pokusu $\mathbf{B}$ sa zmení z hodnoty

$$H(\mathbf{B}) = H(P(B_1), P(B_2), \dots, P(B_m))$$

na hodnotu

$$H\big(P(B_1|A_i), P(B_2|A_i), \dots, P(B_m|A_i)\big),$$

ktorú budeme označovať $H(\mathbf{B}|A_i)$.

Podmienená entropia

$\mathbf{B} = \{B_1, B_2, \dots, B_m\}$ pokus na pravdepodobnostnom priestore $(\Omega, \mathcal{A}, P)$.

Predpokladajme, že nastal elementárny jav $\omega \in \Omega$.

Chceme vedieť, ktorý z javov B_j nastal, t. j. pre ktoré $j = 1, 2, \dots m$ je $\omega \in B_j$.

Pre nejaké ohraničenia nemôžeme vykonať pokus $\mathbf{B}$ (tým skôr sa nemôžeme dozvedieť, ktorý jav $\omega \in \Omega$ nastal), ale dozvieme sa výsledok pokusu $\mathbf{A} = \{A_1, A_2, \dots, A_n\}$.

Predpokladajme, že jeho výsledkom je jav A_i. Ak už vieme, že nastal jav A_i, javy B_1, B_2, $\dots$, B_m nastanú s pravdepodobnosťami

$$P(B_1|A_i), P(B_2|A_i), \dots, P(B_m|A_i).$$

Neurčitosť pokusu $\mathbf{B}$ sa zmení z hodnoty

$$H(\mathbf{B}) = H(P(B_1), P(B_2), \dots, P(B_m))$$

na hodnotu

$$H\big(P(B_1|A_i), P(B_2|A_i), \dots, P(B_m|A_i)\big),$$

ktorú budeme označovať $H(\mathbf{B}|A_i)$.

Podmienená entropia

Definícia

Nech $\mathbf{A} = \{A_1, A_2, \dots, A_n\}$, $\mathbf{B} = \{B_1, B_2, \dots, B_m\}$ *sú dva pokusy.* **Podmienenou entropiu pokusu B za predpokladu, že nastal jav** A_i *(alebo len za podmienky* A_i*) je*

$$H(\mathbf{B}|A_i) = H\big(P(B_1|A_i), P(B_2|A_i), \dots, P(B_m|A_i)\big) = \\ = -\sum_{j=1}^{m} P(B_j|A_i).\log_2(P(B_j|A_i)). \qquad (22)$$

Podmienená entropia – príklad

Hádžeme hracou kockou. Označme $\mathbf{B} = \{B_1, B_2, \dots, B_6\}$ pokus, v ktorom jav B_i znamená „padlo i bodov“ pre $i = 1, 2, \dots 6$.
Všetky javy majú rovnakú pravdepodobnosť $P(B_i) = 1/6$.

$$H(\mathbf{B}) = H(1/6, 1/6, \dots, 1/6) = \log_2(6) = 2.585 \text{ bitu}.$$

Predpokladajme, že sa po uskutočnení pokusu dozvieme, že padlo nepárne číslo.
Označme $A_1 = B_1 \cup B_3 \cup B_5$, $A_2 = B_2 \cup B_4 \cup B_6$.
Jav A_1 znamená „padlo nepárne číslo“, jav A_2 znamená „padlo párne číslo“.
Správa $\mathcal{S} =$„padlo nepárne číslo“ t.j. „nastal jav A_1“
nesie so sebou $-\log_2(P(A_1)) = -\log_2(1/2) = 1$ bit informácie.

Po správe $\mathcal{S}$ sa naša neurčitosť o výsledku pokusu zmení z $H(\mathbf{B})$ na

$$H(\mathbf{B}|A_1) = H\big(P(B_1|A_1), P(B_2|A_1), P(B_3|A_1), P(B_4|A_1), P(B_5|A_1), P(B_6|A_1)\big) = H(1/3, 0, 1/3, 0, 1/3, 0) = H(1/3, 1/3, 1/3) = \log_2(3) = 1.585 \text{ bitu}.$$

Správa $\mathcal{S}$ nesúca 1 bit informácie znžila našu neurčitosť o výsledku pokusu z $H(\mathbf{B}) = 2.585$ na $H(\mathbf{B}|A_1) = 1.585$ – práve o množstvo informácie, ktoré so sebou niesla. POZOR! Toto nie je všeobecne platná skutočnosť.

Podmienená entropia – príklad

Hádžeme hracou kockou. Označme $\mathbf{B} = \{B_1, B_2, \dots, B_6\}$ pokus, v ktorom jav B_i znamená „padlo i bodov" pre $i = 1, 2, \dots 6$.
Všetky javy majú rovnakú pravdepodobnosť $P(B_i) = 1/6$.

$$H(\mathbf{B}) = H(1/6, 1/6, \dots, 1/6) = \log_2(6) = 2.585 \text{ bitu.}$$

Predpokladajme, že sa po uskutočnení pokusu dozvieme, že padlo nepárne číslo.
Označme $A_1 = B_1 \cup B_3 \cup B_5$, $A_2 = B_2 \cup B_4 \cup B_6$.
Jav A_1 znamená „padlo nepárne číslo", jav A_2 znamená „padlo párne číslo".
Správa $\mathcal{S} =$„padlo nepárne číslo" t.j. „nastal jav A_1"
nesie so sebou $-\log_2(P(A_1)) = -\log_2(1/2) = 1$ bit informácie.

Po správe $\mathcal{S}$ sa naša neurčitosť o výsledku pokusu zmení z $H(\mathbf{B})$ na

$$H(\mathbf{B}|A_1) = H\big(P(B_1|A_1), P(B_2|A_1), P(B_3|A_1), P(B_4|A_1), P(B_5|A_1), P(B_6|A_1)\big) = H(1/3, 0, 1/3, 0, 1/3, 0) = H(1/3, 1/3, 1/3) = \log_2(3) = 1.585 \text{ bitu.}$$

Správa $\mathcal{S}$ nesúca 1 bit informácie znžila našu neurčitosť o výsledku pokusu z $H(\mathbf{B}) = 2.585$ na $H(\mathbf{B}|A_1) = 1.585$ – práve o množstvo informácie, ktoré so sebou niesla. POZOR! Toto nie je všeobecne platná skutočnosť.

Podmienená entropia

Michail Schumacher bol fenomenálny pilot formuly 1, ktorý získal v rokoch 1994, 1995 a 2000–2004 sedem titulov majstra sveta. V roku 2004 vyhral 13 pretekov z celkového počtu 18, takže pravdepodobnosť jeho víťazstva bola takmer 3/4. Na základe tejto skutočnosti vytvorme nasledujúci modelový príklad.

Na štarte je 17 jazdcov –
Schumacher s pravdepodobnosťou víťazstva 3/4
a ďalších 16 rovnocenných jazdcov, z ktorých má každý šancu na víťazstvo rovnajúcu sa 1/64.

Označme $\mathbf{B} = \{B_1, B_2, \dots, B_{17}\}$ pokus, v ktorom B_1 je jav znamenajúci, že vyhral Schumacher, B_i pre $i = 2, 3, \dots, 17$ je jav, že vyhral i-ty jazdec.

Nech $P(B_1) = 3/4$, $P(B_2) = P(B_3) = \cdots = P(B_{17}) = 1/64$.
Ak sa po skončení preteku dozvieme, vyhral Schumacher, dostaneme $-\log_2(P(B_1)) = -\log_2(0.75) = 0.415$ bitov informácie.

