INFORMAČNÍ BULLETIN

České statistické společnosti

Ročník 35, číslo 4, prosinec 2024

ZOBECNĚNÁ ENTROPIE (RÉNYIHO α-ENTROPIE A DALŠÍ)

ENTROPY (RÉNYI'S AND OTHERS)

Zdeněk Půlpán

Adresa: Univerzita Pardubice, Dopravní fakulta Jana Pernera, Studentská 95, 532 10 Pardubice 2; Na Brně 1952, 500 09 Hradec Králové 9

E-mail: `zdenek.pulpan@upce.cz, zdenek.pulpan@post.cz`

Abstrakt: Entropie je jedna z charakteristik náhodné veličiny. V práci jsou ukázána některá zobecnění původní Shannonovy entropie jako je Rényiho nebo Tsallisova entropie. Naznačeny jsou další možnosti zobecnění.

Klíčová slova: entropie, Shannonova entropie, Rényiho entropie, Tsallisova entropie, lineární entropie.

Abstract: Entropy is one of the characteristics of a random variable. Some generalized original Shannon entropies such as Rényi's or Tsalli's entropy are shown in the work. Other possibilities of generalization are indicated too.

Keywords: entropy, Shannon entropy, Rényi entropy, Tsallis entropy, linear entropy.

1. Úvod

Když v roce 1928 definoval Hartley svůj vzorec pro určení kvantity neurčitosti v popisu přenosu zpráv, možná netušil, jak významná byla jeho myšlenka použití logaritmu na počet všech očekávaných možných situací. Shannonovou zásluhou o 20 let později pak bylo zavedení míry neurčitosti (později nazvané entropie) a velikosti informace (informace jako rozdíl entropie před a po přijetí zprávy) rozšířením Hartleyovy myšlenky na rozdělení pravděpodobnosti navzájem se vylučujících očekávaných variant zprávy. Studium neurčitosti při přenosu zpráv vedlo k objevu řady pozoruhodných vztahů mezi naším chápáním neurčitosti a jejím zachycením pomocí pravděpodobnosti a pak entropie. Tak se vytvořila nová charakteristika, odvozená z pravděpodobnostního rozdělení (na množině všech variant zprávy) a pojem entropie se stal součástí teorie pravděpodobnosti.

Známý maďarský matematik Alfréd Rényi (1921–1970) stačil za svůj krátký život Shannonovu myšlenku dále tak zobecnit, že předchozí možnosti odhadu neurčitosti z entropie se staly za určitých podmínek pouze speciálními případy Rényiho zobecnění Shannonovy entropie. Rényiho parametrizované zobecnění má (vhodnou volbou hodnoty parametru) větší naději pro uplatnění v různých experimentálních situacích (tedy již ne jenom při přenosu a kódování informace) než poněkud užší pojetí Shannonovo. Základním krokem aplikace v konkrétní experimentální situaci je volba určité náhodné veličiny (mající vztah k experimentu) a její, pro interpretaci nejvhodnější, míry neurčitosti. Měr neurčitosti se ve dvacátém století objevilo velké množství, my se zde zmíníme o těch z našeho hlediska nejzajímavějších [1, 2, 4–8, 12–15].

2. Rényiho zobecněná entropie

Uvažujme pravděpodobnostní schéma s diskrétní náhodnou veličinou X a její pravděpodobnostní funkci

$$P_n = \begin{pmatrix} x_1, \dots, x_n \\ p_1, \dots, p_n \end{pmatrix}, \quad \sum_1^n p_i = 1, \quad 0 \le p_i \le n,$$

v pravděpodobnostním prostoru Ω_n, A_n, P_n, kde $\Omega_n = \{x_1, x_2, \dots, x_n\}$, $n \ge 2$, je základní prostor, A_n algebra všech jeho podmnožin a P_n pravděpodobnost na A_n.

Definice 1. Zobecněnou Rényiho entropií řádu α, kde $\alpha \ge 0,\ \alpha \ne 1$, pro diskrétní náhodnou veličinu X s pravděpodobnostní funkcí $P_n(X = x_i) = p_i,\ i = 1, 2, \dots, n$, nazýváme výraz

$$H_\alpha(X) = \frac{1}{1-\alpha} \log_z \left(\sum_1^n p_i^\alpha \right), \text{ o základu logaritmu } z \ge 2, \tag{1}$$

který je nejčastěji volený $z = 2$, resp. $z = 10$, resp. $z = e$. Podle toho se odlišují jednotky, jak ukážeme.

Dále budeme diskrétní náhodnou veličinu X (dále jen n. v. X) označovat také pomocí jejího pravděpodobnostního rozdělení takto:

$$X = (p_1, p_2, \dots, p_n).$$

Pro každé přípustné α představuje $H_\alpha(X)$ určitý „metr“ měření entropie (pro odhad míry neurčitosti) diskrétní n. v. X. Entropie souvisí s informací n. v. X tak, že informace je rozdíl veličiny $H_\alpha(X)$ s rozdělením před sledovanou událostí a po ní. Základ logaritmu je v podstatě libovolný (určuje ale také „metr“, souvisí tedy s interpretací veličiny $H_\alpha(X)$), často se užívá dvojkový logaritmus ($z = 2$) protože má nejjednodušší interpretaci. Jednotku entropie při $z = 2$ bychom měli označovat α-*bit*.

Poznámka:
Místo $H(X) = H((p_1, p_2, \ldots, p_n))$ píšeme dále jen $H(p_1, p_2, \ldots, p_n)$.

Speciální případy:

1) Je-li $P_n(X = x_i) = \frac{1}{n}$, $i = 1, 2, \ldots, n$, pak

$$H_\alpha(X) = \frac{1}{1-\alpha} \log_z(n \cdot (n^{-\alpha})) = \log_z n = H_0(X)$$

nezávisle na volbě α. Výsledná Rényiho entropie je zde Hartleyovou mírou. Volíme-li ve výrazu pro $H_\alpha(X)$ parametr $\alpha = 0$, dostaneme tutéž hodnotu $\log_z n$. Proto označujeme v tomto případě Rényiho entropii $H_0(X)$.

2) V definici Rényiho zobecněné entropie je vyloučena hodnota $\alpha = 1$. Limitním přechodem pro $\alpha \to 1$ však dostaneme Shannonovu entropii:

$$\begin{aligned}\lim_{\alpha\to 1} H_\alpha(X) &= \lim_{\alpha\to 1} \frac{\log_2 \sum_1^n p_i^\alpha}{1-\alpha} = \lim_{\alpha\to 1} \frac{1}{\ln 2} \frac{\ln \sum_1^n p_i^\alpha}{1-\alpha} = \\ &= \lim_{\alpha\to 1} \frac{-1}{\ln 2} \cdot \frac{1}{\sum_1^n p_i^\alpha} \cdot \sum_1^n p_i^\alpha \ln p_i = -\sum_1^n p_i \log_2 p_i.\end{aligned}$$

Definujeme zde $0 \cdot \log 0$ limitou $\lim_{p\to 0^+} p \cdot \log p$. Hodnotu $H_\alpha(X)$ lze tedy dodefinovat i pro $\alpha = 1$ Shannonovou entropií. Proto v tomto případě se značí

$$H_1(X) = -\sum_1^n p_i \log_2 p_i \quad [bit].$$

Pro základ logaritmu 10, resp. e se jednotka nazývá *dit*, resp. *nit*.

3) Volíme-li $\alpha = 2$, dostaneme tzv. kolizní entropii ve tvaru

$$H_2(X) = -\log_2 \sum_1^n p_i^2 \quad [2\text{-}bit].$$

4) Přechodem k limitě $\alpha \to \infty$, dostaneme Rényiho entropii postupně ve tvaru

$$\begin{aligned}H_\infty(X) &= \lim_{\alpha\to\infty} H_\alpha(X) = \min_i(-\log_2 p_i) = \\ &= -\max_i \log_2 p_i = \log_2 \max_i p_i \quad [\infty\text{-}bit].\end{aligned}$$

Z důvodů, které dále ukážeme, se tato entropie nazývá minimální.

Ještě se někdy uvažuje o limitě $\alpha \to -\infty$ (vzhledem k limitě předchozí) a pak se definuje

$$H_{-\infty}(X) = \lim_{\alpha\to-\infty} H_\alpha(X) = -\log_2 \min_i p_i \quad [-\infty\text{-}bit].$$

Tato entropie se nazývá maximální.

Nyní se podíváme na závislost hodnot entropie $H_\alpha(X)$ na α. Platí následující věta.

Věta 1. Rényiho entropie $H_\alpha(X)$ je nerostoucí funkcí v proměnné α, $\alpha > 0,\ \alpha \neq 1$.

