

INFORMAČNÍ BULLETIN



České statistické společnosti

Ročník 33, číslo 1, březen 2022

ZPRÁVA O ČINNOSTI ČSTS V ROCE 2021

CZECH STATISTICAL SOCIETY IN 2021

Ondřej Vencálek

E-mail: ondrej.vencalek@upol.cz

Členská základna K 31. 12. 2021 měla ČStS 195 členů. V průběhu roku 2020 přibýlo 13 nových členů, ale protože 18 členům bylo členství ukončeno, celkový počet členů se snížil o 5.

Členská schůze ČStS a volby Členská schůze ČStS se uskutečnila online dne 10. června 2021. Zúčastnilo se celkem 32 členů. Byla přečtena a schválena zpráva o činnosti za rok 2020 a zpráva o hospodaření za rok 2020. Proběhly volby výboru společnosti, předsedy společnosti a revizora společnosti. Oproti období let 2019–2021 došlo ve složení výboru ke dvěma změnám. Do výboru se rozhodli nekandidovat M. Malý a J. Klaschka. Nově byli do výboru zvoleni K. Helisová a T. Hlavsa. Předsedou byl na následující dva roky opět zvolen O. Vencálek. Funkci revizora bude i nadále zastávat D. Blatná. Detailní výsledky voleb byly zveřejněny v Informačním bulletinu ČStS č. 2/2021.

Akce ČStS Mikuklášský statistický den byl pořádán dne 7. prosince 2021 v prostorech MFF UK v Praze na Karlíně. Akce byla streamována. Zúčastnilo se 12 účastníků v Praze a dalších 26 online. Bylo předneseno šest příspěvků.

Vznik odborné skupiny ČStS věnující se problematice sběru a analýzy dat souvisejících s epidemií COVID-19 Výbor ČStS schválil hlasováním per-rollam zřízení odborné skupiny ČStS věnující se problematice sběru a analýzy dat souvisejících s epidemií COVID-19.

Zakládající členové skupiny O. Vencálek, T. Fürst, A. Komárek, M. Kulich, S. Katina a G. Dohnal zveřejnili 1. 2. 2021 prohlášení skupiny „Pět kroků (nejen) ke zvládnutí epidemie“. Vznik skupiny reflektovala četná média a vzbudil velký ohlas ve veřejném prostoru. Činnost skupiny je dokumentována na stránce <https://www.statapol.cz/odborna-skupina-statistiku-k-epidemii-covid-19>.

Spolupráce s Českým statistickým úřadem Český statistický úřad je dlouhodobě partnerem ČStS a podporuje její činnost například zajištěním tisku Informačního bulletinu ČStS, poskytnutím sídla společnosti a prostoru pro konání členských schůzí ČStS.

Členství v mezinárodních organizacích Naše společnost je od roku 2011 členem Federace evropských národních statistických společností FENStatS (The Federation of European National Statistical Societies). Členové ČStS jsou průběžně informováni o dění ve FENStatS. Na členský příspěvek ve výši 200 EUR jsme úspěšně čerpali dotaci prostřednictvím Rady vědeckých společností ČR. Podobně hodláme postupovat i v roce 2022. Valné shromáždění FENStatS se konalo online 13. září 2021. ČStS zastupoval předseda společnosti.

Členství v Radě vědeckých společností ČR ČStS je členem Rady vědeckých společností ČR (RVS). Plenární zasedání RVS se konalo online na platformě ZOOM 24. února 2021. Byly schváleny změny stanov RVS týkající se zejména možnosti pořádat online zasedání a konání hlasování. Do RVS byly přijaty další dvě společnosti. Na konci roku byla RVS odevzdána zpráva o činnosti naší společnosti.

Další činnosti

- Byla vydána 4 řádná čísla Informačního bulletinu ČStS a jedno číslo mimořádné.
- Byla pravidelně aktualizována webová stránka společnosti – bylo zveřejněno celkem 25 příspěvků v sekci novinky a byly provedeny úpravy na dalších stránkách.
- Předseda společnosti reprezentoval ČStS na 20. slovenské statistické konferenci a 18. slovenské demografické konferenci konané ve dnech 9. – 10. září 2021 v Banské Bystrici.