Ak sa však dozvieme, že vyhral jazdec číslo 17, dostaneme
$-\log_2(P(B_{17})) = -\log_2(1/64) = 6$ bitov informácie.

Podmienená entropia

Michail Schumacher bol fenomenálny pilot formuly 1, ktorý získal v rokoch 1994, 1995 a 2000–2004 sedem titulov majstra sveta. V roku 2004 vyhral 13 pretekov z celkového počtu 18, takže pravdepodobnosť jeho víťazstva bola takmer 3/4. Na základe tejto skutočnosti vytvorme nasledujúci modelový príklad.

Na štarte je 17 jazdcov –
Schumacher s pravdepodobnosťou víťazstva 3/4
a ďalších 16 rovnocenných jazdcov, z ktorých má každý šancu na víťazstvo rovnajúcu sa 1/64.

Označme $\mathbf{B} = \{B_1, B_2, \ldots, B_{17}\}$ pokus, v ktorom B_1 je jav znamenajúci, že vyhral Schumacher, B_i pre $i = 2, 3, \ldots, 17$ je jav, že vyhral i-ty jazdec.

Nech $P(B_1) = 3/4$, $P(B_2) = P(B_3) = \cdots = P(B_{17}) = 1/64$.
Ak sa po skončení preteku dozvieme, vyhral Schumacher, dostaneme
$-\log_2(P(B_1)) = -\log_2(0.75) = 0.415$ bitov informácie.

Ak sa však dozvieme, že vyhral jazdec číslo 17, dostaneme
$-\log_2(P(B_{17})) = -\log_2(1/64) = 6$ bitov informácie.

Podmienená entropia

Entropia pokusu **B** je

$$H(\mathbf{B}) = H(3/4, 1/64, 1/64, \dots, 1/64) = 1.811.$$

Majme pokus $\mathbf{A} = \{A_1, A_2\}$, kde

- A_1 je jav „vyhral Schumacher“ a
- A_2 je jav „nevyhral Schumacher“.

Je $P(A_1) = 3/4$, $P(A_2) = 1/4$.
Predpokladajme, že sa po preteku dozvieme, že tentokrát Schumacher nevyhral – nastal jav A_2.
Táto správa nesie so sebou $-\log_2(P(A_2)) = -\log_2(1/4) = 2$ bity informácie.
Naša neurčitosť po tejto správe sa zmení a $H(\mathbf{B}) = 1.811$ na $H(\mathbf{B}|A_2)$.

$$H(\mathbf{B}|A_2) = H\big(P(B_1|A_2), P(B_2|A_2), \dots, P(B_{17}|A_2)\big) =$$
$$= H(0, 1/16, 1/16, \dots, 1/16) = H(1/16, 1/16, \dots, 1/16) = 4.$$

Správa „nastal jav A_2“ (t. j. „Schumacher nevyhral“) doniesla 2 bity informácie, a napriek tejto správe naša neurčitosť o výsledku preteku stúpla z $H(\mathbf{B}) = 1.811$ na hodnotu $H(\mathbf{B}|A_2) = 4$.

Podmienená entropia

Entropia pokusu **B** je

$$H(\mathbf{B}) = H\,(3/4, 1/64, 1/64, \dots, 1/64) = 1.811.$$

Majme pokus $\mathbf{A} = \{A_1, A_2\}$, kde

- A_1 je jav „vyhral Schumacher“ a
- A_2 je jav „nevyhral Schumacher“.

Je $P(A_1) = 3/4$, $P(A_2) = 1/4$.
Predpokladajme, že sa po preteku dozvieme, že tentokrát Schumacher nevyhral – nastal jav A_2.
Táto správa nesie so sebou $-\log_2(P(A_2)) = -\log_2(1/4) = 2$ bity informácie.
Naša neurčitosť po tejto správe sa zmení a $H(\mathbf{B}) = 1.811$ na $H(\mathbf{B}|A_2)$.

$$H(\mathbf{B}|A_2) = H\big(P(B_1|A_2), P(B_2|A_2), \dots, P(B_{17}|A_2)\big) =$$
$$= H(0, 1/16, 1/16, \dots, 1/16) = H(1/16, 1/16, \dots, 1/16) = 4.$$

Správa „nastal jav A_2“ (t. j. „Schumacher nevyhral“) doniesla 2 bity informácie, a napriek tejto správe naša neurčitosť o výsledku preteku stúpla z $H(\mathbf{B}) = 1.811$ na hodnotu $H(\mathbf{B}|A_2) = 4$.

Podmienená entropia

Výsledok A_2 nastane s pravdepodobnosťou $1/4$ a vtedy $H(\mathbf{B}|A_2) = 4$.
Výsledok A_1 nastane s pravdepodobnosťou $3/4$ a vtedy $H(\mathbf{B}|A_1) = 0$.

Stredná hodnota zvyškovej neurčitosti pokusu **B** po vykonaní pokusu **A** sa bude rovnať

$$P(A_1).H(\mathbf{B}|A_1) + P(A_2).H(\mathbf{B}|A_2) = (3/4).0 + (1/4).4 = 1.$$

Stredná hodnota zvyškovej entropie pokusu **B** po vykonaní pokusu **A** bude 1 bit.

Definícia

Nech sú dané dva pokusy

$$\mathbf{A} = \{A_1, A_2, \ldots, A_n\}, \quad \mathbf{B} = \{B_1, B_2, \ldots, B_m\}.$$

Podmienenou entropiu pokusu B za predpokladu, vykonania pokusu A *(alebo len za podmienky* **A***) je*

$$H(\mathbf{B}|\mathbf{A}) = \sum_{i=1}^{n} P(A_i).H(\mathbf{B}|A_i). \tag{23}$$

Podmienená entropia

Výsledok A_2 nastane s pravdepodobnosťou $1/4$ a vtedy $H(\mathbf{B}|A_2) = 4$.
Výsledok A_1 nastane s pravdepodobnosťou $3/4$ a vtedy $H(\mathbf{B}|A_1) = 0$.

Stredná hodnota zvyškovej neurčitosti pokusu **B** po vykonaní pokusu **A** sa bude rovnať

$$P(A_1).H(\mathbf{B}|A_1) + P(A_2).H(\mathbf{B}|A_2) = (3/4).0 + (1/4).4 = 1.$$

Stredná hodnota zvyškovej entropie pokusu **B** po vykonaní pokusu **A** bude 1 bit.