Důkaz: Spočteme derivaci $H_\alpha(X)$ podle α:

$$\begin{aligned}\frac{\mathrm{d}H_\alpha(X)}{\mathrm{d}\alpha} &= \frac{1}{(1-\alpha)^2}\cdot\log_2\sum_1^n p_i^\alpha + \frac{1}{1-\alpha}\sum_1^n \frac{p_i^\alpha}{\sum_j p_j^\alpha}\log_2 p_i = \\ &= \frac{1}{(1-\alpha)^2}\left\{\log_2\sum_1^n p_i^\alpha + \sum_1^n p_i^\alpha \frac{1}{\sum_j p_j^\alpha}\log_2 p_i - \right.\\ &\quad \left. - \sum_1^n p_i^\alpha \frac{1}{\sum_j p_j^\alpha}\log_2 p_i^\alpha\right\} = \frac{1}{(1-\alpha)^2}\left\{1\cdot\log_2\sum_1^n p_i^\alpha + \right.\\ &\quad \left. + \sum_1^n \frac{p_i^\alpha}{\sum_j p_j^\alpha}\cdot\log_2\frac{p_i}{p_i^\alpha}\right\}.\end{aligned}$$

Nyní použijeme vztah

$$1 = \frac{\sum_1^n p_i^\alpha}{\sum_1^n p_j^\alpha} = \sum_1^n p_i^\alpha \frac{1}{\sum_{j=1}^n p_j^\alpha}.$$

A dosadíme do posledního výrazu pro derivaci a dostaneme:

$$\begin{aligned}\frac{\mathrm{d}H_\alpha(X)}{\mathrm{d}\alpha} &= \frac{1}{(1-\alpha)^2}\left\{\sum_i \frac{p_i^\alpha}{\sum_j p_j^\alpha}\log_2\sum_i p_i^\alpha + \sum_i \frac{p_i^\alpha}{\sum_j p_j^\alpha}\log_2\frac{p_i}{p_i^\alpha}\right\} = \\ &= \frac{1}{(1-\alpha)^2}\left\{\sum_i \frac{p_i^\alpha}{\sum_j p_j^\alpha}\log_2\frac{p_i}{p_i^\alpha}\sum_i p_i^\alpha\right\} = \\ &= -\frac{1}{(1-\alpha)^2}\cdot\sum_i \frac{p_i^\alpha}{\sum_j p_j^\alpha}\cdot\log_2\frac{p_i^\alpha}{p_i\sum_j p_j^\alpha} = \frac{-1}{(1-\alpha)^2}\sum_i z_i\log_2\frac{z_i}{p_i},\end{aligned}$$

kde jsme položili $z_i = \frac{p_i^\alpha}{\sum_j p_j^\alpha}$. Derivace $\frac{\mathrm{d}H_\alpha(X)}{\mathrm{d}\alpha}$ je záporná, protože představuje $\frac{-1}{(1-\alpha)^2}$ násobek Kullback-Leiblerovy divergence, která je vždy nezáporná [4].

Poznámka: Po dodefinování H_1 i pro $\alpha = 1$, je $H_\alpha(X)$ nerostoucí spojitou funkcí α.

Důsledek: Pro jednotlivé entropie $H_0(X),\ H_1(X),\ H_2(X),\ H_\infty(X)$ platí nerovnosti

$$H_0(X) \geq H_1(X) \geq H_2(X) \geq H_\infty(X).$$

Poznámka: Uvedené nerovnosti vyplývají také z následujících nerovností:

$$H_1(X) \geq H_2(X) \text{ protože platí } \sum_i p_i \log_2 p_i \leq \log_2 \sum_i p_i^2,$$
$$H_\infty(X) \leq H_2(X) \text{ protože platí } \log_2 \sum_i p_i^2 \leq \log_2 \sup_i p_i.$$

Věta 2. Pro entropie $H_2(X)$ a $H_\infty(X)$ platí nerovnost

$$H_2(X) \leq 2 \cdot H_\infty(X).$$

Důkaz: Podle definice je $H_2\,(X) = \frac{1}{1-2} \log_2 \sum_i p_i^2 = -\log_2 \sum_i p_i^2$. Protože $\log_2 \sum_i p_i^2 \geq \log_2 \max_i p_i^2 = 2 \log_2 \max_i p_i$, je $H_2\,(X) \leq -2 \log_2 \max_i p_i = 2 \cdot H_\infty(X)$.

Příklad 1a. Mějme n. v. X s rozdělením $P_2\,(X = x_1) = p$, $P_2\,(X = x_2) = 1 - p = q$, $0 \leq p,\ q \leq 1$. Pak je

$$\begin{aligned} H_0\,(X) &= \log_2 2 = 1 \\ H_1\,(X) &= -p \log_2 p - (1-p) \log_2 (1-p) \\ H_2\,(X) &= -\log_2 (p^2 + q^2) \\ H_\infty\,(X) &= -\log_2 \max(p, q). \end{aligned}$$

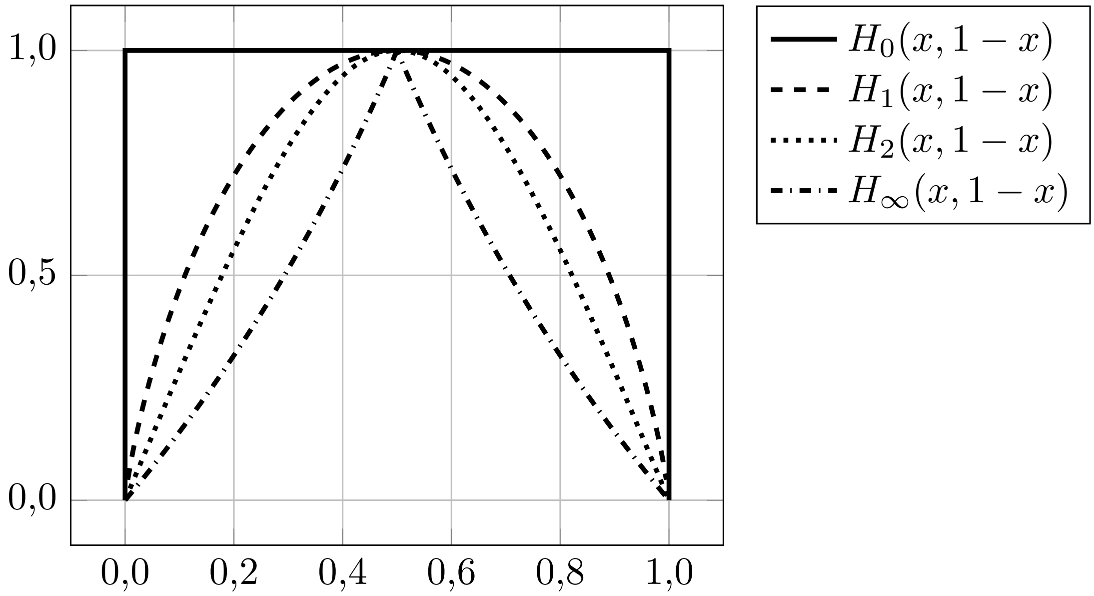
Na Obr. 1 jsou zakresleny grafy funkcí $H_0\,(X)$, $H_1\,(X)$, $H_2\,(X)$ a $H_\infty(X)$, kde $X = (p, 1-p)$, $0 \leq p \leq 1$.

O dalších vlastnostech entropie $H_\alpha(X)$ se lze dozvědět z [9] nebo také z [18].

Sledujeme-li Obr. 1, na první pohled se zdá, že by mohlo existovat pro n. v. $X = (p, 1-p)$ reálné $\alpha_0 > 2$ tak, že

$$H_{\alpha_0}\,(X) = 2 \cdot \min(p, 1-p)$$

by byla po částech lineární. Platí ale následující Věta 3.



Obrázek 1: Grafy hodnot $H_i\,(p, 1-p)$ pro $i = 0, 1, 2, \infty$
v závislosti na $p \in \langle 0; 1\rangle$.

Věta 3. Pro žádné reálné $\alpha > 2$ nemůže být

$$H_\alpha\,(p, 1-p) = \frac{1}{1-\alpha}\log_2(p^\alpha) + (1-p^\alpha) = 2\cdot\min\,(p, 1-p)\,, \quad 0 \le p \le 1.$$

Důkaz: Předpokládejme, že takové α existuje pro všechna $p \in (0, 1)$. Pak musí pro tato p platit

$$\frac{1}{1-\alpha}\log_2(p^\alpha + (1-p^\alpha)) = 2\cdot\min\,(p, 1-p) = 2p,$$

protože vztah je symetrický k p a $1-p$.

Ekvivalentními úpravami tohoto výrazu dostaneme mezi p a α vztah

$$2^{2p(1-\alpha)} = p^\alpha + (1-p^\alpha),$$

který má být platný pro jediné α a všech možných volbách $p \in (0; 0{,}5)$. Pro $p = 0; 0{,}5; 1$ vztah zřejmě platí pro jakékoliv α. Volme proto nejprve $p = 0{,}25$ a pak $p = 0{,}125$ a dosaďme do předchozí rovnice.

Pro $p = 0{,}25$ dostaneme podmínku pro α ve tvaru

$$2^{1{,}5\alpha+0{,}5} = 1 + 3^\alpha$$

a pro $p = 0{,}125$ pak

$$2^{2{,}75\alpha+0{,}25} = 1 + 7^\alpha.$$

Podle předpokladu oba vztahy pro určení α musí mít stejné řešení pro $\alpha > 2$. Ukážeme, že nemohou mít stejná řešení.

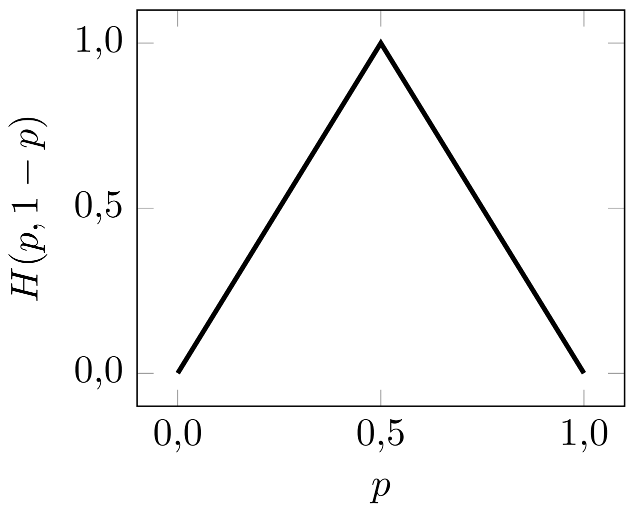
Snadno zjistíme numerickými prostředky, že první rovnice má jediný kořen pro $\alpha > 2$ v intervalu $(5; 6)$ a druhá v intervalu $(3; 4)$. Společný kořen $\alpha > 2$ tedy neexistuje. Proto také pro žádné reálné $\alpha > 2$ nemůže nastat rovnost z Věty 3.