Akce chystané v roce 2022

- Setkání odborné skupiny ČStS k epidemii COVID-19 – 9. 2. 2022.
- Členská schůze ČStS – 3. 5. 2022.
- 22. letní škola JČMF Robust – 12. – 17. června 2022.
- Statistické dny v Telči – 23. – 25. září 2022.
- Mikuklášský den ČStS – kolem 6. 12. 2022.

V Prostějově dne 18. 2. 2022 (aktualizováno 18. 3. 2022).

Ondřej Vencálek
předseda ČStS

INFORMACE A NEURČITOSTI V EXPERTNÍCH ODHADECH

INFORMATION AND UNCERTAINTIES IN EXPERT ESTIMATIONS

Zdeněk Půlpán

Adresa: Prof. RNDr. Zdeněk Půlpán, CSc., Na Brně 1952, 500 09 Hradec Králové 9

E-mail: zdenek.pulpan@post.cz; zdenek.pulpan@upce.cz

Abstrakt: Každé rozhodování je možné z hlediska psychologie redukovat na rozhodování mezi variantou hlavní, uvažovanou, a variantami alternativními. Sémantizovaná informace pro fuzzy množiny nebo intuicionistické fuzzy množiny nám umožňuje přejít od kvalitativní i ke kvantitativní stránce rozhodování, zaměřené na volbu pouze jediné ze dvou možných variant. Je popsána metoda, jak na základě odhadu elementárních nejistot kvantifikovat neurčitost a potřebnou informaci pro rozhodování v tomto jednoduchém modelu. Z popsaného je také možné odvodit zobecnění pro komplikovanější rozhodovací situaci.

Klíčová slova: rozhodování, neurčitosti, odhad informace, fuzzy množiny, intuicionistické fuzzy množiny.

Abstract: Every decision is possible with the terms of psychology reducible to deciding between alternative main consideration and alternative options. The semantic information of intuitionistic fuzzy sets allows us to focus on the quantitative aspect of this type of decision-making that is based on only one main alternative. The article describes the method of estimation based on the fundamental uncertainties and makes quantify uncertainty and information that is needed for decision-making in such a simple model. From the described method is also possible to derive a generalization of complex decision-making situation.

Keywords: decision making, uncertainties, estimation of information, fuzzy sets, intuitionistics fuzzy sets.

1. Úvod

Žijeme v době intenzivního používání dotazníků (testů). Dotazníkům se přisuzuje dost vysoká vypovídací hodnota, na jejich výsledku závisí úspěšnost přijetí i absolvování střední školy a pak i vstup na školu vysokou. Kvalifikační zkoušky v různých oborech se realizují pomocí dotazníků, probíhají různá

marketingová šetření, výzkumy veřejného mínění, atd. Kdo je podroben dotazování, musí se *rozhodnout* pro některou z variant odpovědi (dokonce i tak, že nebude odpovídat).

Každé rozhodování je volbou mezi několika alternativami a předpokládá existenci určitého (velmi často jen konečného) souboru dobře identifikovatelných a od sebe dostatečně odlišitelných alternativ. Psychologie nás poučuje, že rozhodování pro právě jednu z více uvažovaných alternativ si respondent redukuje na posloupnost rozhodování vždy jen mezi dvěma alternativami, z nichž obvykle jedna je z určitých subjektivních důvodů respondentem preferována a považována za hlavní, nejdůležitější, správnou, atd. Rozhodovací proces lze pak modelovat orientovaným stromovým grafem, kde z každého uzlu vycházejí jen dvě hrany, jejichž koncový uzel je obrazem určité alternativy. Každá z alternativ má pak u respondenta jistou míru volby, postup v grafu je možný na základě rozhodnutí, které je u respondenta podloženo jistými důvody. Rozhodování je poznamenáno nejistotou a důvody pro určitá rozhodnutí závisejí na mnoha i nepředvídatelných faktorech. Tušíme, že zde hraje roli náhoda, kterou je možné odhadnout pravděpodobností. Ta může mít svůj původ subjektivní (například z osobní zkušenosti) nebo může být objektivizována odhadem z relativních četností mimo nás realizovaných určitých rozhodnutí za přibližně stejných podmínek.