Definícia

Nech sú dané dva pokusy

$$\mathbf{A} = \{A_1, A_2, \dots, A_n\}, \quad \mathbf{B} = \{B_1, B_2, \dots, B_m\}.$$

Podmienenou entropiu pokusu B za predpokladu, vykonania pokusu A *(alebo len za podmienky* **A***) je*

$$H(\mathbf{B}|\mathbf{A}) = \sum_{i=1}^{n} P(A_i).H(\mathbf{B}|A_i). \qquad (23)$$

Podmienená entropia

Platí:

$$\begin{aligned}
\sum_{i=1}^{n} P(A_i).H(\mathbf{B}|A_i) &= \sum_{i=1}^{n} P(A_i).\underbrace{H\big(P(B_1|A_i), P(B_2|A_i), \dots, P(B_m|A_i)\big)}_{\sum_{j=1}^{m} P(B_j|A_i).\log_2(P(B_j|A_i))} = \\
&= -\sum_{i=1}^{n}\sum_{j=1}^{m} P(A_i).P(B_j|A_i).\log_2(P(B_j|A_i)) = \\
&= -\sum_{i=1}^{n}\sum_{j=1}^{m} P(A_i).\frac{P(A_i \cap B_j)}{P(A_i)}.\log_2\left(\frac{P(A_i \cap B_j)}{P(A_i)}\right) = \\
&= -\sum_{i=1}^{n}\sum_{j=1}^{m} P(A_i \cap B_j).\log_2\left(\frac{P(A_i \cap B_j)}{P(A_i)}\right).
\end{aligned}$$

Môžeme teda tiež písať

$$\boxed{H(\mathbf{B}|\mathbf{A}) = -\sum_{i=1}^{n}\sum_{j=1}^{m} P(A_i \cap B_j).\log_2\left(\frac{P(A_i \cap B_j)}{P(A_i)}\right)} \qquad (24)$$

Definícia

Nech $\mathbf{A} = \{A_1, A_2, \dots, A_n\}$, $\mathbf{B} = \{B_1, B_2, \dots, B_m\}$ *sú dva pokusy na pravdepodobnostnom priestore* $(\Omega, \mathcal{A}, P)$.
Kombinovaným pokusom pokusov A, B *nazveme pokus*

$$\mathbf{A} \wedge \mathbf{B} = \left\{A_i \cap B_j \mid A_i \in \mathbf{A},\ B_j \in \mathbf{B}\right\} \tag{25}$$

Ak najprv vykonáme pokus $\mathbf{A}$ a potom pokus $\mathbf{B}$, (alebo aj najprv $\mathbf{B}$ a potom $\mathbf{A}$), dozvieme sa to isté, t. j. získame rovnakú informáciu, ako keby sme vykonali pokus $\mathbf{A} \wedge \mathbf{B}$.
Ak už vykonáme pokus $\mathbf{A}$ a jeho výsledok je A_i, podmienená entropia pokusu $\mathbf{B}$ za predpokladu, že nastal jav A_i, je $H(\mathbf{B}|A_i)$. Keďže jav A_i má pravdepodobnosť $P(A_i)$, jeho príspevok k celkovej strednej hodnote pokusu $\mathbf{B}$ za predpokladu, že je známy výsledok pokusu $\mathbf{A}$, je $P(A_i).H(\mathbf{B}|A_i)$ a podmienená entropia pokusu $\mathbf{B}$ za predpokladu, že poznáme výsledok pokusu $\mathbf{A}$, je $H(\mathbf{B}|\mathbf{A}) = \sum_{i=1}^{n} P(A_i).H(\mathbf{B}|A_i)$.

Entropia kombinovaného pokusu

Definícia

Nech $\mathbf{A} = \{A_1, A_2, \dots, A_n\}$, $\mathbf{B} = \{B_1, B_2, \dots, B_m\}$ *sú dva pokusy na pravdepodobnostnom priestore* $(\Omega, \mathcal{A}, P)$.
Kombinovaným pokusom pokusov A, B *nazveme pokus*

$$\mathbf{A} \wedge \mathbf{B} = \big\{A_i \cap B_j \mid A_i \in \mathbf{A},\ B_j \in \mathbf{B}\big\} \tag{25}$$

Ak najprv vykonáme pokus **A** a potom pokus **B**, (alebo aj najprv **B** a potom **A**), dozvieme sa to isté, t. j. získame rovnakú informáciu, ako keby sme vykonali pokus $\mathbf{A} \wedge \mathbf{B}$.
Ak už vykonáme pokus **A** a jeho výsledok je A_i, podmienená entropia pokusu **B** za predpokladu, že nastal jav A_i, je $H(\mathbf{B}|A_i)$. Keďže jav A_i má pravdepodobnosť $P(A_i)$, jeho príspevok k celkovej strednej hodnote pokusu **B** za predpokladu, že je známy výsledok pokusu **A**, je $P(A_i).H(\mathbf{B}|A_i)$ a podmienená entropia pokusu **B** za predpokladu, že poznáme výsledok pokusu **A**, je $H(\mathbf{B}|\mathbf{A}) = \sum_{i=1}^{n} P(A_i).H(\mathbf{B}|A_i)$.

Entropia kombinovaného pokusu

Nech pre $i = 1, 2, \ldots, n$ máme $p_i = q_{i1} + q_{i2} + \cdots + q_{im_i} > 0$. Potom

$$H(q_{11}, q_{12} \ldots q_{1m_1}, q_{21}, q_{22}, \ldots, q_{2m_2}, \ldots, q_{n1}, q_{n2}, \ldots, q_{nm_n}) =$$

$$= H(p_1, p_2, \ldots, p_n) + \sum_{i=1}^{n} p_i . H\left(\frac{q_{i1}}{p_i}, \frac{q_{i2}}{p_i}, \ldots, \frac{q_{im_i}}{p_i}\right)$$

Entropia kombinovaného pokusu

Vezmime pokus $\mathbf{A} \wedge \mathbf{B}$. Označme $q_{ij} = P(A_i \cap B_j)$, $p_i = P(A_i)$. Potom platí

$$p_i = P(A_i) = \sum_{j=1}^{m} p(A_i \cap B_j) = \sum_{j=1}^{m} q_{ij}.$$

Predpoklady vety 6 sú teda splnené a preto je

$$\begin{aligned}
H(\mathbf{A} \wedge \mathbf{B}) &= H\left(\underbrace{q_{11}, q_{12}, \dots q_{1m}}_{p_1}, \underbrace{q_{21}, q_{22}, \dots q_{2m}}_{p_2}, \dots, \underbrace{q_{n1}, q_{n2}, \dots q_{nm}}_{p_n}\right) = \\
&= H(p_1, p_2, \dots, p_n) + \sum_{j=1}^{m} p_i . H\left(\frac{q_{i1}}{p_i}, \frac{q_{i2}}{p_i}, \dots, \frac{q_{im}}{p_i}\right) = \\
&= H\big(P(A_1), P(A_2), \dots, P(A_n)\big) + \\
&+ \sum_{i=1}^{m} P(A_i) H\left(\frac{P(A_i \cap B_1)}{P(A_i)}, \frac{P(A_i \cap B_2)}{P(A_i)}, \dots, \frac{P(A_i \cap B_m)}{P(A_i)}\right) = \\
&= H(\mathbf{A}) + H(\mathbf{B}|\mathbf{A})
\end{aligned}$$

Entropia kombinovaného pokusu

Teda platí nasledujúca veta:

Veta

$$H(\mathbf{A} \wedge \mathbf{B}) = H(\mathbf{A}) + H(\mathbf{B}|\mathbf{A}) \qquad (26)$$

$H(\mathbf{B}|\mathbf{A})$ je zvyšková entropia kombinovaného pokusu $\mathbf{A} \wedge \mathbf{B}$ po vykonaní pokusu $\mathbf{A}$.

Vidíme tiež, že o čo je entropia $H(\mathbf{A})$ pokusu $\mathbf{A}$ väčšia, o to menšia je podmienená entropia $H(\mathbf{B}|\mathbf{A})$.

