

1 Úvod

1.1 Motivace

Zavedli jsme entropii. V termodynamice měla všechny atributy nepořádku, chaosu apod.: v izolovaném systému samovolně mohla jen růst nevratným procesem. Ve statistické fyzice souvisela s pravděpodobností daného makrostavu v konkurenci k makrostavům ostatním, majícím jiné počty jim odpovídajících mikrostavů. Na rozdíl od multiplikativní pravděpodobnosti (nezávislých dějů) je entropie (neinteragujících stavů) aditivní.

Chtěli bychom z definice entropie „odpreparovat“ všechno ostatní a ponechat jenom pravděpodobnost; tím by byl takový pojem použitelný všude, kde se s pravděpodobností pracuje – např. právě v informatice. Ukážeme, že pokus o kvantifikování pojmů „neurčitost“, „nejistota“ vede právě tímto směrem – k pojmu entropie, známému z termodynamiky a statistické fyziky.

1.2 Základní představy a označení

Značme (v souladu s normou IEC/ISO 80000, Díl 13)

- $\log_u$ logaritmus o základu u ,
- $\lg \equiv \log_{10}$ logaritmus o základu 10; jednotka informace hartley, Hart;
- $\ln \equiv \log_e$ logaritmus o základu e (logaritmus naturalis); jednotka informace nat, nat;
- $\text{lb} \equiv \log_2$ logaritmus o základu 2 (binární); jednotka informace shannon, Sh.

Pro jevy $\mathcal{J}$ zavedeme pravděpodobnosti p , takové, že

- jev jistý $\mathcal{I}$ má pravděpodobnost $p = 1$,
- jev vyloučený $\mathcal{O}$ má pravděpodobnost $p = 0$,
- každý jiný jev $\mathcal{J}$ má pravděpodobnost $0 < p < 1$.
- pravděpodobnost p , že nastane některý z navzájem se vylučujících jevů $\mathcal{J}_i, \mathcal{J}_j$, je rovna součtu jejich pravděpodobností: $p = p_i + p_j$.

Uvažujme úplný soubor jevů $\mathcal{S}_N^{\mathcal{E}}$, tvořený N *elementárními* jevy $\mathcal{E}_i$ s pravděpodobnostmi p_i (pro $i = 1 \dots N$), tj.

- s určitostí nastane některý z nich,
- nikdy nenastane současně více z nich.
- všechny mají stejnou pravděpodobnost p_i ; zřejmě platí $p_i = \frac{1}{N}$.

Dále uvažujme úplný soubor $\mathcal{S}_K$ tvořený K jevy $\mathcal{J}_i$ s pravděpodobnostmi p_i (pro $i = 1 \dots K$), tj.

- s určitostí nastane některý z nich,
- nikdy nenastane současně více z nich.
- jev $\mathcal{J}_i$ má pravděpodobnost p_i ; zřejmě platí $\sum_{i=1}^N p_i = 1$.

Množinu všech pravděpodobností p_i pro $i = 1, \dots, N$ budeme značit buď výpisem všech p_i , anebo $\{p_i\}_1^N$ (tedy také $\{\frac{1}{4}\}_1^4$ pro $N = 4$), případně $\mathbf{p}$.

2 Nejistota

2.1 Značení

Každý úplný soubor $\mathcal{S}_K$ tvořený K jevy $\mathcal{J}_i$ je tedy charakterizován konečnou posloupností $\{p_i\}_1^K$ tvořenou K pravděpodobnostmi $\{p_i\}_1^K$. Naším záměrem bude vyslyšet Galilea¹ a přiřadit tomuto souboru číslo, které by měřilo naši nejistotu, máme-li uhádnout, který z jevů $\mathcal{J}_i$ v jistém konkrétním případě nastane. Nejistotu označíme $H(\mathbf{p})$, kde $\mathbf{p} \equiv \{p_i\}_1^K$. Příležitostně uijeme označení typu $H_K()$ pro zdůraznění počtu jevů, případně označení typu H^A pro konkrétní hodnotu H .

Velmi častý případ nejistoty souboru tvořeného N stejně pravděpodobnými (elementárními) jevy budeme stručně značit $\mathbf{H}_N \equiv H(\{\frac{1}{N}\}_1^N)$.

♣ Písmeno S užívané pro entropii se nám pro nejistotu nehodí, protože ho užíváme pro soubory. Naše písmeno H ovšem neznamena termodynamickou entalpii. Kdo by mocí mermo chtěl mít nějakou tu mnemoniku, tak ať si představí, že H měří *horrorovitost* situace.