3. Další entropie

Připomeňme na začátek ještě jedno velmi obecné zavedení entropie tzv. Fadějevovými axiomy z roku 1956. Fadějev [17] určuje entropii $H(p_1, p_2, \ldots, p_n)$, $\sum_i p_i = 1$, $0 \leq p_i \leq 1$ pro všechna i až na multiplikativní konstantu jednoznačně následujícími podmínkami:

1. $H(p, 1-p)$ je spojitá funkce pro $p \in (0; 1)$ a kladná v aspoň jednom bodě;
2. $H(p_1, p_2, \ldots, p_n)$ je symetrická funkce svých proměnných;
3. $H(p_1, p_2, \ldots, p_{n-1}, q_1, q_2) = H(p_1, p_2, \ldots, p_n) + p_n \cdot H\left(\frac{q_1}{p_n}, \frac{q_2}{p_n}\right)$, kde je $p_n = q_1 + q_2 > 0$. [12]

Nyní se vrátíme k Obr. 1 a uvědomíme si, že žádná Rényiho entropie $H_\alpha(X)$ nemůže být pro dichotomický rozhodovací proces po částech lineární ve tvaru, viz také Obr. 2.



Obrázek 2: Graf funkce $H(p, 1-p) = 2 \cdot \min(p, 1-p),\ 0 \leq p \leq 1$.

Pokusme se ale předchozí vztah rozšířit pomocí Fadějevových axiomů na novou entropii pro n. v. X.

Začneme podmínkou 3. Podle ní musí platit pro $n = 3$

$$\begin{aligned} H\left(p_1, p_2, p_3\right) &= H\left(p_1, p_2 + p_3\right) + \left(p_2 + p_3\right) \cdot H\left(\frac{p_2}{p_2 + p_3}, \frac{p_3}{p_2 + p_3}\right) = \\ &= 2 \cdot \min(p_1, p_2 + p_3) + (p_2 + p_3) \cdot 2 \cdot \min\left(\frac{p_2}{p_2 + p_3}, \frac{p_3}{p_2 + p_3}\right) = \\ &= 2 \cdot \min(p_1, p_2 + p_3) + 2 \cdot \min(p_2, p_3). \end{aligned}$$

A podobně pak bude

$$\begin{aligned} H\left(p_1, p_2, p_3, p_4\right) &= H\left(p_1, p_2, p_3 + p_4\right) + \left(p_3 + p_4\right) H\left(\frac{p_3}{p_3 + p_4}, \frac{p_4}{p_3 + p_4}\right) = \\ &= 2 \cdot \min\left(p_1, p_2 + p_3 + p_4\right) + 2 \cdot \min\left(p_2, p_3 + p_4\right) + \\ &\quad + 2 \cdot \min(p_3, p_4), \text{ atd.} \end{aligned}$$

Uvedený postup odpovídá rozhodování podle schématu na Obr. 3.

$$\begin{pmatrix} x_1 & x_2 & x_3 \\ p_1 & p_2 & p_3 \end{pmatrix} \longrightarrow \begin{pmatrix} x_1 & \{x_2, x_3\} \\ p_1 & 1 - p_1 \end{pmatrix} \longrightarrow \begin{pmatrix} x_2 & \{x_3\} \\ \frac{p_2}{p_2+p_3} & \frac{p_3}{p_2+p_3} \end{pmatrix}$$

Obrázek 3: Schéma rozhodování ve dvou krocích,
týkající se n. v. $X = (p_1, p_2, p_3)$.

Takto definovaná funkce $H\left(p_1, p_2, p_3\right)$ ale není symetrickou funkcí svých pravděpodobností:

$$\begin{aligned} H\,(0{,}5; 0{,}25; 0{,}25) &= 2 \cdot \min\,(0{,}5; 0{,}5) + 2 \cdot \min\,(0{,}25; 0{,}25) = 1 + 0{,}5 = 1{,}5; \\ H\,(0{,}25; 0{,}5; 0{,}25) &= 2 \cdot \min\,(0{,}25; 0{,}75) + 2 \cdot \min\,(0{,}5; 0{,}25) = 0{,}5 + 0{,}5 = 1. \end{aligned}$$

Poznámka: Někdy je důležité respektovat uspořádání pravděpodobností $p_1, p_2, \ldots, p_n$ například s hlediska logiky příslušného rozhodovacího procesu. V některých aplikacích by mohlo záležet na tom, v jaké posloupnosti se jednotlivé varianty v rozhodovacím procesu identifikují nebo která z variant je pro interpretaci variantou preferovanou. [10, 11]

Ukážeme, že lze předchozí vztah pro $H(p_1, p_2, p_3)$ upravit tak, aby jeho hodnota nezávisela na permutaci pravděpodobností (p_1, p_2, p_3): $H(p_1, p_2, p_3) = H(p_{i_1}, p_{i_2}, p_{i_3})$, kde $(p_{i_1}, p_{i_2}, p_{i_3})$ je taková permutace pravděpodobností, že

$$p_{i_1} \geq p_{i_2} \geq p_{i_3}.$$

Zavedeme proto tuto obecnou definici entropie $H(p_1, p_2, \ldots, p_n)$:

Definice 2. Pro diskrétní n. v. $X = (p_1, p_2, \ldots, p_n)$, $\sum_i p_i = 1$, $0 \le p_i \le 1$ pro všechna i, definujeme entropii $H(p_1, p_2, \ldots, p_n)$ takto:

a) $H(p_1, p_2) = 2 \cdot \min(p_1, p_2)$,

b) pro $n \geq 3$ platí

$$H(p_1, p_2, \ldots, p_n) = H\left(p_{i_1}, p_{i_2}, \ldots, p_{i_{n-1}} + p_{i_n}\right) + \\ + \left(p_{i_{n-1}} + p_{i_n}\right) \cdot H\left(\frac{p_{i_{n-1}}}{p_{i_{n-1}} + p_{i_n}}, \frac{p_{i_n}}{p_{i_{n-1}} + p_{i_n}}\right), \tag{2}$$

kde $p_{i_1} \geq p_{i_2} \geq \cdots \geq p_{i_{n-1}} \geq p_{i_n} \geq 0$ a $p_{i_{n-1}} + p_{i_n} > 0$.

Použijeme-li tuto definici, dostaneme pro $H(p_1, p_2, \ldots, p_n)$ vztah (2) ve tvaru

$$\begin{aligned} H(p_1, p_2, \ldots, p_n) &= 2 \cdot \min\left(p_{i_1}, p_{i_2}, \ldots, p_{i_{n-1}} + p_{i_n}\right) + \\ &\quad + \left(p_{i_{n-1}} + p_{i_n}\right) \cdot 2 \cdot \min\left(\frac{p_{i_{n-1}}}{p_{i_{n-1}} + p_{i_n}}, \frac{p_{i_n}}{p_{i_{n-1}} + p_{i_n}}\right) = \\ &= 2 \cdot \min\left(p_{i_1}, p_{i_2} + \cdots + p_{i_{n-1}} + p_{i_n}\right) + 2 \cdot \min\left(p_{i_2}, p_{i_3} + \right. \\ &\quad \left. + \cdots + p_{i_{n-1}} + p_{i_n}\right) + \cdots + 2 \cdot \min(p_{i_{n-1}}, p_{i_n}), \end{aligned}$$

kde již můžeme vypustit podmínku $p_{i_{n-1}} + p_{i_n} > 0$.

Definice 2 vyhovuje všem Fadějevovým podmínkám, ale není Shannonovou entropií:

1. Funkce $H(p_1, p_2) = 2 \cdot \min(p_1, p_2)$, $p_1 + p_2 = 1$, je spojitou funkcí svých proměnných a kladná např. pro $p_1 = p_2 = 0{,}5$;
2. $H(p_1, p_2, \ldots, p_n)$ je symetrickou funkcí svých proměnných;
3. určení $H(p_1, p_2, \ldots, p_n)$ je jednoznačné.

Entropie $H(p_1, p_2, \ldots, p_n)$ má následující vlastnosti:

a) $H(0{,}5; 0{,}5) = 2 \cdot \min(0{,}5; 0{,}5) = 1$. Proto jsme kdysi nazvali tuto jednotku *lit* [3].

b) $H(1, 0, 0, \ldots, 0) = 0$ [*lit*],

c) $H\left(\frac{1}{n}, \frac{1}{n}, \ldots, \frac{1}{n}\right) = 2 \cdot \frac{n-1}{n} = 2 \cdot \left(1 - \frac{1}{n}\right) < 2$ [*lit*],

d) $0 \leq H(p_1, p_2, \ldots, p_n) < 2$ [*lit*],

e) $H(p_1, p_2, 0, 0, \ldots, 0) = H(p_1, p_2)$, $p_1 + p_2 = 1$, $0 \le p_1, p_2 \le 1$.

Příklad 1b. Nechť n. v. X $(p = 0{,}25; 1 - p = 0{,}75)$ pak máme

$$\begin{aligned} H(p, 1-p) &= H(1-p, p) = 2 \cdot \min(p, 1-p) = 0{,}5 \; [\mathit{lit}] \\ H_0(X) &= \log_2 2 = 1 \; [\mathit{bit}] \\ H_1(X) &= 0{,}811 \; [\mathit{bit}] \\ H_2(X) &= -\log_2\left(0{,}25^2 + 0{,}75^2\right) = 0{,}67807 \; [2\text{-}\mathit{bit}] \\ H_\infty(X) &= -\log_2 \max(0{,}25; 0{,}75) = 0{,}41503 \; [\infty\text{-}\mathit{bit}]. \end{aligned}$$

Nyní se podíváme na další z entropií, tzv. Tsallisovu entropii. Je zajímavá také tím, že ve své formuli neobsahuje logaritmus. Brazilský fyzik řeckého původu C. Tsallis (*1943) zavedl tuto entropii při zobecňování Boltzmann--Gibbsovy fyzikální entropie [5].