Definícia

Nech $\mathbf{A} = \{A_1, A_2, \dots, A_n\}$, $\mathbf{B} = \{B_1, B_2, \dots, B_m\}$ *sú dva pokusy na pravdepodobnostnom priestore* $(\Omega, \mathcal{A}, P)$.
Hovoríme, že pokusy $\mathbf{A}$, $\mathbf{B}$ *sú* **štatisticky nezávislé** *(alebo len nezávislé), ak pre každé* $i = 1, 2, \dots, n$, $j = 1, 2, \dots, m$ *sú* A_i, B_j *nezávislé javy.*

Entropia kombinovaného pokusu

Teda platí nasledujúca veta:

Veta

$$H(\mathbf{A} \wedge \mathbf{B}) = H(\mathbf{A}) + H(\mathbf{B}|\mathbf{A}) \tag{26}$$

$H(\mathbf{B}|\mathbf{A})$ je zvyšková entropia kombinovaného pokusu $\mathbf{A} \wedge \mathbf{B}$ po vykonaní pokusu $\mathbf{A}$.

Vidíme tiež, že o čo je entropia $H(\mathbf{A})$ pokusu $\mathbf{A}$ väčšia, o to menšia je podmienená entropia $H(\mathbf{B}|\mathbf{A})$.

Definícia

Nech $\mathbf{A} = \{A_1, A_2, \dots, A_n\}$, $\mathbf{B} = \{B_1, B_2, \dots, B_m\}$ *sú dva pokusy na pravdepodobnostnom priestore* $(\Omega, \mathcal{A}, P)$.
Hovoríme, že pokusy $\mathbf{A}$, $\mathbf{B}$ *sú* **štatisticky nezávislé** *(alebo len nezávislé), ak pre každé* $i = 1, 2, \dots, n$, $j = 1, 2, \dots, m$ *sú* A_i, B_j *nezávislé javy.*

Entropia kombinovaného pokusu

Teda platí nasledujúca veta:

Veta

$$H(\mathbf{A} \wedge \mathbf{B}) = H(\mathbf{A}) + H(\mathbf{B}|\mathbf{A}) \tag{26}$$

$H(\mathbf{B}|\mathbf{A})$ je zvyšková entropia kombinovaného pokusu $\mathbf{A} \wedge \mathbf{B}$ po vykonaní pokusu $\mathbf{A}$.

Vidíme tiež, že o čo je entropia $H(\mathbf{A})$ pokusu $\mathbf{A}$ väčšia, o to menšia je podmienená entropia $H(\mathbf{B}|\mathbf{A})$.

Definícia

Nech $\mathbf{A} = \{A_1, A_2, \dots, A_n\}$, $\mathbf{B} = \{B_1, B_2, \dots, B_m\}$ *sú dva pokusy na pravdepodobnostnom priestore* $(\Omega, \mathcal{A}, P)$.
Hovoríme, že pokusy $\mathbf{A}$, $\mathbf{B}$ *sú* **štatisticky nezávislé** *(alebo len nezávislé), ak pre každé* $i = 1, 2, \dots, n$, $j = 1, 2, \dots, m$ *sú* A_i, B_j *nezávislé javy.*

Spoločná informácia pokusov

Zaujímame sa o výsledok pokusu **B** s entropiou $H(\mathbf{B})$.

Tento pokus z nejakých dôvodov nemôžeme vykonať, ale vykonáme pokus **A**.

Výsledky pokusu **A** zmenia neurčitosť pokusu **B** z $H(\mathbf{B})$ na $H(\mathbf{B}|\mathbf{A})$

$H(\mathbf{B}|\mathbf{A})$ je stredné množstvo dodatočnej informácie, ktorú možno získať z pokusu **B** po vykonaní pokusu **A**.

Rozdiel $H(\mathbf{B}) - H(\mathbf{B}|\mathbf{A})$ možno považovať za stredné množstvo informácie o pokuse **B** obsiahnuté v pokuse **A**.

Definícia

Stredné množstvo informácie $I(\mathbf{A}, \mathbf{B})$ o pokuse B v pokuse A *je*

$$I(\mathbf{A}, \mathbf{B}) = H(\mathbf{B}) - H(\mathbf{B}|\mathbf{A}) \tag{27}$$

Spoločná informácia pokusov

Zaujímame sa o výsledok pokusu $\mathbf{B}$ s entropiou $H(\mathbf{B})$.

Tento pokus z nejakých dôvodov nemôžeme vykonať, ale vykonáme pokus $\mathbf{A}$.

Výsledky pokusu $\mathbf{A}$ zmenia neurčitosť pokusu $\mathbf{B}$ z $H(\mathbf{B})$ na $H(\mathbf{B}|\mathbf{A})$

$H(\mathbf{B}|\mathbf{A})$ je stredné množstvo dodatočnej informácie, ktorú možno získať z pokusu $\mathbf{B}$ po vykonaní pokusu $\mathbf{A}$.

Rozdiel $H(\mathbf{B}) - H(\mathbf{B}|\mathbf{A})$ možno považovať za stredné množstvo informácie o pokuse $\mathbf{B}$ obsiahnuté v pokuse $\mathbf{A}$.

Definícia

Stredné množstvo informácie $I(\mathbf{A}, \mathbf{B})$ o pokuse B v pokuse A *je*

$$I(\mathbf{A}, \mathbf{B}) = H(\mathbf{B}) - H(\mathbf{B}|\mathbf{A}) \tag{27}$$

Spoločná informácia pokusov

Veta

$$I(\mathbf{A}, \mathbf{B}) = H(\mathbf{A}) + H(\mathbf{B}) - H(\mathbf{A} \wedge \mathbf{B}) \tag{28}$$

Podľa definície je

$$I(\mathbf{A}, \mathbf{B}) = H(\mathbf{B}) - H(\mathbf{B}|\mathbf{A})$$

Pred časom sme dokázali

$$H(\mathbf{A} \wedge \mathbf{B}) = H(\mathbf{A}) + H(\mathbf{B}|\mathbf{A})$$

Dosadením za $H(\mathbf{B}|\mathbf{A}) = H(\mathbf{A} \wedge \mathbf{B}) - H(\mathbf{A})$ do prvého vzťahu dostaneme žiadaný vzťah. □

Zo vzťahu (28) vidíme, že $I(\mathbf{A}, \mathbf{B}) = I(\mathbf{B}, \mathbf{A})$, t. j., že informácia o pokuse **B** obsiahnutá v pokuse **A** sa rovná informácii o pokuse **A** obsiahnutej v pokuse **B**.

Preto sa niekedy hodnote $I(\mathbf{A}, \mathbf{B})$ hovorí aj

spoločná informácia pokusov A, B

.

Spoločná informácia pokusov

Veta

$$I(\mathbf{A},\mathbf{B}) = H(\mathbf{A}) + H(\mathbf{B}) - H(\mathbf{A} \wedge \mathbf{B}) \tag{28}$$

Podľa definície je

$$I(\mathbf{A},\mathbf{B}) = H(\mathbf{B}) - H(\mathbf{B}|\mathbf{A})$$

Pred časom sme dokázali

$$H(\mathbf{A} \wedge \mathbf{B}) = H(\mathbf{A}) + H(\mathbf{B}|\mathbf{A})$$

Dosadením za $H(\mathbf{B}|\mathbf{A}) = H(\mathbf{A} \wedge \mathbf{B}) - H(\mathbf{A})$ do prvého vzťahu dostaneme žiadaný vzťah. □

Zo vzťahu (28) vidíme, že $I(\mathbf{A},\mathbf{B}) = I(\mathbf{B},\mathbf{A})$, t. j., že informácia o pokuse **B** obsiahnutá v pokuse **A** sa rovná informácii o pokuse **A** obsiahnutej v pokuse **B**.