2.2 Konkrétní příklad: Jiřík a Zlatovláska

Jak známo, Jiřík poté, co zdolal všechny překážky, si měl vybrat svou Zlatovlásku ze dvanácti ($N = 12$) na pohled stejných panen; jeho nezáviděníhodná nejistota byla nejprve

$$\mathbf{H}_{12} = H(\mathbf{p}) = H\left(\frac{1}{12}, \frac{1}{12}, \frac{1}{12}, \frac{1}{12}, \frac{1}{12}, \frac{1}{12}, \frac{1}{12}, \frac{1}{12}, \frac{1}{12}, \frac{1}{12}, \frac{1}{12}, \frac{1}{12}\right). \quad (1)$$

Tuto nejistotu bychom v dalším rádi kvantifikovali. Díky dobrotě srdce Jiříkova v předchozím dění mu však zlatá muška sednuvší na třetí pannu zleva změnila podmínky soutěže na mnohem příznivější situaci:

$$H^{B12} = H(\mathbf{p}^{B12}) = H(0, 0, 1, 0, 0, 0, 0, 0, 0, 0, 0, 0). \quad (2)$$

Jiřík měl nyní naprostou jistotu, že zvolí Zlatovlásku a nikoli povitru. Nejistotu měl tedy nulovou: $H^{B12} = 0$. Fakticky byl ve stejné situaci, jako kdyby od Zlatovlásky dále nebyl nikdo:

$$H^{B3} = H(\mathbf{p}^{B3}) = H(0, 0, 1) = H^{B12} \quad (3)$$

anebo rovnou, kdyby tam byla jediná dívka, z níž by měl volit²

$$\mathbf{H}_1 = H(1) = 0. \quad (4)$$

¹„Co se dá změřit, má být změřeno, a co se nedá měřit, má být převedeno na měřitelné.“

²Princip voleb za totality (Národní fronta) už současní studenti znají jen z vyprávění.

Připustíme však alternativní pohádku, kdy je letní parno, plno much kolem hradu i uvnitř a znavený Jiřík si není jistý, zda velká zlatá moucha přiletěví náhle na první dívku je opožděnou pravou, velkou zlatou muškou či obyčejnou masařkou; nyní je situace

$$H^{C12} = H(\mathbf{p}^{C12}) = H\left(\frac{1}{2}, 0, \frac{1}{2}, 0, 0, 0, 0, 0, 0, 0, 0\right), \quad (5)$$

fakticky stejná se situací

$$H^{C2} = H(\mathbf{p}^{C2}) = H\left(\frac{1}{2}, \frac{1}{2}\right) = \mathbf{H}_2, \quad (6)$$

odhlédneme-li od kandidátek evidentně klamných.

Kdyby se mu přece jen velká a tlustá moucha zdála méně vhodná ke Zlatovlásce (ale koneckonců kdoví, čeho se může nadít později ...), mohlo by to být třeba

$$H^{D1,3} = H(\mathbf{p}^{D1,3}) = H\left(\frac{1}{100}, 0, \frac{99}{100}, 0, 0, 0, 0, 0, 0, 0, 0\right), \quad (7)$$

fakticky shodně se situací

$$H^{D2} = H(\mathbf{p}^{D2}) = H\left(\frac{1}{100}, \frac{99}{100}\right). \quad (8)$$

Pro úplnost připomeňme, že pořadí kandidátek v sevřené rojnici není jistě podstatné, takže záměna první a třetí dívky

$$H^{D3,1} = H(\mathbf{p}^{D3,1}) = H\left(\frac{99}{100}, 0, \frac{1}{100}, 0, 0, 0, 0, 0, 0, 0, 0\right), \quad (9)$$

je opět fakticky stejná se situací

$$H^D = H(\mathbf{p}^D) = H\left(\frac{99}{100}, \frac{1}{100}\right). \quad (10)$$

Mění ovšem Jiříkovo rozhodnutí, nikoli však jeho šance, a tedy ani jeho nejistotu.

Úhrnně můžeme kvalifikovat pořadí nejistot:

$$0 = \mathbf{H}_1 = H^{B12} = H^{B3} < H^{D1,3} = H^{D2} = H^{D3,1} = H^D < \mathbf{H}_2 = H^{C12} = H^{C2} < \mathbf{H}_{12}, \quad (11)$$

Pro kvantifikaci provedeme v dalším podrobnější rozbor. Zatím si prosím znovu pročtěte a promyslete Jiříkovy alternativy spolu s jeho kvalitativním hodnocení šancí podle rel. (11).