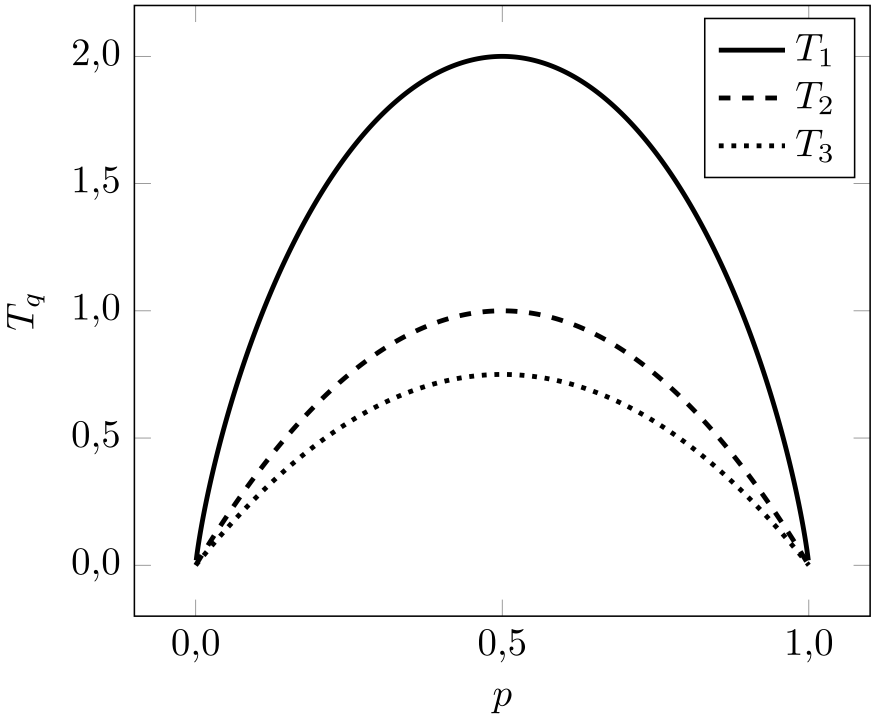
Definice 3. Tsallisovou entropií řádu q nazýváme entropii, která se určuje pro diskrétní náhodnou veličinu $X(p_1, \ldots, p_n)$, $\sum_i p_i = 1$, $0 \le p_i \le 1$, $i = 1, 2, \ldots, n$ ze vzorce

$$T_q(X) = \frac{k}{q-1}\left(1 - \sum_1^n p_i^q\right), \quad \text{kde } k > 0,\ q > 0,\ q \ne 1. \tag{3}$$

Volbou parametru k se upravuje jednotka této entropie. Jednotka této entropie nemá název, ale pro pořádek ji nazveme (k, q)-*tit.*

Vlastnosti $T_q(X)$:

1) Entropie $T_q(X)$ je invariantní vůči kterékoliv permutaci $(p_1, \ldots, p_n)$.
2) $\lim_{q \to 1} T_q(X) = \lim_{q \to 1} \frac{k}{q-1}(1 - \sum_1^n p_i^q) = \lim_{q \to 1} k \cdot \frac{-\sum_i p_i^q \cdot \ln p_i}{1} = k \cdot (-\sum_i p_i \ln p_i)$. Proto se tato entropie značí $T_1(X) = H_1(X)$.
3) $T_2(X) = k \cdot (1 - \sum_i p_i^2)$ protože $\sum_i p_i^2 \le 1$, je $0 \le T_2(X) \le k$. V případě $n = 2$ je pak $T_2(X) = k \cdot 2 \cdot p \cdot (1-p)$, tedy $0 \le T_2(X) \le k \cdot 0{,}5$. Proto je někdy vhodné volit $k = 2$, aby $0 \le T_2(X) \le 1$.
4) $\frac{\mathrm{d}T_q}{\mathrm{d}q} = \frac{\mathrm{d}}{\mathrm{d}q}\left(\frac{k}{q-1}(1 - \sum_1^n p_i^q)\right) = k \cdot (q-1)^{-2} \cdot (-\sum_i p_i^q \cdot \ln p_i) \le 0$, tedy $T_q(X)$ je nerostoucí funkcí pro všechna přípustná q. Vezmeme-li v úvahu ještě vlastnost 2), platí to i pro $q > 0$.
5) Pro n. v. $X(1, 0, \ldots, 0)$ je $T_q(X) = 0$ pro všechna přípustná k a q.



Obrázek 4: Graf $T_q\,(X)$ pro n. v. $X(p, 1-p)$ a $k = 2$, $q = 1, 2, 3$.
$T_1 = -2 \cdot \{p \ln p + (1-p) \ln(1-p)\}$; $T_2 = 4 \cdot p \cdot (1-p)$; $T_3 = 3 \cdot p \cdot (1-p)$.

Z Fadějevových axiomů Tsallisova entropie $T_q\,(X)$ splňuje samozřejmě první a druhý. Obecně pro všechna $q > 0$, $q \neq 1$ třetí axiom nesplňuje, protože platí

$$p_n^q = (q_1 + q_2)^q \neq q_1^q + q_2^q.$$

Není tedy pro uvedená q entropií ve Fadějevově smyslu. Entropie $T_1\,(X)$ samozřejmě Fadějevovy axiomy splňuje.

Uvažujme ještě jednu možnost odhadu neurčitosti diskrétní n. v. X (opět bez logaritmu v definičním vztahu). Je to tzv. lineární entropie $L\,(X)$.

Definice 4. Lineární entropie pro n. v. $X\,(p_1, \ldots, p_n)$, $\sum_i p_i = 1$, $0 \leq p_i \leq 1$, $i = 1, 2, \ldots, n$, je určena vztahem

$$L\,(X) = k \cdot \sum_1^n p_i\,(1 - p_i)\,, \quad k > 0. \tag{4}$$

Vlastnosti $L\,(X)$:

1) Lin. entropie $L\,(X)$ je symetrickou funkcí svých proměnných $p_1, \ldots, p_n$.
2) Pro lineární entropii $L\,(X)$ platí: $L\,(X) = k \cdot \left(\sum_i p_i - \sum_i p_i^2\right) \geq 0$, rovnost nastává pro $X = (1, 0, \ldots, 0)$.

3) Pro $n = 2$ je $L(X) = L(p, 1-p) = k \cdot p(1-p) = T_2(X)$.
4) Pro n. v. $X = \left(\frac{1}{n}, \frac{1}{n}, \ldots, \frac{1}{n}\right)$ je $L(X) = k \cdot \left(1 - \frac{1}{n}\right) < k$.
5) Platí $H_1(X) = H_1(p_1, \ldots, p_n) = -\sum_i p_i \ln p_i \geq L(X) = \sum_i p_i(1-p_i)$. To vyplývá z toho, že $p_i - 1 \geq \ln p_i$ a pak je $p_i(1-p_i) \leq -p_i \ln p_i$ pro všechna i.
6) Jednotka této entropie také zavedena není, budeme ji zde označovat lit^2.

Lineární entropie $L(X)$ odpovídá neurčitosti n nezávislých dichotomických rozhodování:

$$x_1 - non\, x_1; \quad x_2 - non\, x_2; \quad x_n - non\, x_n.$$

Poznámka: V práci [16] je ještě diskutována a-entropie h_a zobecněného pravděpodobnostního schématu. Přitom je třetí Fadějevův požadavek zobecněn na tvar

$$h_a(X) = h_a(tp_1, (1-t)\,p_1, p_2, \ldots, p_n) = h_a(p_1, \ldots, p_n) + p_1^a \cdot h_a(t, 1-t).$$

4. Aplikace

Aplikace není snadná. Obtíže již může činit vymezení prvotního (observačního) jevového pole Ω, A, kde Ω je idealizovaný pozorovaný prostor, A je σ-algebra podmnožin z Ω. Dříve zmíněná množina Ω_n není počátkem zkoumání, je jen součástí schématu pro popis výsledku realizace n. v. X, která přísluší schématu $\mathrm{P} = [\Omega_n, A_n, P_n]$. Míra jistoty realizace n. v. X, patřící prostoru P je entropie. Smyslem aplikace je určení vztahu mezi prvotním jevovým polem $[\Omega, A]$ a konečnými množinami Ω_n, A_n. Tento vztah je dán náhodnou veličinou Y, která má hodnoty v Ω_n a platí pro pravděpodobnostní prostory $[\Omega, A, P]$ a $[\Omega_n, A_n, P_n]$, že

$$P(\{\omega \in \Omega; Y(\omega) = x_i\}) = P_n(X = x_i) = p_i.$$

Pokud se podaří aspoň dostatečně přesně definovat n. v. Y, musíme se pokusit na základě měření o odhad pravděpodobností p_i. Nemůžeme, až na výjimky, stanovit skutečnou hodnotu příslušné entropie, ale jen její odhad, vycházející z relativních četností $\widehat{p_i} = \frac{N_i}{N}$ jednotlivých variant x_i, kde N_i je počet jejich výskytů v souboru N pozorování. Protože vycházíme z odhadů pravděpodobností, vede použití příslušných vzorců pouze k odhadu odpovídající entropie. Takový odhad ale může být vychýlený, zvlášť když výraz pro entropii obsahuje logaritmus, tj. se systematickou chybou. Každý výpočet by měl obsahovat nejen odhad pro entropii, ale i chybu tohoto odhadu.