Nabízí se rozhodování modelovat konečným souborem alternativ $\Omega = \{\alpha_1, \alpha_2, \dots, \alpha_n\}$ a pravděpodobnostmi jejich realizace $p_i = p(\alpha_i)$, $i = 1, 2, \dots, n$, $0 \leq p_i \leq 1$, kde $\sum_1^n p_i = 1$. Množina alternativ Ω spolu s pravděpodobností $P(A) = \sum_{\alpha_i \in A} p_i$ pro $A \subset \Omega$ může být základem pro formalizaci rozhodování v prostoru alternativ Ω . Dostáváme tak pravděpodobnostní prostor rozhodování $[\Omega, 2^\Omega, P]$.

2. Různé míry neurčitosti a odpovídající informace

Neurčitost H (nazýváme ji také entropie) rozhodnutí pro určitou alternativu v pravděpodobnostním prostoru $[\Omega, 2^\Omega, P]$ může být určena ze známého Shannonova vztahu [3]

$$H = - \sum_1^n p_i \log_z p_i, \quad 0 \leq H \leq \log_z n, \quad (1)$$

výraz $0 \cdot \log_z 0$ definujeme limitou $\lim_{x \rightarrow 0^+} x \log_z x = 0$, kde $z > 1$ určuje jednotku neurčitosti. Je-li $z = 2$, pak jednotka neurčitosti je 1 bit, je-li $z = e$, základ přirozeného logaritmu, je jednotkou 1 nit, v ostatních případech nazveme jednotku 1 lit. (Převod nitů na bity lze realizovat, obsahuje-li stejný

výraz logaritmy, platí totiž $\log_2 x = \log_e x / \log_e 2$; $\log_e x$ označujeme také jako $\ln x$.)

Představa neurčitosti rozhodování o velikosti 1 bit je velice jednoduchá. Je to neurčitost rozhodování mezi dvěma stejně pravděpodobnými alternativami:

$$\Omega = \{\alpha, \beta\}, \quad p(\alpha) = p(\beta) = 0,5; \quad H = -0,5 \log_2 0,5 - 0,5 \log_2 0,5 = 1 \text{ [bit]}.$$

Nejjednodušší model rozhodování předpokládá volbu právě jedné ze dvou relevantních (možných, uvažovaných) variant α, β . Neurčitost je pak určitou mírou „síly volby“ některé z variant. Neurčitost jako jedna z charakteristik rozhodovacího procesu může být kvantifikována například na základě subjektivně nebo objektivně odhadnuté pravděpodobnosti $p \in \langle 0, 1 \rangle$ volby jedné ze dvou variant, např. volbou α (když druhá varianta je $\neg\alpha = \beta$, Shannonovým vztahem (a viz Obr. 1))

$$H(p) = -p \cdot \log_2 p - (1 - p) \log_2 (1 - p) \text{ [bit]} \quad (1b)$$

kde je $H(p) \in \langle 0, 1 \rangle$ (přitom klademe $0 \cdot \log_2 0 = 0$).

Jedna ze základních vlastností neurčitosti (1b) je její symetrie vůči variantám α, β :

$$H(p) = H(1 - p); \quad H(0,5) = 1 = \max_p \{H(p)\}; \quad H(0) = H(1) = 0. \quad (2)$$