Preto sa niekedy hodnote $I(\mathbf{A},\mathbf{B})$ hovorí aj

spoločná informácia pokusov A, B

.

Spoločná informácia pokusov

Veta

$$I(\mathbf{A},\mathbf{B}) = H(\mathbf{A}) + H(\mathbf{B}) - H(\mathbf{A}\wedge\mathbf{B}) \tag{28}$$

Podľa definície je

$$I(\mathbf{A},\mathbf{B}) = H(\mathbf{B}) - H(\mathbf{B}|\mathbf{A})$$

Pred časom sme dokázali

$$H(\mathbf{A}\wedge\mathbf{B}) = H(\mathbf{A}) + H(\mathbf{B}|\mathbf{A})$$

Dosadením za $H(\mathbf{B}|\mathbf{A}) = H(\mathbf{A}\wedge\mathbf{B}) - H(\mathbf{A})$ do prvého vzťahu dostaneme žiadaný vzťah. □

Zo vzťahu (28) vidíme, že $I(\mathbf{A},\mathbf{B}) = I(\mathbf{B},\mathbf{A})$, t. j., že informácia o pokuse $\mathbf{B}$ obsiahnutá v pokuse $\mathbf{A}$ sa rovná informácii o pokuse $\mathbf{A}$ obsiahnutej v pokuse $\mathbf{B}$.

Preto sa niekedy hodnote $I(\mathbf{A},\mathbf{B})$ hovorí aj

spoločná informácia pokusov A, B

.

Spoločná informácia pokusov

Veta

$$I(\mathbf{A}, \mathbf{B}) = H(\mathbf{A}) + H(\mathbf{B}) - H(\mathbf{A} \wedge \mathbf{B}) \tag{28}$$

Podľa definície je

$$I(\mathbf{A}, \mathbf{B}) = H(\mathbf{B}) - H(\mathbf{B}|\mathbf{A})$$

Pred časom sme dokázali

$$H(\mathbf{A} \wedge \mathbf{B}) = H(\mathbf{A}) + H(\mathbf{B}|\mathbf{A})$$

Dosadením za $H(\mathbf{B}|\mathbf{A}) = H(\mathbf{A} \wedge \mathbf{B}) - H(\mathbf{A})$ do prvého vzťahu dostaneme žiadaný vzťah. □

Zo vzťahu (28) vidíme, že $I(\mathbf{A}, \mathbf{B}) = I(\mathbf{B}, \mathbf{A})$, t. j., že informácia o pokuse **B** obsiahnutá v pokuse **A** sa rovná informácii o pokuse **A** obsiahnutej v pokuse **B**.

Preto sa niekedy hodnote $I(\mathbf{A}, \mathbf{B})$ hovorí aj

spoločná informácia pokusov A, B

.

Spoločná informácia pokusov

Veta

$$I(\mathbf{A},\mathbf{B}) = H(\mathbf{A}) + H(\mathbf{B}) - H(\mathbf{A} \wedge \mathbf{B}) \tag{28}$$

Podľa definície je

$$I(\mathbf{A},\mathbf{B}) = H(\mathbf{B}) - H(\mathbf{B}|\mathbf{A})$$

Pred časom sme dokázali

$$H(\mathbf{A} \wedge \mathbf{B}) = H(\mathbf{A}) + H(\mathbf{B}|\mathbf{A})$$

Dosadením za $H(\mathbf{B}|\mathbf{A}) = H(\mathbf{A} \wedge \mathbf{B}) - H(\mathbf{A})$ do prvého vzťahu dostaneme žiadaný vzťah. □

Zo vzťahu (28) vidíme, že $I(\mathbf{A},\mathbf{B}) = I(\mathbf{B},\mathbf{A})$, t. j., že informácia o pokuse **B** obsiahnutá v pokuse **A** sa rovná informácii o pokuse **A** obsiahnutej v pokuse **B**.

Preto sa niekedy hodnote $I(\mathbf{A},\mathbf{B})$ hovorí aj

spoločná informácia pokusov A, B

.

Spoločná informácia pokusov

Veta

Nech $\mathbf{A} = \{A_1, A_2, \dots, A_n\}$, $\mathbf{B} = \{B_1, B_2, \dots, B_m\}$ *sú dva pokusy na pravdepodobnostnom priestore* $(\Omega, \mathcal{A}, P)$. *Potom*

$$I(\mathbf{A}, \mathbf{B}) = \sum_{i=1}^{n} \sum_{j=1}^{m} P(A_i \cap B_j). \log_2 \left(\frac{P(A_i \cap B_j)}{P(A_i).P(B_j)} \right) . \tag{29}$$

Dôkaz.
Pretože $\mathbf{A} = \{A_1, A_2, \dots, A_n\}$ je rozklad priestoru Ω je

$$B_j = B_j \cap \Omega = B_j \cap \bigcup_{i=1}^{n} A_i = \bigcup_{i=1}^{n} A_i \cap B_j .$$

Pretože zjednotenie na pravej strane posledného výrazu je disjunktné, je

$$P(B_j) = \sum_{i=1}^{n} P(A_i \cap B_j) .$$

Spoločná informácia pokusov

Veta

Nech $\mathbf{A} = \{A_1, A_2, \ldots, A_n\}$, $\mathbf{B} = \{B_1, B_2, \ldots, B_m\}$ *sú dva pokusy na pravdepodobnostnom priestore* $(\Omega, \mathcal{A}, P)$. *Potom*

$$I(\mathbf{A}, \mathbf{B}) = \sum_{i=1}^{n} \sum_{j=1}^{m} P(A_i \cap B_j).\log_2 \left(\frac{P(A_i \cap B_j)}{P(A_i).P(B_j)} \right) \ . \tag{29}$$

Dôkaz.
Pretože $\mathbf{A} = \{A_1, A_2, \ldots, A_n\}$ je rozklad priestoru Ω je

$$B_j = B_j \cap \Omega = B_j \cap \bigcup_{i=1}^{n} A_i = \bigcup_{i=1}^{n} A_i \cap B_j \ .$$

Pretože zjednotenie na pravej strane posledného výrazu je disjunktné, je

$$P(B_j) = \sum_{i=1}^{n} P(A_i \cap B_j) \ .$$

Spoločná informácia pokusov

Dosadením za $H(\mathbf{B}|\mathbf{A})$ zo vzťahu (24) do definičnej rovnosti $I(\mathbf{A},\mathbf{B})$ dostávame

$$I(\mathbf{A},\mathbf{B}) = H(\mathbf{B}) - H(\mathbf{B}|\mathbf{A}) =$$

$$= -\sum_{j=1}^{m} P(B_j).\log_2 P(B_j) + \sum_{i=1}^{n}\sum_{j=1}^{m} P(A_i \cap B_j).\log_2\left(\frac{P(A_i \cap B_j)}{P(A_i)}\right) =$$

$$= -\sum_{j=1}^{m}\sum_{i=1}^{n} P(A_i \cap B_j).\log_2 P(B_j) + \sum_{i=1}^{n}\sum_{j=1}^{m} P(A_i \cap B_j).\log_2\left(\frac{P(A_i \cap B_j)}{P(A_i)}\right) =$$

$$= \sum_{i=1}^{n}\sum_{j=1}^{m} P(A_i \cap B_j).\left[\log_2\left(\frac{P(A_i \cap B_j)}{P(A_i)}\right) - \log_2 P(B_j)\right] =$$

$$= \sum_{i=1}^{n}\sum_{j=1}^{m} P(A_i \cap B_j).\log_2\left(\frac{P(A_i \cap B_j)}{P(A_i).P(B_j)}\right)$$