2.3 Požadavky na kvantitativní hodnocení

Stanovme nejprve, co očekáváme od nejistoty H jako takové:

symetrie: Hodnota funkce $H(\{p_i\})$ nezávisí na permutaci indexů:

$$H(\{\dots, p_i, \dots, p_j, \dots\}) = H(\{\dots, p_j, \dots, p_i, \dots\}) \quad \text{pro každá } i, j; \quad (12)$$

spojitost: Funkce $H(\{p_i\})$ je spojitou funkcí každé ze svých proměnných p_i ;

monotonie (vůči počtu možností): Pro nejistoty stejně pravděpodobných jevů (v souborech $\mathcal{S}_N^\varepsilon$) platí:

$$0 = H_1(1) < H_2\left(\frac{1}{2}, \frac{1}{2}\right) < H_3\left(\frac{1}{3}, \frac{1}{3}, \frac{1}{3}\right) \dots \quad (13)$$

tedy

$$(0 =) \quad \mathbf{H}_1 < \mathbf{H}_2 < \dots < \mathbf{H}_N < \mathbf{H}_{N+1} < \dots; \quad (14)$$

aditivita: Zapišeme-li posloupnost jevů jako skupiny skupin těchto jevů, pak se všechny nejistoty všech skupin sečtou (s příslušnými vahami).

Tato formulace není asi příliš srozumitelná na první čtení. Osvětlíme ji v následujícím odstavci.

2.4 Aditivita (rozděl a panuj)

Představme si místo dobrého Jiříka zhýčkaného Žoržíka, kterému žádná zlatá muška nehodlá pomoci. Veze si s sebou proto tři oddané sluhy, aby se sám nemusel namáhat a aby oni převzali aspoň kus jeho práce i odpovědnosti. Teď se pokouší, zda by něco získal rozdělením nejistot mezi své sluhy. On sám má nejistotu $H^{\tilde{Z}}$; podle rov. (1) je však $H^{\tilde{Z}} = \mathbf{H}_{12}$.

$$\underbrace{\left(\frac{1}{12}, \frac{1}{12}, \frac{1}{12}, \frac{1}{12}, \frac{1}{12}, \frac{1}{12}, \frac{1}{12}, \frac{1}{12}, \frac{1}{12}, \frac{1}{12}, \frac{1}{12}, \frac{1}{12} \right)}_{\tilde{Z}} \quad (15)$$

Zkusí rozdělit odpovědnost tak, že Alfons navrhne mezi prvními pěti pannami, Bolton mezi dalšími čtyřmi a Cyril mezi zbývajících třemi pannami.

$$\underbrace{\left(\frac{1}{12}, \frac{1}{12}, \frac{1}{12}, \frac{1}{12}, \frac{1}{12} \right)}_A, \quad \underbrace{\left(\frac{1}{12}, \frac{1}{12}, \frac{1}{12}, \frac{1}{12} \right)}_B, \quad \underbrace{\left(\frac{1}{12}, \frac{1}{12}, \frac{1}{12} \right)}_C \quad (16)$$

Alfons má šanci (může mít šťastnou ruku) v $\frac{5}{12}$ případů a má nejistotu $H^A = \mathbf{H}_5$, Bolton má šanci $\frac{4}{12}$ s nejistotou $H^B = \mathbf{H}_4$ a Cyril má nejmenší šanci $\frac{3}{12}$, ale s nejistotou jen $H^C = \mathbf{H}_3$. Žoržík se pak rozhodne mezi třemi návrhy svých sluhů, tedy mezi šancemi $\left(\frac{5}{12}, \frac{4}{12}, \frac{3}{12} \right)$.

$$\underbrace{\underbrace{\left(\frac{1}{12}, \frac{1}{12}, \frac{1}{12}, \frac{1}{12}, \frac{1}{12} \right)}_A, \quad \underbrace{\left(\frac{1}{12}, \frac{1}{12}, \frac{1}{12}, \frac{1}{12} \right)}_B, \quad \underbrace{\left(\frac{1}{12}, \frac{1}{12}, \frac{1}{12} \right)}_C}_{\tilde{Z}^*} \quad (17)$$

Uváží-li, jakou měl který sluha šanci uhádnout Zlatovlásku, přináší Žoržík sám nejistotu

$$H^{\tilde{Z}^*} = H_3\left(\frac{5}{12}, \frac{4}{12}, \frac{3}{12}\right) \quad (18)$$