Je-li $h(p_1, p_2, \ldots, p_n)$ entropie se spojitými parciálními derivacemi $\frac{\partial h}{\partial p_i}$, $i = 1, 2, \ldots, n$, je možné určit chybu Δh měření v bodě $(\widehat{p}_1,\ \widehat{p}_2, \ldots, \widehat{p}_n)$ ze vztahu

$$\Delta h \approx \left| \sum_i \frac{\partial h(\widehat{p}_1, \widehat{p}_2, \ldots, \widehat{p}_n)}{\partial p_i} \Delta p_i \right| .$$

Pro odhad $|\widehat{p}_i - p_i| = \Delta p_i$ je možné použít pro velká N přibližného vztahu, platného s pravděpodobností $1 - \alpha$

$$\Delta p_i \approx u_{1-\frac{\alpha}{2}} \cdot \sqrt{\frac{\widehat{p}_i}{N}(1 - \widehat{p}_i)},$$

kde $u_{1-\frac{\alpha}{2}}$ je $\left(1 - \frac{1}{2}\alpha\right)$-tý kvantil normovaného normálního rozdělení. Podle tohoto vztahu je při $\alpha = 0{,}05$ pro $u_{1-\frac{\alpha}{2}} = 1{,}96$. Odhady chyb jednotlivých entropií se pak budou počítat podle následujících vztahů

$$\begin{aligned}
\Delta H_\alpha\left(X\right) &\approx \frac{1}{\ln 2} \frac{\alpha}{|1-\alpha|} \cdot \frac{1}{\sum_i p_i^\alpha} \cdot \sum_i p_i^{\alpha-1} \Delta p_i \quad [\alpha\text{-}bit] \\
\Delta H\left(X\right) &\approx n\left(n-1\right) \cdot \max \Delta p_i \\
\Delta T_q\left(X\right) &\approx \frac{kq}{|q-1|} \cdot \sum_i p_i^{q-1} \cdot \Delta p_i \\
\Delta L(X) &\approx 2 \cdot \sum_i |1 - 2p_i| \cdot \Delta p_i,
\end{aligned}$$

kde všude dosazujeme za p_i odhad $\widehat{p}_i$.

Příklad 2. Mějme n. v. $X = \left(p_1, p_2, p_3\right)$, kde $0 \le p_i \le 1$, pro $i = 1, 2, 3$, $\sum_1^3 p_i = 1$. Na vzorku o rozsahu $N = 100$ byl proveden odhad pravděpodobností p_i s výsledkem $\widehat{p}_1 = 0{,}2$; $\widehat{p}_2 = 0{,}5$; $\widehat{p}_3 = 0{,}3$. Odhadneme některé z předchozích entropií. Budeme volit $\alpha = 0{,}05$. Pak je pro $\Delta p_1 \approx 1{,}96 \cdot 0{,}1 \cdot \sqrt{0{,}2 \cdot 0{,}8} = 0{,}078$; $\Delta p_2 \approx 1{,}96 \cdot 0{,}1 \cdot \sqrt{0{,}5 \cdot 0{,}5} = 0{,}098$; $\Delta p_3 \approx 1{,}96 \cdot 0{,}1 \cdot \sqrt{0{,}3 \cdot 0{,}7} = 0{,}090$.

Pro odhady jednotlivých entropií pak máme

$$\Delta H_2\left(X\right) \approx 2 \cdot \frac{1}{0{,}38} \cdot (0{,}2 \cdot 0{,}078 + 0{,}5 \cdot 0{,}098 + 0{,}3 \cdot 0{,}090) = 0{,}482.$$

Pro entropii $H_2(X)$ tak je

$$\begin{aligned}
H_2(X) &\approx -\log_2 0{,}38 \pm 0{,}482 = 1{,}40 \pm 0{,}482 \quad [2\text{-}bit].\\
H(X) &\approx 2 \cdot \min(0{,}5; 0{,}5) + 2 \cdot \min(0{,}3; 0{,}2) \pm 6 \cdot 0{,}1 = 1{,}4 \pm 0{,}6 \quad [lit].\\
T_2(X) &\approx 2 \cdot 2(1 - 0{,}38) \pm (0{,}2 \cdot 0{,}078 + 0{,}5 \cdot 0{,}098 + 0{,}3 \cdot 0{,}090) =\\
&= 2{,}48 \pm 0{,}09 \quad [(2{,}2) - tit] \text{ (volili jsme tedy } k = 2).\\
L(X) &= 2 \cdot (1 - 0{,}38) \pm 2 \cdot \{(1 - 2 \cdot 0{,}2) \cdot 0{,}078 + (1 - 2 \cdot 0{,}5) \cdot 0{,}098 +\\
&\quad + (1 - 2 \cdot 0{,}3) \cdot 0{,}090\} = 1{,}24 \pm 0{,}16 \quad [lit^2].
\end{aligned}$$

Jak je vidět, všechny odhady entropií jsou zatíženy velkou chybou. Je proto třeba používat poměrně rozsáhlých náhodných výběrů. Aby byla v našem příkladu chyba desetkrát menší, což by bylo přijatelné, měli bychom rozsah šetření N zvětšit na 10 000.

5. Závěr

K odhadu neurčitosti máme poměrně velké množství měr [1, 5, 14]. To je proto, že ne každá experimentální situace je pro určitou míru vhodná. Rozhoduje intuice a zkušenost experimentátora, která určuje vhodnost konkrétní míry.

Všechny míry neurčitosti diskrétní náhodné veličiny X jsou nezáporné a jejich podstatnou vlastností je, že pro rovnoměrné rozdělení této veličiny nabývají nejvyšších možných hodnot (nebo se jim blíží) a v případě $X = (1, 0, 0, \ldots, 0)$ jsou nulové. Vždy při odhadu neurčitosti se předpokládá buď znalost rozdělení n. v. X (a to spíše výjimečně) nebo jen statistický odhad rozdělení (neurčitost je zde její pravděpodobnostní charakteristika). Naznačili jsme, že je potřebí realizovat odhad neurčitosti rozsáhlejším experimentem.

Někdy místo znalosti rozdělení náhodné veličiny X je možné využít i expertního odhadu (zvláště v humanitních vědách), pak je možné se opřít o teorii fuzzy množin a využít některou z fuzzy měr neurčitosti [10, 11]. Jsou však i teorie, které neurčitosti počítají z jiných znalostí, například Dempster-Schaferova teorie [3].

Poděkování

Autor děkuje ing. Pavlu Střížovi, Ph.D., za technické dopracování tohoto příspěvku včetně obrázků.

Reference

[1] Arndt, C.: *Information measures: Information and its Description in Science and Engeneering*, Springer, 2004, ISBN 978-3-540-40855-0. *cit. 4, 16*

[2] Cover, T. M., Thomas, J. A.: *Elements of Information Theory,* 2. Edition, Wiley-Interscience, 2006, ISBN 0-471-24195-4. *cit. 4*

[3] Půlpán, Z.: *K problematice vágnosti v humanitních vědách*, Academia, Studie 2/1997, ISBN 80-200-0648-6. *cit. 11, 16*

[4] Rényi, A.: On measures of information and entropy, *Proceedings of the 4th Berkley Symposium on Mathematics, Statistics and Probability,* 1961, pp. 547–561. *cit. 4, 6*

[5] Tsallis, C.: On measures of information and entropy, *Journal of Statistical Physics* 52, 1988, pp. 479–487. *cit. 4, 12 a 16*

[6] Kukal, J.: Entropie nejen ve fyzice, *Rozhledy matematicko-fyzikální* č. 4, roč. 86, 2011, str. 13–20. *cit. 4*

[7] Ben-Bassat, M., Raviv, J.: Rényis entropy and the probability of error, IEEE, *Transactions on Information Theory*, Vol. 24, No. 3, pp. 324–331, May 1978. *cit. 4*

[8] Rényi, A.: *Teorie pravděpodobnosti,* Academia, Praha 1972. *cit. 4*

[9] Rényi, A.: On the dimension and entropy of probability distributions, *Acta Math. Acad. Sci. Hung.* 10, 1959, pp. 193–215. *cit. 7*

[10] Půlpán, Z.: *Odhad informace z dat vágní povahy*, Academia, Praha 2012, ISBN 978-80-200-2076-5. *cit. 10, 16*

[11] Půlpán, Z.: Informace a neurčitosti v expertních odhadech, *Informační Bulletin České statistické společnosti*, roč. 33, 1, str. 5–19. *cit. 10, 16*

[12] Černý, J.: *Entropia a informácia v kybernetike,* Bratislava, Alfa 1981. *cit. 4, 9*

[13] Kullback, S.: *Information Theory and Statistics*, New York, John Willey 1959. *cit. 4*

[14] Vajda, I.: Teória informácie a štatistického rozhodovania, Bratislava, Alfa 1982. *cit. 4, 16*

[15] Havrda, J.: *Stochastické procesy a teorie informace,* ČVUT, Praha 1982. *cit. 4*

[16] Vajda, I.: Axiomy a-entropie zobecněného pravděpodobnostního schématu, *Kybernetika*, č. 2, roč. 4, 1968. *cit. 14*

[17] Fadějev, D. K.: K ponjatiu entropii koněčnoj věrojatnostnoj schemy, *Uspěchi matěmatičeskich nauk,* č. 11, 1956. *cit. 9*

[18] Zvárová, J.: On asymptotic Behavior of Sample Estimator of Rényi's Information of order α, In: *Transaction of the 6th Prague Conference on Information Theory*, Academia, Prague 1973. *cit. 7*

ROBUSTNÍ MODEL KONFERENCE SE SMÍŠENÝMI EFEKTY

REPORT ON THE ROBUST 2024 CONFERENCE

Ondřej Vencálek

E-mail: ondrej.vencalek@upol.cz

Dvacátá třetí letní škola JČMF ROBUST 2024 se uskutečnila ve dnech 8. až 13. září 2024 v Bardějově. Při celkovém počtu přibližně 90 účastníků jde o jednu z největších (ne-li největší) letošních konferencí českých a slovenských matematiků-statistiků. Jedním ze spoluorganizátorů konference je tradičně i Česká statistická společnost. Rád bych se s odstupem dvou měsíců ke konferenci vrátil a sepsal několik svých dojmů a postřehů. Jak již název tohoto textu napovídá, půjde tentokrát o dojmy „smíšené“.