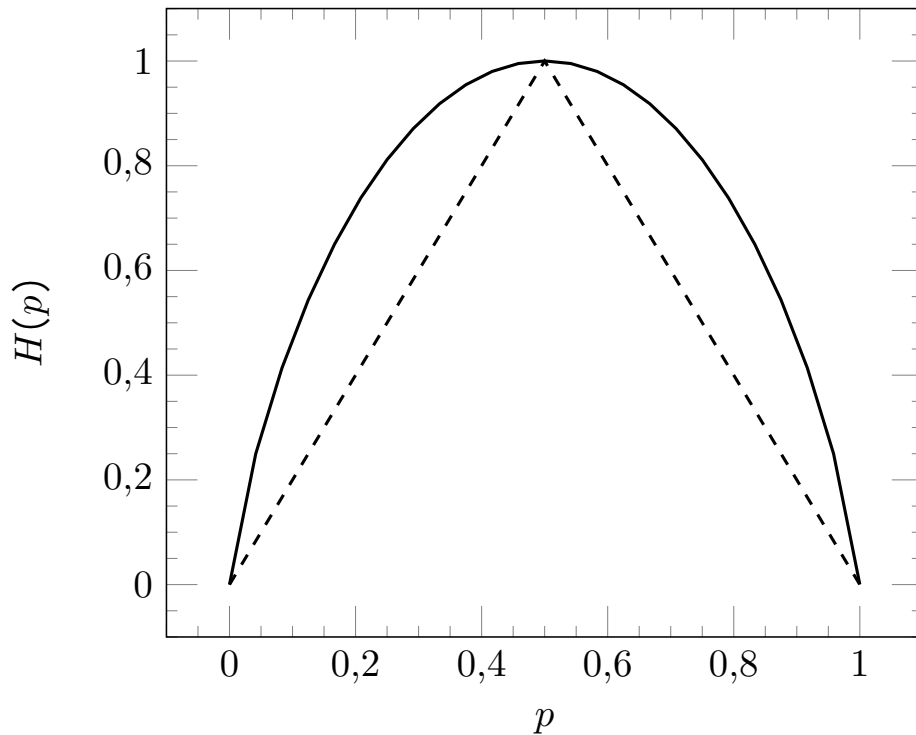
Veličinu $I(p) = 1 - H(p)$ je možné interpretovat jako míru informace, „spotřebované“ pro volbu jedné z variant. Chceme-li informaci sémantizovat, tj. polarizovat ji vzhledem k určité variantě, např. α , která například odpovídá správné variantě, volíme sémantizovanou informaci $I^s(p)$ ve tvaru ([2])

$$I^s(p) = \begin{cases} -I(p); & 0 \leq p \leq 0,5 \\ I(p); & 0,5 < p \leq 1 \end{cases} \quad \text{a platí} \quad \begin{matrix} I^s(0) = -1 \leq I^s(p) \leq 1 = I^s(1); \\ I^s(0,5) = 0. \end{matrix} \quad (3)$$

Neurčitost $H(p)$ je také možné definovat vztahem (4), který je vlastně hrubou aproximací (1b)

$$H(p) = 2 \cdot (1 - \max(p; 1 - p)) \text{ [lit]}. \quad (4)$$

Sémantizovanou informaci je pak možné z předchozího vztahu určit také jak je uvedeno v (3). Všechny hodnoty veličin, které byly získané pomocí (4) jsou v jednotkách lit. [3]



Obrázek 1: Závislost neurčitosti $H(p)$ podle (1b) (graf má nelineární průběh, plná linka) a $H(p)$ podle (4) (graf neurčitosti je po částech lineární v závislosti na pravděpodobnosti p , přerušovaná linka).

3. Příklad

Mějme dotazník, skládající se z k položek (otázek). Ke každé položce jsou nabídnuty varianty odpovědí; za varianty odpovědí nemusíme jen uvažovat jejich reálné konstrukce, ale například varianty „správně – špatně“, z nichž právě jedna je správná. Z odpovědí respondentů na i -tou položku stanovíme pomocí relativních četností odhady pravděpodobností (například $\hat{p}_i, 1 - \hat{p}_i$ volby každé z variant. Podle (3) pak získáme odhad $I_i^s(\hat{p}_i)$. Součet $\sum_1^k I_i^s(\hat{p}_i)$ je horním odhadem informace, kterou využil „průměrný“ respondent ke správným odpovědím. Pro lepší interpretaci je třeba tuto hodnotu normovat vydělením maximální možnou hodnotou k . Čím je hodnota $\frac{1}{k} \sum_1^k I_i^s(\hat{p}_i)$ blíží 1, tím je rozhodování jedince ve prospěch správné odpovědi jednoznačnější, záporná hodnota by znamenala velkou „neznalost“, ...