Z nejistot každého z jeho sluhů se ovšem uplatní jen příslušná poměrná část. Celková výsledná nejistota koumáckého Žoržíka je tedy

$$\begin{aligned} H^{\tilde{Z}^*} &+ \frac{5}{12} H^A + \frac{4}{12} H^B + \frac{3}{12} H^C = \\ H_3\left(\frac{5}{12}, \frac{4}{12}, \frac{3}{12}\right) &+ \frac{5}{12} \mathbf{H}_5 + \frac{4}{12} \mathbf{H}_4 + \frac{3}{12} \mathbf{H}_3 \end{aligned} \quad (19)$$

My ovšem víme, že spravedlnost vítězí³, čili Žoržík si tím nemůže ani polepšit, ani pohoršit. Musí tedy platit

$$(H^{\tilde{Z}^*} =) \quad \mathbf{H}_{12} = H_3\left(\frac{5}{12}, \frac{4}{12}, \frac{3}{12}\right) + \frac{5}{12} \mathbf{H}_5 + \frac{4}{12} \mathbf{H}_4 + \frac{3}{12} \mathbf{H}_3 \quad (20)$$

a to je náš předpoklad spravedlivého skládání nejistot v tomto konkrétním případě.

Každý matfyzák to ovšem ihned přepíše obecně: Pro úplný soubor $\mathcal{S}_K$ tvořený K jevy $\mathcal{J}_i$ s pravděpodobnostmi p_i (pro $i = 1 \dots K$) převeditelnými na společný jmenovatel tak, že $p_i = \frac{n_i}{N}$, kde n_i , N jsou přirozená čísla, platí

$$\mathbf{H}_N = H_K\left(\{p_k\}_1^K\right) + \sum_{k=1}^K p_k \mathbf{H}_{n_k}, \quad (21)$$

³(ve všech pohádkách i ve školních učebnicích, včetně teoretické fyziky)

resp. po opuštění našich zkratk

$$H_N\left(\left\{\frac{1}{N}\right\}_{i=1}^N\right) = H_K\left(\left\{\frac{n_k}{N}\right\}_{k=1}^K\right) + \sum_{k=1}^K \frac{n_k}{N} H_{n_k}\left(\left\{\frac{1}{n_k}\right\}_{j=1}^{n_k}\right), \quad (22)$$

O situace, kdy pravděpodobnosti p_i jsou iracionální a nelze je tedy převést na společný jmenovatel, se pak postará náš požadavek spojitosti.

3 Řešení problému

3.1 Úloha

Požadavky jsou formulovány; stačí tedy „jen“ je splnit, tj. najít obecný vzorec pro

$$H_K\left(\{p_k\}\right) \equiv H_K\left(\left\{\frac{n_k}{N}\right\}_{k=1}^K\right), \quad (23)$$

pro libovolnou posloupnost racionálních p_k (se společným jmenovatelem N), a případně najít řešení i pro iracionální p_i .

3.2 Zjednodušení

Ukážeme, že namísto vyjádření (23) stačí vyjádřit všechna $\mathbf{H}_N$ pro $N = 1, 2, \dots$, tedy nejistoty pro soubory $\mathcal{S}_N^\mathcal{E}$ elementárních jevů pro všechna přirozená N . Hledaná (23) totiž při znalosti $\mathbf{H}_N$ dostaneme pro libovolnou konečnou posloupnost $\{p_k\}$ s racionálními p_k z rov. (21). To je ovšem znamenité zjednodušení.

3.3 Speciální případ: $N = 2^n$

Případ $N = 1$ je jasný: $\mathbf{H}_1 = H_1(1) = 0$, žádná nejistota není.

Ani pro případ $N = 2$ nám rov. (21) nepřinese nic nového: rozdělíme-li 2 jevy do 2 skupin po 1 jevu, dostaneme

$$\mathbf{H}_2 = H_2\left(\frac{1}{2}, \frac{1}{2}\right) + \frac{1}{2}\mathbf{H}_1 + \frac{1}{2}\mathbf{H}_1 = \mathbf{H}_2 + 0 + 0 = \mathbf{H}_2, \quad (24)$$

tedy identitu. (Aspoň máme potvrzeno, že $\mathbf{H}_1 = 0$.)