Stalo se jakousi tradicí, že ozvěny konference Robust začínají vždy zhodnocením odlehlosti lokality, kde se konference uskutečnila. To není vzhledem k názvu konference překvapivé, vždyť problém „odlehlých pozorování“ stál u zrodu celého odvětví statistiky známého jako „robustní metody“. Lze konstatovat, že lokalita Bardějov, v níž se letos konference uskutečnila, je mezi všemi dějišti Robustu jednak nejvzdálenější od Prahy (vzdušnou čarou 500 km, přičemž tato nejkratší spojnice vede mj. přes území Polska), ale rovněž je natolik vzdálená od hlavního města Slovenska (vzdušnou čarou 340 km), že je jen málo míst v zemi našich východních sousedů, která by byla od Bratislavy dále. To se odrazilo mj. v tom, že i mnozí účastníci konference z metropole na Dunaji byli v Bardějově vůbec poprvé.

Účastníci, kteří na Robust jezdí opakovaně a rádi, se pravděpodobně většinově shodnou na tom, že volba zajímavé a ne příliš známé lokality je jednou z hlavních předností Robustu. Také proto je této volbě organizátory tradičně věnována velká pozornost a péče.

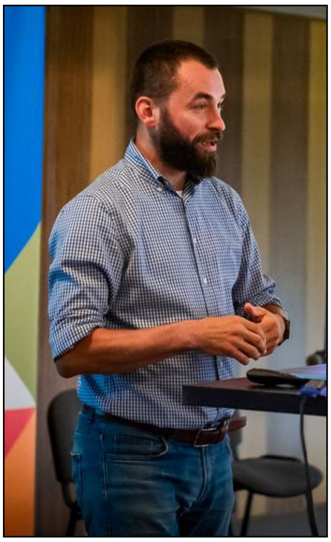
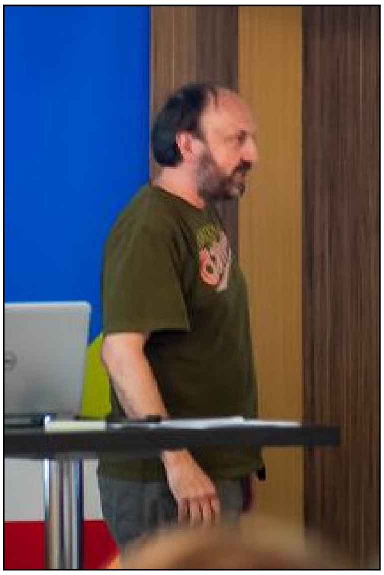
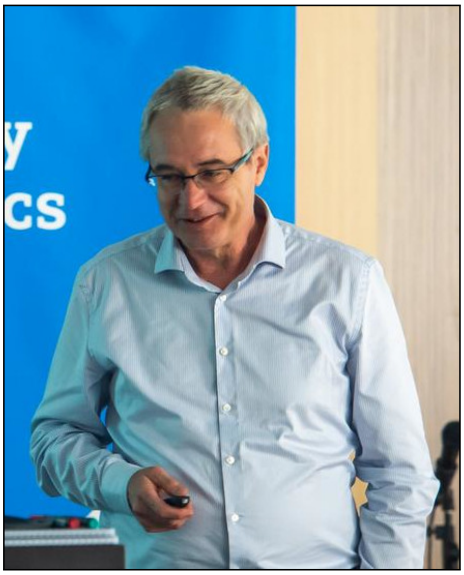
Organizátoři se také tradičně snaží udělat konferenci co nejlevněji, aby umožnili účast co největšímu počtu zájemců. Stojí za zaznamenání, že šestidenní konferenci dokázali v roce 2024 uspořádat při konferenčním poplatku *zahrnujícím ubytování a stravu* ve výši 6 250 Kč (resp. 250 Euro), přičemž pro studenty byla tato částka ještě o něco nižší (5 500 Kč, resp. 220 Euro). Tedy všechna čest, toto se (opět) povedlo.

Důležitým rysem, který podstatně ovlivnil atmosféru prvních dní letošního Robustu, byla účast významných zahraničních hostů, a to v rámci „paralelně“ konaného workshopu AMISTAT 2024 (Analytical Methods in Statistics). Slovo paralelně uvádím v uvozovkách, neboť zapojení workshopu do programu bylo spíše „sériové“, nikoliv „paralelní“. Toto zapojení mělo řadu pravděpodobně nezamýšlených efektů. Posluchač vždy chtě nechtě porovnává kvalitu přednášek. Organizátoři Robustu si mohou blahopřát, že v konkurenci zahraničních hostů se významně prosadili zejména (ovšem nikoliv výlučně) zvaní řečníci Robustu. Ti byli z řad „domácích“: Daniel Klein z Košic, David Kraus z Brna, Tomáš Mrkvička z Českých Budějovic, Samuel Rosa z Bratislavy či Michal Pešta z Prahy. Z přednášek zahraničních účastníků zaujal zejména Marc Hallin a Peter Filzmoser.

Další Robust se kvapem blíží – zbývá něco málo přes rok. Kde bude? To ještě nevím, ale těším se, že opět v „nějaké díře“, jak zpívá Zdeněk Fabián.Důležitější otázkou než „Kde bude?“ je však otázka „Jaký bude?“ To záleží na tom, jak se budou cítit jeho účastníci. Věřím, že organizátoři udělají maximum pro to, aby se všichni účastníci na Robustu cítili dobře.

V Prostějově dne 7. listopadu 2024, Ondřej Vencálek




PERNÉ STATISTICKÉ DNY POD PÁLAVOU 2024
REPORT ON STATISTICAL DAYS 2024

Kolektiv Ústavu matematiky FSI VUT v Brně

Adresa: Fakulta strojního inženýrství Vysokého učení technického v Brně, Technická 2896/2, 616 69 Brno-Královo Pole

E-mail: hubnerova@fme.vutbr.cz, matej.benko@vutbr.cz

I letos jsme navázali na tradici Statistických dnů pořádaných ČStS od roku 1994 a setkali jsme se v obci Perná na Pálavě 24. – 26. května 2024. Letošní 16. ročník setkání organizoval Ústav matematiky Fakulty strojního inženýrství Vysokého učení technického v Brně ve spolupráci se společností SC&C Partner, s.r.o.

1. den: příjezd, první přednášky a ochutnávka vín

Hned v prvních přednáškách jsme se dozvěděli zajímavosti a způsob implementace Shewhartových regulačních diagramů v podání předsedy ČStS Ondřeje Vencálka a Tomáše Jurczyka. Po přestávce nás Zdeněk Karpíšek obeznámil s různými mírami závislostí náhodných jevů a po něm David Kraus představil své výsledky o neurčitosti rekonstrukcí funkcionálních trajektorií. Po poslední přestávce tohoto odpoledne jsme se vrhli k aplikacím statistiky v softwaru Minitab, a to konkrétně využití Random Forests, TreeNet, a CARTu v podání Tomáše Trávnička z SC&C Partner. Po vydatné večeři jsme se vybrali na ochutnávku vín do Vinařství u Matoušků. Ochutnali jsme spolu třináct vín a to konkrétně osm bílých, dvě růžové a tři červené. Všechna vína měla společné to, že byla výtečná.

2. den: výlet do přírody a soutěživý večer

Po snídani jsme se vrhli do vědecké činnosti, konkrétně do maratonu přednášek. Je potřeba zmínit, že velice zajímavých, pokrývajících jak teoretickou matematickou statistiku, tak i užitečné aplikace. Petr Volf nám prezentoval problém optimálního hledání ztracené osoby formulovaný pomocí teorie pravděpodobnosti a po něm Matej Benko pohovořil o stochastických diferenciálních rovnicích, jejich souvislostech a aplikacích v jiných oblastech jako parciální diferenciální rovnice nebo strojové učení.

Poté nás Michal Fusek a Jan Holešovský seznámili s problematikou odhadu míry lokální závislosti extrémních hodnot v časových řadách. Na závěr nás obohatil pan profesor Jaromír Antoch z Univerzity Karlovy svými zkušenostmi s výběrovými šetřeními a kontrolami pro odhalení řídce se vyskytujících závad.

Po obědě se většina z nás vydala na turistický výlet do Pavlova a zbytek fandil českým hokejistům v boji se Švédskem v semifinále MS v hokeji. Fandové hokeje měli štěstí, jelikož se vyhnuli vydatnému osvěžení v podobě bouře a bláta. Ale za to přišli o krásné výhledy (sice částečně pod mrakem) a přírodu Jižní Moravy. Obdiv patří neodvážnějším, kteří došli celý plánovaný okruh zpět do Perné.