Neurčitost rozdělení $\vec{p} = \begin{pmatrix} \alpha_1 & \cdots & \alpha_n \\ p_1 & \cdots & p_n \end{pmatrix}$ na prostoru Ω n variant můžeme definovat vztahem

$$H(\vec{p}) = 1 - \frac{n}{n-1} \sum_1^n \left(p_i - \frac{1}{n}\right)^2 [\text{lit}], \quad \sum_1^n p_i = 1, \quad 0 \leq p_i \leq 1, \quad (5)$$

kde $H(1; 0; 0; \dots; 0) = 0 \leq H(\vec{p}) \leq 1 = H(\frac{1}{n}, \frac{1}{n}, \dots, \frac{1}{n})$.

Pro $n = 2$ je

$$H(p, 1-p) = 1 - 2 \cdot \left[\left(p - \frac{1}{2}\right)^2 + \left(1-p - \frac{1}{2}\right)^2 \right] = 4 \cdot p \cdot (1-p) [\text{lit}] \quad (5a)$$

Jednotkou této neurčitosti je opět neurčitost volby mezi dvěma stejně pravděpodobnými variantami, tedy 1 bit. Stupnice pro (5a) vzhledem k pravděpodobnosti p se však od Shannonovy entropie (1b) liší. (Nelze tedy jednotky z (4), resp. z (5a) převést na bity z (1), resp. (1a).) Neurčitost $H(p, 1-p)$ je pro $p \in \langle 0; 1 \rangle$ konkávní funkcí, pro $p \in (0; 0,5)$ je rostoucí, pro $p \in (0,5; 1)$ klesající, odpovídá aproximaci (1b).

Chceme-li posuzovat neurčitost z hlediska rozhodování mezi variantou (jevem) α_i a některou z variant zbývajících (jevy $\alpha_j \neq \alpha_i$), známe-li pravděpodobnost p_i , volíme pravděpodobnostní prostor pro $n = 2$ s elementárními jevy $\alpha_i, \neg\alpha_i$ a s pravděpodobnostmi $p_i, 1-p_i$. Odpovídající neurčitost $h(p_i)$ pak můžeme definovat i ve tvaru poměru (viz také Obr. 2)

$$h(p_i) = \begin{cases} \frac{p_i}{1-p_i}, & \text{když } 0 \leq p_i < 0,5, \\ \frac{1-p_i}{p_i}, & \text{když } 0,5 \leq p_i \leq 1. \end{cases} \quad (6a)$$

Pak je $0 \leq h(p_i) \leq 1$, pro $p_i \in (0; 0,5)$ rostoucí konvexní funkcí, pro $p_i \in (0,5; 1)$ klesající konvexní funkcí. Informaci, kterou přináší zpráva o tom, že jev α_i nastal, můžeme kvantifikovat hodnotou