$\square$

Veta

Nech $\mathbf{A} = \{A_1, A_2, \ldots, A_n\}$, $\mathbf{B} = \{B_1, B_2, \ldots, B_m\}$ *sú dva pokusy na pravdepodobnostnom priestore* $(\Omega, \mathcal{A}, P)$*. Potom*

$$0 \leq I(\mathbf{A}, \mathbf{B}), \tag{30}$$

pričom rovnosť nastáva práve vtedy, keď sú pokusy $\mathbf{A}$, $\mathbf{B}$ *štatisticky nezávislé.*

Dôkaz. $\log_2 x = \ln 2 . \ln x$
Použijeme vzťah (29) z vety 14 a nerovnosť $\ln x \leq x - 1$, ktorá platí pre všetky $x > 0$, pričom rovnosť nastáva práve vtedy, keď $x = 1$.

$$P(A_i \cap B_j). \log_2 \left(\frac{P(A_i).P(B_j)}{P(A_i \cap B_j)} \right) = P(A_i \cap B_j). \frac{1}{\ln(2)}. \ln \left(\frac{P(A_i).P(B_j)}{P(A_i \cap B_j)} \right) \leq$$

$$\leq P(A_i \cap B_j). \frac{1}{\ln(2)}. \left[\left(\frac{P(A_i).P(B_j)}{P(A_i \cap B_j)} \right) - 1 \right] = \frac{1}{\ln(2)}. [P(A_i).P(B_j) - P(A_i \cap B_j)],$$

pričom rovnosť platí práve vtedy, keď $\dfrac{P(A_i).P(B_j)}{P(A_i \cap B_j)} = 1$, t. j. vtedy, keď sú javy A_i, B_j nezávislé.

Spoločná informácia pokusov

Veta

Nech $\mathbf{A} = \{A_1, A_2, \dots, A_n\}$, $\mathbf{B} = \{B_1, B_2, \dots, B_m\}$ *sú dva pokusy na pravdepodobnostnom priestore* $(\Omega, \mathcal{A}, P)$. *Potom*

$$0 \le I(\mathbf{A}, \mathbf{B}), \tag{30}$$

pričom rovnosť nastáva práve vtedy, keď sú pokusy **A**, **B** *štatisticky nezávislé.*

Dôkaz. $\log_2 x = \ln 2 . \ln x$
Použijeme vzťah (29) z vety 14 a nerovnosť $\ln x \le x - 1$, ktorá platí pre všetky $x > 0$, pričom rovnosť nastáva práve vtedy, keď $x = 1$.

$$P(A_i \cap B_j). \log_2 \left(\frac{P(A_i).P(B_j)}{P(A_i \cap B_j)} \right) = P(A_i \cap B_j). \frac{1}{\ln(2)}. \ln \left(\frac{P(A_i).P(B_j)}{P(A_i \cap B_j)} \right) \le$$

$$\le P(A_i \cap B_j). \frac{1}{\ln(2)}. \left[\left(\frac{P(A_i).P(B_j)}{P(A_i \cap B_j)} \right) - 1 \right] = \frac{1}{\ln(2)}. \left[P(A_i).P(B_j) - P(A_i \cap B_j) \right],$$

pričom rovnosť platí práve vtedy, keď $\dfrac{P(A_i).P(B_j)}{P(A_i \cap B_j)} = 1$, t. j. vtedy, keď sú javy A_i, B_j nezávislé.

Veta

Nech $\mathbf{A} = \{A_1, A_2, \dots, A_n\}$, $\mathbf{B} = \{B_1, B_2, \dots, B_m\}$ *sú dva pokusy na pravdepodobnostnom priestore* $(\Omega, \mathcal{A}, P)$. *Potom*

$$0 \le I(\mathbf{A}, \mathbf{B}), \tag{30}$$

pričom rovnosť nastáva práve vtedy, keď sú pokusy $\mathbf{A}$, $\mathbf{B}$ *štatisticky nezávislé.*

Dôkaz. $\log_2 x = \ln 2 . \ln x$
Použijeme vzťah (29) z vety 14 a nerovnosť $\ln x \le x - 1$, ktorá platí pre všetky $x > 0$, pričom rovnosť nastáva práve vtedy, keď $x = 1$.

$$P(A_i \cap B_j). \log_2 \left(\frac{P(A_i).P(B_j)}{P(A_i \cap B_j)} \right) = P(A_i \cap B_j). \frac{1}{\ln(2)}. \ln \left(\frac{P(A_i).P(B_j)}{P(A_i \cap B_j)} \right) \le$$

$$\le P(A_i \cap B_j). \frac{1}{\ln(2)}. \left[\left(\frac{P(A_i).P(B_j)}{P(A_i \cap B_j)} \right) - 1 \right] = \frac{1}{\ln(2)}. \left[P(A_i).P(B_j) - P(A_i \cap B_j) \right],$$

pričom rovnosť platí práve vtedy, keď $\dfrac{P(A_i).P(B_j)}{P(A_i \cap B_j)} = 1$, t. j. vtedy, keď sú javy A_i, B_j nezávislé.

Spoločná informácia pokusov

Použitím práve dokázanej nerovnosti máme

$$
\begin{aligned}
-I(\mathbf{A},\mathbf{B}) &= \sum_{i=1}^{n}\sum_{j=1}^{m} P(A_i \cap B_j).\log_2\left(\frac{P(A_i).P(B_j)}{P(A_i \cap B_j)}\right) \leq \\
&\leq \frac{1}{\ln(2)}.\left[\sum_{i=1}^{n}\sum_{j=1}^{m}\left(P(A_i).P(B_j) - P(A_i \cap B_j)\right)\right] = \\
&= \frac{1}{\ln(2)}.\left[\sum_{i=1}^{n}\sum_{j=1}^{m} P(A_i).P(B_j) - \underbrace{\sum_{i=1}^{n}\sum_{j=1}^{m} P(A_i \cap B_j)}_{=1}\right] = \\
&= \frac{1}{\ln(2)}.\left[\sum_{i=1}^{n} P(A_i)\underbrace{\sum_{j=1}^{m} P(B_j)}_{=1} - 1\right] = \frac{1}{\ln(2)}.\left[\underbrace{\sum_{i=1}^{n} P(A_i)}_{=1} - 1\right] = 0,
\end{aligned}
$$

pričom rovnosť platí práve vtedy, keď sú všetky dvojice javov A_i, B_j pre $i = 1, 2, \ldots, n$, $j = 1, 2, \ldots, m$ nezávislé. □