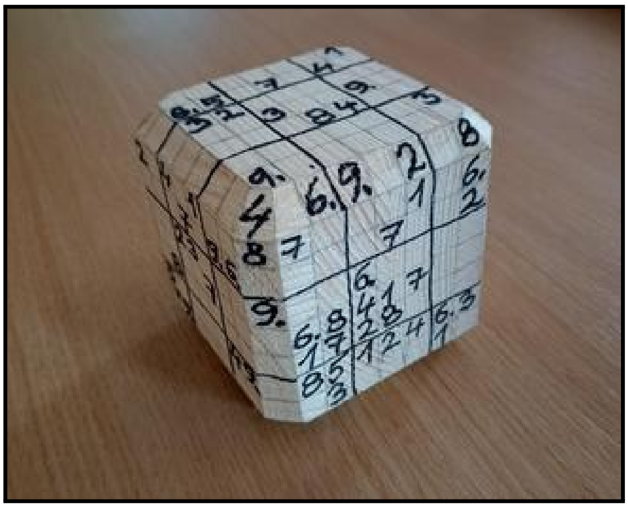
Po večeři jsme se vrhli do soutěžení. Libor Žák z FSI VUT nám připravil již tradiční souboj ve znalostech chutí vína. Tentokrát byl test našich chuťových buněk komplexní. Byli jsme testováni z rozpoznání vín Muškát moravský, Pálava, Rulandské šedé, Chardonnay a Tramín červený. Nestačilo jen rozpoznat víno, ale také jejich cukry, kyselinky, alkohol a rok. Po Matejovi Benkovi z VUT pomyslný pohár vítězů převzal Viliam Páleník, a na závěr lze jen dodat „sláva vítězům a čest poraženým“.



3. den: Poslední přednášky a návrat domů

Poslední den jsme si poslechli přednášky z aplikací statistiky a zejména zpracování dat a poté o zajímavých hrách. Iveta Stankovičová a Iveta Syneková nám představily Integrovaný statistický informační systém, který využívá Statistický úřad Slovenské republiky.

Po nich nám Viliam Páleník ukázal, jak disponibilní důchody domácností využít jako alternativní indikátor reálné konvergence. Po poslední přestávce nám Josef Košťálek popsal vyvinutý projekt v programu Microsoft Excel pro výuku statistiky, nepochybně zajímavou byla přednáška Eriky Liptákové o postoji studentů ekonomické fakulty k výuce statistiky a na závěr nám Pavel Stříž představil některé varianty sudoku.



Závěr

Na závěr bychom chtěli poděkovat všem organizátorům, společnosti SC&C Partner a také účastníkům, zejména těm, kteří si připravili přednášku do vědecké části programu. Prožili jsme krásné tři dny plné zážitků a statistiky a těšíme se na příští ročník.

NOVINKY V TEXLIVE 2024

NEWS FROM TEXLIVE 2024

Pavel Stříž

E-mail: pavel@striz.cz

Aktualizací a nových balíčků je tradičně hodně, opět jsem se pokusil vybrat klíčové záležitosti.

Spolupráce s nástroji

Příjemným překvapením roku 2024 jsou balíčky podporující komunikaci s jinými výpočetními nástroji či hlubší propojení s jazykem Lua. Jedná se o balíčky numerica, numerica-plus, numerica-tables; zx-calculus, resolsysteme, lt3luabridge (Vít Novotný); functional a latex2pydata.

Průběžné ukládání

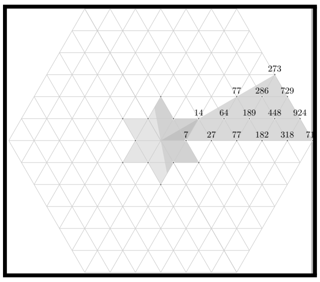
Aby se výpočty a kresby neopakovaly při každém běhu TEXu, vznikají balíčky řešící tuto situaci. Jedná se především o balíčky scontents, robust-externalize, hvextern a memoize.

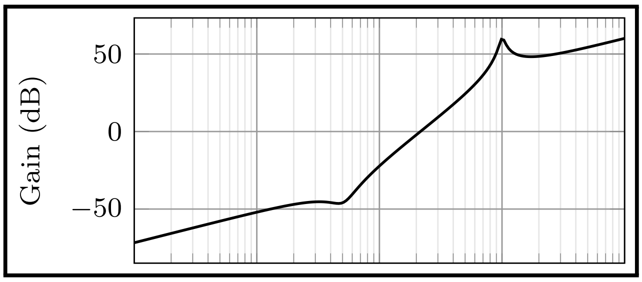
Užitečné nástroje

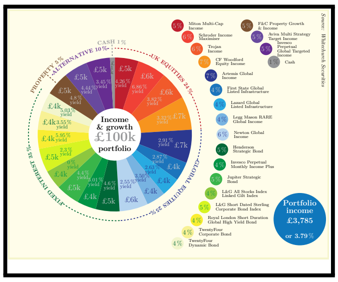
Nástroj Xindex je zde již nějakou dobu, s pomocí balíčku Lua-UCA (Michal Hoftich) zvládá podporu mnoha jazyků. Mít jazyk navázaný přes definice firmy Unicode byl dlouho dobu otevřený problém. První vážný pokus o optimalizaci sazby přes strany je shrnut v balíčku lua-widow-control (Max Chernoff). Více o otevřených TEXových problémech viz `https://tug.org/tc/devfund/criteria.html`. Zvláštní skupinu tvoří nástroje vyvíjené Petrem Olšákem, viz OPtex, OPmac a OPbible. Od našince stojí dále za pozornost nový balíček Michala Hofticha responsive a markdown od Víta Novotného.

Grafy a grafika

Zaujaly mě balíčky rank-2-roots, bodeplot, wheelchart (obrázky níž), mcf2graph, mptrees, tikz-bricks a liftarms.







Typografie

V této oblasti jsem si poznačil balíčky lua-ul (varianta k balíčkům soul a ulem), circularglyphs, fvextra (nadstavba balíčku fancyvrb) a witharrows, viz ukázky.

$$
\begin{aligned}
2xy' - 3y = \sqrt{x} &\iff 2x(K'y_0 + Ky_0') - 3Ky_0 = \sqrt{x} \\
&\iff 2xK'y_0 + K(2xy_0' - 3y_0) = \sqrt{x} \\
&\iff 2xK'y_0 = \sqrt{x} && \text{we replace } y_0 \text{ by its value} \\
&\iff 2xK'x^{\frac{3}{2}} = x^{\frac{1}{2}} && \text{simplification of the } x \\
&\iff K' = \tfrac{1}{2x^2} && \text{antiderivation} \\
&\iff K = -\tfrac{1}{2x}
\end{aligned}
$$

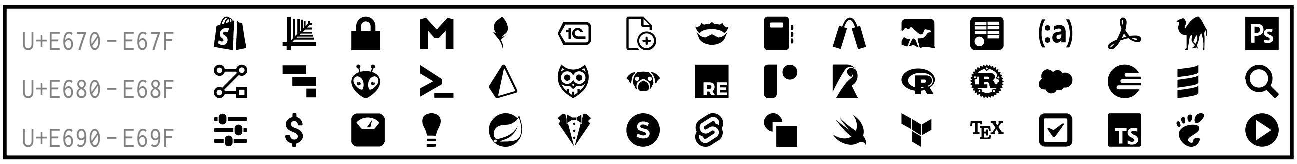
$$
\begin{aligned}
A &= B \\
&= C \\
&= D \qquad \text{text} \\
&= E \\
&= F
\end{aligned}
$$

$$
\begin{aligned}
3(2x+4) &= 6 && \div 3 \\
2x + 4 &= 2 && -4 \\
2x &= -2 && \div 2 \\
2x &= -1
\end{aligned}
$$

Znaky, glyfy a písma

Pro pobavení stojí balíčky tikzpingus, tikzlings a rpgicons. Mnohem důležitější je rodina písem New Computer Modern, viz balíčky newcm a fontsetup, srovnání lze provést za pomoci balíčku unicodefonttable (rozšíření pomůcky `testfont`).

Nezarmoutilo ani písmo Inconsolata Nerd (horní ukázka, balíček inconsolata-nerd-font), balíček simpleicons (spodní ukázka) a podpora pro písmo Junicode (balíček stejného jména), vhodný především pro humanitní obory.



	wolfram	`\simpleicon{wolfram}`
	wolframlanguage	`\simpleicon{wolframlanguage}`
	wolframmathematica	`\simpleicon{wolframmathematica}`

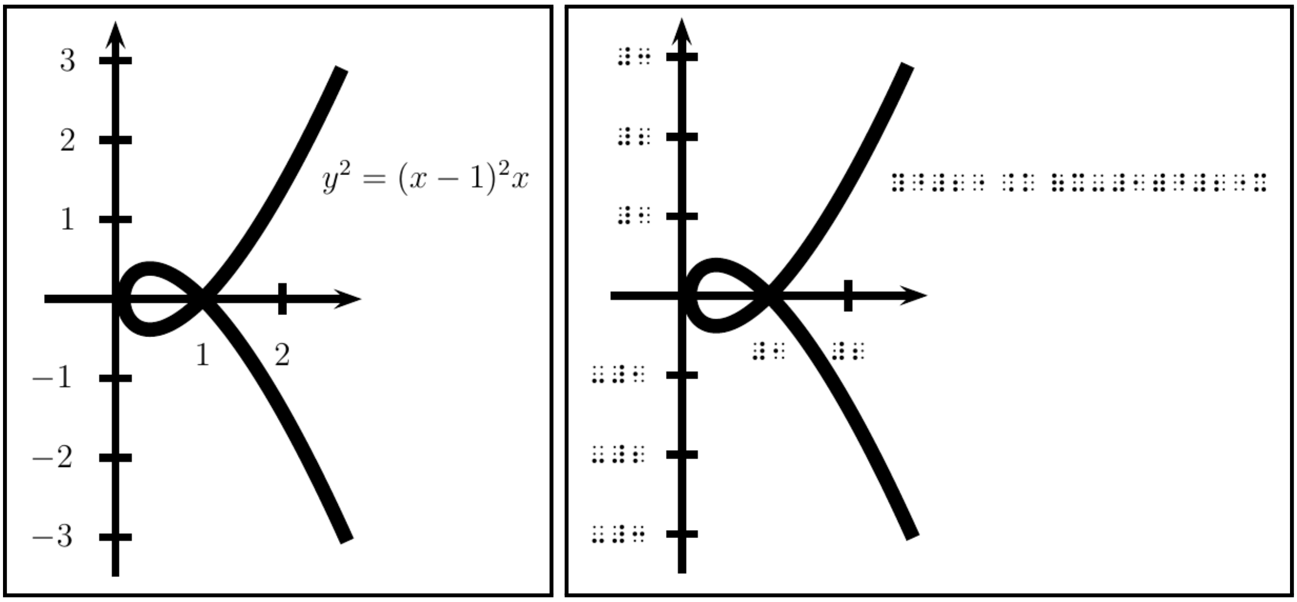
Podpora pro zrakově postižené

Poslední, ale nikoliv nevýznamnou oblastí, je TeXová podpora v této oblasti. Ze starších projektů je známý balíček braille, z nových bych rád upozornil na balíčky spelatex (Speech-enabled LaTeX) a colorblind, více viz výzkum Paula Tola, `https://personal.sron.nl/~pault/`.