Až další případ, $N = 4$, dává něco nového: čtyři elementární jevy rozdělíme na dvě skupiny po dvou a dostaneme

$$\mathbf{H}_4 \equiv H_4\left(\frac{1}{4}, \frac{1}{4}, \frac{1}{4}, \frac{1}{4}\right) = H_2\left(\frac{1}{2}, \frac{1}{2}\right) + \frac{1}{2}\mathbf{H}_2 + \frac{1}{2}\mathbf{H}_2 = 2\mathbf{H}_2. \quad (25)$$

Podobně postupujeme dále:

$$\mathbf{H}_8 \equiv H_8\left(\left\{\frac{1}{8}, \dots\right\}\right) = H_2\left(\frac{1}{2}, \frac{1}{2}\right) + \frac{1}{2}\mathbf{H}_4 + \frac{1}{2}\mathbf{H}_4 = 3\mathbf{H}_2, \quad (26)$$

$$\mathbf{H}_{16} \equiv H_{16}\left(\left\{\frac{1}{16}, \dots\right\}\right) = H_2\left(\frac{1}{2}, \frac{1}{2}\right) + \frac{1}{2}\mathbf{H}_8 + \frac{1}{2}\mathbf{H}_8 = 4\mathbf{H}_2, \quad (27)$$

a snadno dokážeme úplnou indukcí, že

$$\mathbf{H}_{2^n} = n\mathbf{H}_2, \quad (28)$$

neboli

$$\mathbf{H}_N = \text{lb}(N) \mathbf{H}_2, \quad (29)$$

alespoň pokud je $\text{lb } N = n$ přirozené číslo. Logaritmická závislost ve speciálním případě budí podezření, že vzorec jde zobecnit. Zkusíme proto dostat obecnou funkcionální rovnici.

3.4 Další speciální případ: $N = a \cdot b$

V předchozím odstavci jsme zvětšovali počet jevů násobením dvěma. Můžeme ovšem násobit i jiným číslem, třeba třemi.

Jestliže je $N = 3b$, kde b je přirozené číslo, pak můžeme N jevů rozdělit na 3 disjunktní skupiny po b elementárních jevech a rov. (21) dává

$$\mathbf{H}_{3b} = \mathbf{H}_3 + 3 \cdot \frac{1}{3} \mathbf{H}_b = \mathbf{H}_3 + \mathbf{H}_b \quad (30)$$

Nedá práci nahlédnout, že pro libovolná přirozená čísla platí

$$\mathbf{H}_{ab} = \mathbf{H}_a + \mathbf{H}_b \quad (31)$$

Stačí tedy znát jen $\mathbf{H}_p$ pro všechna prvočísla p .

3.5 Obecné řešení pro $\mathbf{H}_N$ a $H(\{p_i\})$

Funkcionální rov. (31) má však známé řešení:

$$\mathbf{H}_a = \log a \quad \text{při libovolném základu logaritmu.} \quad (32)$$

V informatice obvykle pracujeme s binární soustavou, a proto bereme logaritmy se základem 2, tedy $\text{lb}(x) = \log_2(x)$.

Připomeňme, že všechny logaritmy si jsou úměrné, takže platí např. $\text{lb } 27 = \lg 27 / \lg 2$. Volba základu logaritmu je tedy pouze otázkou volby konstanty — měrné jednotky (viz norma IEC/ISO 80000, Díl 13). U binárních logaritmů je $\mathbf{H}_2 = 1$; jednotkou je proto shannon, Sh. Volba přirozených logaritmů $\ln$ se základem e dává jednotku nat, nat. Volba dekadických logaritmů $\lg$ dává jednotku hartley, Hart. Přibližné převodní vztahy jsou

- $1 \text{ Sh} \approx 0,693 \text{ nat} \approx 0,301 \text{ Hart}$
- $1 \text{ nat} \approx 1,433 \text{ Sh} \approx 0,434 \text{ Hart}$
- $1 \text{ Hart} \approx 3,322 \text{ Sh} \approx 2,303 \text{ nat}$

Dosazením do rov. (21) dostaneme

$$\begin{aligned} H_K(\{p_k\}_1^K) &= -\mathbf{H}_N - \sum_{k=1}^K p_k \mathbf{H}_{n_k} \quad \text{pro } p_k = n_k/N; \text{ tedy} \\ &= -\text{lb}(N) - \sum_{k=1}^K \frac{n_k}{N} \text{lb}(n_k) \\ &= -\sum_{k=1}^K \frac{n_k}{N} \text{lb}\left(\frac{n_k}{N}\right) \\ H_K(\{p_k\}_1^K) &= -\sum_{k=1}^K p_k \text{lb}(p_k) \end{aligned} \quad (33)$$

Tento vzorec jsme odvodili jen pro racionální pravděpodobnosti p_k , ale díky požadavku spojitosti H musí platit i pro p_k zcela obecná.