Spoločná informácia pokusov

Použitím práve dokázanej nerovnosti máme

$$\begin{aligned} -I(\mathbf{A},\mathbf{B}) &= \sum_{i=1}^{n}\sum_{j=1}^{m} P(A_i\cap B_j).\log_2\left(\frac{P(A_i).P(B_j)}{P(A_i\cap B_j)}\right) \le \\ &\le \frac{1}{\ln(2)}.\left[\sum_{i=1}^{n}\sum_{j=1}^{m}\left(P(A_i).P(B_j) - P(A_i\cap B_j)\right)\right] = \\ &= \frac{1}{\ln(2)}.\left[\sum_{i=1}^{n}\sum_{j=1}^{m} P(A_i).P(B_j) - \underbrace{\sum_{i=1}^{n}\sum_{j=1}^{m} P(A_i\cap B_j)}_{=1}\right] = \\ &= \frac{1}{\ln(2)}.\left[\sum_{i=1}^{n} P(A_i)\underbrace{\sum_{j=1}^{m} P(B_j)}_{=1} - 1\right] = \frac{1}{\ln(2)}.\left[\underbrace{\sum_{i=1}^{n} P(A_i)}_{=1} - 1\right] = 0, \end{aligned}$$

pričom rovnosť platí práve vtedy, keď sú všetky dvojice javov A_i, B_j pre $i = 1, 2, \dots, n$, $j = 1, 2, \dots, m$ nezávislé. □

Spoločná informácia pokusov

Použitím práve dokázanej nerovnosti máme

$$
\begin{aligned}
-I(\mathbf{A},\mathbf{B}) &= \sum_{i=1}^{n}\sum_{j=1}^{m} P(A_i\cap B_j).\log_2\left(\frac{P(A_i).P(B_j)}{P(A_i\cap B_j)}\right) \leq \\
&\leq \frac{1}{\ln(2)}.\left[\sum_{i=1}^{n}\sum_{j=1}^{m}\left(P(A_i).P(B_j) - P(A_i\cap B_j)\right)\right] = \\
&= \frac{1}{\ln(2)}.\left[\sum_{i=1}^{n}\sum_{j=1}^{m} P(A_i).P(B_j) - \underbrace{\sum_{i=1}^{n}\sum_{j=1}^{m} P(A_i\cap B_j)}_{=1}\right] = \\
&= \frac{1}{\ln(2)}.\left[\sum_{i=1}^{n} P(A_i)\underbrace{\sum_{j=1}^{m} P(B_j)}_{=1} - 1\right] = \frac{1}{\ln(2)}.\left[\underbrace{\sum_{i=1}^{n} P(A_i)}_{=1} - 1\right] = 0,
\end{aligned}
$$

pričom rovnosť platí práve vtedy, keď sú všetky dvojice javov A_i, B_j pre $i = 1, 2, \ldots, n$, $j = 1, 2, \ldots, m$ nezávislé. □

Spoločná informácia pokusov

Použitím práve dokázanej nerovnosti máme

$$
\begin{aligned}
-I(\mathbf{A},\mathbf{B}) &= \sum_{i=1}^{n}\sum_{j=1}^{m} P(A_i \cap B_j).\log_2\left(\frac{P(A_i).P(B_j)}{P(A_i \cap B_j)}\right) \le \\
&\le \frac{1}{\ln(2)}.\left[\sum_{i=1}^{n}\sum_{j=1}^{m}\left(P(A_i).P(B_j) - P(A_i \cap B_j)\right)\right] = \\
&= \frac{1}{\ln(2)}.\left[\sum_{i=1}^{n}\sum_{j=1}^{m} P(A_i).P(B_j) - \underbrace{\sum_{i=1}^{n}\sum_{j=1}^{m} P(A_i \cap B_j)}_{=1}\right] = \\
&= \frac{1}{\ln(2)}.\left[\sum_{i=1}^{n} P(A_i)\underbrace{\sum_{j=1}^{m} P(B_j)}_{=1} - 1\right] = \frac{1}{\ln(2)}.\left[\underbrace{\sum_{i=1}^{n} P(A_i)}_{=1} - 1\right] = 0,
\end{aligned}
$$

pričom rovnosť platí práve vtedy, keď sú všetky dvojice javov A_i, B_j pre $i = 1, 2, \dots, n$, $j = 1, 2, \dots, m$ nezávislé. □

Entropia kombinovaného pokusu

Veta

$$H(\mathbf{B}|\mathbf{A}) \leq H(\mathbf{B}), \tag{31}$$

pričom rovnosť nastáva práve vtedy, keď sú pokusy **A**, **B** *štatisticky nezávislé.*

Dôkaz.

Pretože $0 \leq I(\mathbf{A}, \mathbf{B})$ s rovnosťou práve vtedy, keď sú pokusy **A**, **B** štatisticky nezávislé, máme

$$\begin{aligned} 0 &\leq I(\mathbf{A}, \mathbf{B}) = H(\mathbf{B}) - H(\mathbf{B}|\mathbf{A}) \\ 0 &\leq H(\mathbf{B}) - H(\mathbf{B}|\mathbf{A}) \\ H(\mathbf{B}|\mathbf{A}) &\leq H(\mathbf{B}), \end{aligned}$$

kde rovnosť nastáva práve vtedy, keď sú pokusy **A**, **B** štatisticky nezávislé. □

Veta

$$H(\mathbf{B}|\mathbf{A}) \leq H(\mathbf{B}), \qquad (31)$$

pričom rovnosť nastáva práve vtedy, keď sú pokusy **A**, **B** *štatisticky nezávislé.*

Dôkaz.
Pretože $0 \leq I(\mathbf{A}, \mathbf{B})$ s rovnosťou práve vtedy, keď sú pokusy **A**, **B** štatisticky nezávislé, máme

$$\begin{aligned} 0 &\leq I(\mathbf{A}, \mathbf{B}) = H(\mathbf{B}) - H(\mathbf{B}|\mathbf{A}) \\ 0 &\leq H(\mathbf{B}) - H(\mathbf{B}|\mathbf{A}) \\ H(\mathbf{B}|\mathbf{A}) &\leq H(\mathbf{B}), \end{aligned}$$

kde rovnosť nastáva práve vtedy, keď sú pokusy **A**, **B** štatisticky nezávislé. □

Entropia kombinovaného pokusu

Veta

$$H(\mathbf{B}|\mathbf{A}) \leq H(\mathbf{B}), \tag{31}$$

pričom rovnosť nastáva práve vtedy, keď sú pokusy **A**, **B** *štatisticky nezávislé.*

Dôkaz.
Pretože $0 \leq I(\mathbf{A}, \mathbf{B})$ s rovnosťou práve vtedy, keď sú pokusy **A**, **B** štatisticky nezávislé, máme

$$\begin{aligned} 0 &\leq I(\mathbf{A}, \mathbf{B}) = H(\mathbf{B}) - H(\mathbf{B}|\mathbf{A}) \\ 0 &\leq H(\mathbf{B}) - H(\mathbf{B}|\mathbf{A}) \\ H(\mathbf{B}|\mathbf{A}) &\leq H(\mathbf{B}), \end{aligned}$$

kde rovnosť nastáva práve vtedy, keď sú pokusy **A**, **B** štatisticky nezávislé. □

Veta

$$H(\mathbf{B}|\mathbf{A}) \leq H(\mathbf{B}), \tag{31}$$

pričom rovnosť nastáva práve vtedy, keď sú pokusy **A**, **B** *štatisticky nezávislé.*

Dôkaz.
Pretože $0 \leq I(\mathbf{A}, \mathbf{B})$ s rovnosťou práve vtedy, keď sú pokusy **A**, **B** štatisticky nezávislé, máme

$$\begin{aligned} 0 &\leq I(\mathbf{A}, \mathbf{B}) = H(\mathbf{B}) - H(\mathbf{B}|\mathbf{A}) \\ 0 &\leq H(\mathbf{B}) - H(\mathbf{B}|\mathbf{A}) \\ H(\mathbf{B}|\mathbf{A}) &\leq H(\mathbf{B}), \end{aligned}$$

kde rovnosť nastáva práve vtedy, keď sú pokusy **A**, **B** štatisticky nezávislé. □