Na balíček unicode-math navazuje balíček unicode-math-braille. Zde je vzorek z dokumentace, v prvním sloupci je TeXový příkaz, v druhém vysázený symbol, v třetím reprezentace v Braillově písmu a anglický ekvivalent.

`\coprod`	$\coprod$	⠫⠨⠏⠻	coproduct operator
`\sum`	$\sum$	⠨⠠⠎	summation operator
`\int`	$\int$	⠮	integral operator
`\iint`	$\iint$	⠮⠮	double integral operator
`\iiint`	$\iiint$	⠮⠮⠮	triple integral operator

Na závěr bych rád upozornil na TUGem podpořený projekt č. 33, který vyústil v balíček `latex2nemeth`. Objevil jsem jejich program pro Javu v adresáři `~/texlive/2024/texmf-dist/scripts/latex2nemeth/`, resp. v adresáři `~/texlive/2024/bin/x86_64-linux/`, a po jistém experimentování se mi konverze písma podařila.



Je možné, že by se konverze dala udělat přímo v LuaTeXu, ale nezkoušel jsem to.

POZVÁNKA NA KONFERENCI OSSCONF 2025
INVITATION TO THE OSSCONF 2025 CONFERENCE

Miroslav Kvaššay

E-mail: miroslav.kvassay@fri.uniza.sk

Otvorený softvér vo vzdelávaní, výskume a v IT riešeniach

Fakulta riadenia a informatiky Žilinskej Univerzity v spolupráci so **Spoločnosťou pre otvorené informačné technológie** (SOIT) a združeniami **CSTUG**, **Česká statistická společnost**, **Geokomunita**

organizujú tento rok už 13. ročník konferencie

OSSConf 2025

Konferencia sa uskutoční v priestoroch Fakulty riadenia a informatiky v Žiline v termíne

1. júla – 3. júla 2025

Cieľom našej konferencie je poskytnúť priestor pre informovanie o novinkách vo vývoji otvoreného softvéru a otvorených technológií, o možnostiach využitia týchto nástrojov vo vede a vzdelávaní a taktiež poskytnúť priestor pre neformálne priateľské stretnutie užívateľov a priaznivcov otvorených informačných technológií.

Pri príležitosti konferencie bude vydaný zborník príspevkov. Príspevky autorov prijímame buď ako nerecenzované príspevky, z ktorých uverejňujeme len abstrakt alebo recenzované príspevky, ktoré budú posudzované dvomi nezávislými recenzentmi. Predpísaná šablóna pre písanie príspevkov v typografickom systéme TEX/LATEX bude k dispozícii na webovej stránke konferencie. Príspevky prijímame v slovenskom, českom alebo anglickom jazyku.

Dôležité termíny

Recenzované príspevky:	do 1. júna 2025.
Abstrakty nerecenzovaných vystúpení:	do 15. júna 2025.

Sekcie

Open AI	**LATEX, R a ich priatelia**	**Vývoj OSS**
OSS vo vzdelávaní	**Open GIS & Open Data**	**Open Hardware**

Našu konferenciu sa snažíme, tak ako po iné roky, udržať bez platenia vložného. Pre záujemcov môžeme za úhradu zabezpečiť obedy v stravovacom zariadení Žilinskej univerzity (aktuálnu cenu uverejníme na web stránke). Účastníci, ktorí majú záujem o ubytovanie si toto vybavujú individuálne podľa pokynov zverejnených na web stránke konferencie. Predpokladaná cena zborníka bude 8 – 10 €.

Bližšie informácie o konferencii budeme postupne zverejňovať na http://ossconf.soit.sk/.

VÁNOČNÍ DÁREK A PF2025!
CHRISTMAS GIFT AND PF2025!

Pavel Stříž

E-mail: `pavel@striz.cz`

Vážené kolegyně, vážení kolegové,

v zoufalé snaze upočítat masivní výpočetní problém pro mého kamaráda, Štefana Pešky z Žiliny, jsem si založil malé výpočetní centrum. Kdyby někdo potřeboval výpočetní sílu, ozvěte se mi. Je to takový malý vánoční dárek.

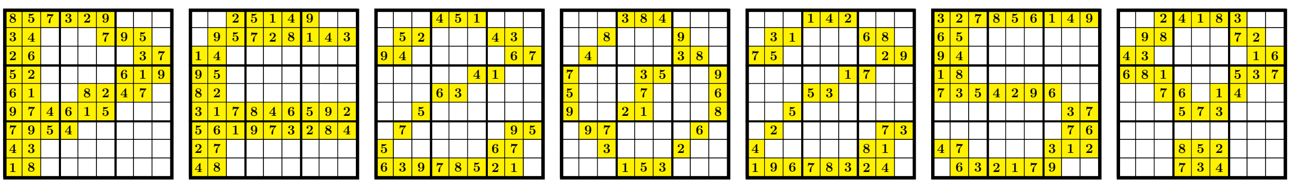
Ale mnohem větší dárek je objev serveru `https://compute.cudo.org/`, kde užití vCPU a vGPU je sice zpoplatněné, ale rozumně, počítá to stabilně a bez intervencí od administrátorů či majitelů serverů. Bez větších problémů jsem počítal na více než 250 silných jádrech a nezbankrotoval jsem. Mohu s klidným svědomím doporučit!

Měl jsem takový příjemný pocit, že jsem objevil „díru“ ve výpočetním světě. Nabídka silných virtuálních strojů je drahá, u menších firem jsem se setkal s tím, že mi výpočty majitelé mazali, a ještě měli tu drzost mi psát, že mi systém někdo naboural, či mi snižovali výkon (dokonce až na třetinu).

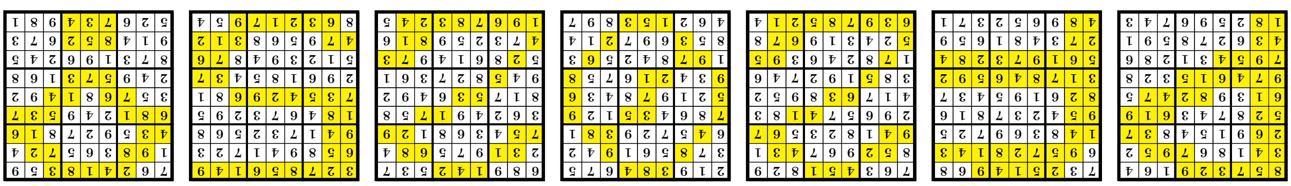
Objevená firma se zaměřuje na výpočty přes vGPU, ale vCPU nabízí, protože ty na řízení vGPU potřebují. Nevadilo jim, že jsem jim na týdny výpočetně zabral všechny volné vCPU.

⋆ ⋆ ⋆

S rokem 2024 se tradičně rozloučíme drobnou symbolickou novoročenkou. Jedná se o nápad z nápověd v sudoku vytvořit písmenka, a to tak, aby sudoku byla platná a hratelná. Zde je PF2025! pro příznivce této logické hry plus řešení. Každé sudoku je řešitelné samostatně, nejsou propojená.



Příjemné prožití vánočních svátků a šťastný nový rok 2025!



Informační bulletin České statistické společnosti vychází čtyřikrát do roka v českém vydání. Příležitostně i mimořádné české a anglické číslo. Vydavatelem je Česká statistická společnost, IČ 00550795, adresa společnosti je Na padesátém 81, 100 82 Praha 10. Evidenční číslo registrace vedené Ministerstvem kultury ČR dle zákona č. 46/2000 Sb. je E 21214. Časopis je sázen v programu TEX, ve formátu LuaHBTEX s písmy balíku CSfonts.

The Information Bulletin of the Czech Statistical Society is published quarterly. The contributions in the journal are published in English, Czech and Slovak languages.

Předseda společnosti: doc. Mgr. Ondřej Vencálek, Ph.D., Katedra matematické analýzy a aplikací matematiky, Přírodovědecká fakulta Univerzity Palackého, 17. listopadu 12, 771 46 Olomouc, e-mail: ondrej.vencalek@upol.cz.

Redakce: prof. RNDr. Gejza Dohnal, CSc. (šéfredaktor), prof. RNDr. Jaromír Antoch, CSc., doc. RNDr. Zdeněk Karpíšek, CSc., RNDr. Marek Malý, CSc., doc. RNDr. Jiří Michálek, CSc., prof. Ing. Jiří Militký, CSc., doc. Ing. Iveta Stankovičová, PhD., doc. Mgr. Ondřej Vencálek, Ph.D.

Redaktor časopisu: doc. Mgr. Ondřej Vencálek, Ph.D., ondrej.vencalek@upol.cz.
Informace pro autory jsou na stránkách společnosti, http://www.statspol.cz/.

DOI: 10.5300/IB, http://dx.doi.org/10.5300/IB
ISSN 1210–8022 (Print), ISSN 1804–8617 (Online)

Toto číslo bylo vytištěno s laskavou podporou Českého statistického úřadu.