Veta

$$H(\mathbf{A} \wedge \mathbf{B}) \leq H(\mathbf{A}) + H(\mathbf{B}), \tag{32}$$

pričom rovnosť nastáva práve vtedy, keď sú pokusy **A**, **B** *štatisticky nezávislé.*

Dôkaz.
Pretože $0 \leq I(\mathbf{A}, \mathbf{B})$ s rovnosťou práve vtedy, keď sú pokusy **A**, **B** štatisticky nezávislé, s využitím vzťahu $I(\mathbf{A}, \mathbf{B}) = H(\mathbf{A}) + H(\mathbf{B}) - H(\mathbf{A} \wedge \mathbf{B})$ máme

$$\begin{aligned} 0 &\leq I(\mathbf{A}, \mathbf{B}) = H(\mathbf{A}) + H(\mathbf{B}) - H(\mathbf{A} \wedge \mathbf{B}) \\ 0 &\leq H(\mathbf{A}) + H(\mathbf{B}) - H(\mathbf{A} \wedge \mathbf{B}) \\ H(\mathbf{A} \wedge \mathbf{B}) &\leq H(\mathbf{A}) + H(\mathbf{B}), \end{aligned}$$

kde rovnosť nastáva práve vtedy, keď sú pokusy **A**, **B** štatisticky nezávislé. □

Veta

$$H(\mathbf{A} \wedge \mathbf{B}) \le H(\mathbf{A}) + H(\mathbf{B}), \tag{32}$$

pričom rovnosť nastáva práve vtedy, keď sú pokusy **A**, **B** *štatisticky nezávislé.*

Dôkaz.
Pretože $0 \le I(\mathbf{A}, \mathbf{B})$ s rovnosťou práve vtedy, keď sú pokusy **A**, **B** štatisticky nezávislé, s využitím vzťahu $I(\mathbf{A}, \mathbf{B}) = H(\mathbf{A}) + H(\mathbf{B}) - H(\mathbf{A} \wedge \mathbf{B})$ máme

$$\begin{aligned} 0 &\le I(\mathbf{A}, \mathbf{B}) = H(\mathbf{A}) + H(\mathbf{B}) - H(\mathbf{A} \wedge \mathbf{B}) \\ 0 &\le H(\mathbf{A}) + H(\mathbf{B}) - H(\mathbf{A} \wedge \mathbf{B}) \\ H(\mathbf{A} \wedge \mathbf{B}) &\le H(\mathbf{A}) + H(\mathbf{B}), \end{aligned}$$

kde rovnosť nastáva práve vtedy, keď sú pokusy **A**, **B** štatisticky nezávislé. □

Entropia kombinovaného pokusu

Veta

$$H(\mathbf{A} \wedge \mathbf{B}) \leq H(\mathbf{A}) + H(\mathbf{B}), \tag{32}$$

pričom rovnosť nastáva práve vtedy, keď sú pokusy **A**, **B** *štatisticky nezávislé.*

Dôkaz.
Pretože $0 \leq I(\mathbf{A}, \mathbf{B})$ s rovnosťou práve vtedy, keď sú pokusy **A**, **B** štatisticky nezávislé, s využitím vzťahu $I(\mathbf{A}, \mathbf{B}) = H(\mathbf{A}) + H(\mathbf{B}) - H(\mathbf{A} \wedge \mathbf{B})$ máme

$$\begin{aligned} 0 &\leq I(\mathbf{A}, \mathbf{B}) = H(\mathbf{A}) + H(\mathbf{B}) - H(\mathbf{A} \wedge \mathbf{B}) \\ 0 &\leq H(\mathbf{A}) + H(\mathbf{B}) - H(\mathbf{A} \wedge \mathbf{B}) \\ H(\mathbf{A} \wedge \mathbf{B}) &\leq H(\mathbf{A}) + H(\mathbf{B}), \end{aligned}$$

kde rovnosť nastáva práve vtedy, keď sú pokusy **A**, **B** štatisticky nezávislé. □

Entropia kombinovaného pokusu

Veta

$$H(\mathbf{A} \wedge \mathbf{B}) \leq H(\mathbf{A}) + H(\mathbf{B}), \qquad (32)$$

pričom rovnosť nastáva práve vtedy, keď sú pokusy **A**, **B** *štatisticky nezávislé.*

Dôkaz.
Pretože $0 \leq I(\mathbf{A}, \mathbf{B})$ s rovnosťou práve vtedy, keď sú pokusy **A**, **B** štatisticky nezávislé, s využitím vzťahu $I(\mathbf{A}, \mathbf{B}) = H(\mathbf{A}) + H(\mathbf{B}) - H(\mathbf{A} \wedge \mathbf{B})$ máme

$$\begin{aligned} 0 &\leq I(\mathbf{A}, \mathbf{B}) = H(\mathbf{A}) + H(\mathbf{B}) - H(\mathbf{A} \wedge \mathbf{B}) \\ 0 &\leq H(\mathbf{A}) + H(\mathbf{B}) - H(\mathbf{A} \wedge \mathbf{B}) \\ H(\mathbf{A} \wedge \mathbf{B}) &\leq H(\mathbf{A}) + H(\mathbf{B}), \end{aligned}$$

kde rovnosť nastáva práve vtedy, keď sú pokusy **A**, **B** štatisticky nezávislé. $\square$

Zhrnutie -ŤAHÁK

$$H(\mathbf{B}|A_i) \stackrel{\text{def}}{=} H\big(P(B_1|A_i), P(B_2|A_i), \dots, P(B_m|A_i)\big) = -\sum_{j=1}^{m} P(B_j|A_i).\log_2(P(B_j|A_i)).$$

$$H(\mathbf{B}|\mathbf{A}) \stackrel{\text{def}}{=} \sum_{i=1}^{n} P(A_i).H(\mathbf{B}|A_i) \qquad H(\mathbf{B}|\mathbf{A}) = -\sum_{i=1}^{n}\sum_{j=1}^{m} P(A_i \cap B_j).\log_2\left(\frac{P(A_i \cap B_j)}{P(A_i)}\right)$$

$$I(\mathbf{A},\mathbf{B}) \stackrel{\text{def}}{=} H(\mathbf{B}) - H(\mathbf{B}|\mathbf{A}) \qquad I(\mathbf{A},\mathbf{B}) = \sum_{i=1}^{n}\sum_{j=1}^{m} P(A_i \cap B_j).\log_2\left(\frac{P(A_i \cap B_j)}{P(A_i).P(B_j)}\right)$$

$$\begin{aligned} H(\mathbf{A}\wedge\mathbf{B}) &= H(\mathbf{A}) + H(\mathbf{B}|\mathbf{A}) \\ I(\mathbf{A},\mathbf{B}) &= H(\mathbf{A}) + H(\mathbf{B}) - H(\mathbf{A}\wedge\mathbf{B}) \end{aligned}$$

V nasledukúcich troch nerovnostiach platí rovnosť práve vtedy, keď **A**, **B** sú štatisticky nezávislé pokusy:

$$\begin{aligned} 0 &\le I(\mathbf{A},\mathbf{B}) \\ H(\mathbf{B}|\mathbf{A}) &\le H(\mathbf{B}), \\ H(\mathbf{A}\wedge\mathbf{B}) &\le H(\mathbf{A}) + H(\mathbf{B}), \end{aligned}$$