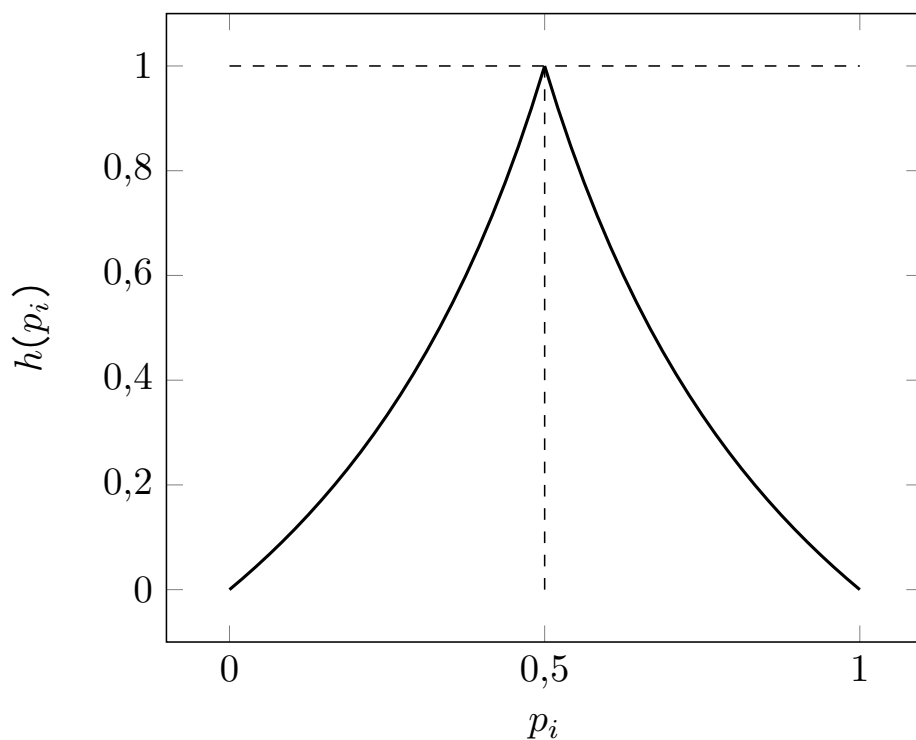
$$I(\alpha_i) = \begin{cases} 1 - \frac{h(p_i)}{2}, & \text{pro } p_i \in \langle 0; 0,5 \rangle, \\ \frac{h(p_i)}{2}, & \text{pro } p_i \in \langle 0,5; 1 \rangle. \end{cases} \quad (7a)$$

Pak je $0 \leq I(\alpha_i) \leq 1$. Přitom $I(\alpha_i) = 0$, když $p_i = 1$ a $I(\alpha_i) = 1$, když $p_i = 0$. Menším hodnotám pravděpodobnosti p_i odpovídá větší hodnota informace. (Porovnejte vztah (7a) a (3).)

Neurčitost pravděpodobnostního rozdělení $\vec{p} = \begin{pmatrix} \alpha_1 & \dots & \alpha_n \\ p_1 & \dots & p_n \end{pmatrix}$ můžeme pak definovat vztahem

$$h(\vec{p}) = \frac{1}{2} \sum_{i=1}^n h(p_i) \text{ [lit]} \quad (6b)$$

kde $h(\vec{p})$ je symetrickou funkcí pravděpodobností p_i a platí $0 = h(1; 0; 0; \dots; 0) \leq h(\vec{p}) \leq 1 = h(0,5; 0,5; 0; \dots; 0)$, ale $h(\frac{1}{n}; \frac{1}{n}; \dots; \frac{1}{n}) = \frac{n}{n-1}$ pro všechna $n \geq 2$. Tato poslední vlastnost ukazuje na nutnost jiné interpretace této neurčitosti než předpokládají předchozí vztahy pro neurčitost (1), (4), (5). Také normování výrazu (6b) jednou polovinou má vztah k základní situaci jen pro $n = 2$. Vztah (6b) pro rozdělení $\vec{p}$ se může proto doporučit pro odhad neurčitosti jen pro $n = 2$.



Obrázek 2: Neurčitost $h(p_i)$, definovaná vztahem (5) pro $n = 2$.

Pravděpodobnost správné odpovědi na určitou dotazníkovou položku závisí jak na úrovni respondenta (tu charakterizuje latentní proměnná *schopnost*), tak i na *obtížnosti* položky (respondenti stejné schopnosti reagují (odpovídají) na rozdílné položky různě, ale podobně na tutéž položku). Experimentálně se zjišťovala závislost počtu správných odpovědí v reprezentativní skupině respondentů na schopnosti (předpokládalo se, že schopnost je úměrná celkovému počtu správných odpovědí dotazníku s větším počtem položek, schopnější je ten, kdo odpoví správně na více položek). Výsledkem experimentu byla esovitá křivka závislosti relativních četností správných odpovědí na určitou položku na schopnosti jedince. Jedna z možných aproximací

této křivky byla gaussovskou distribuční funkcí (získaná stupnice schopností se pak nazvala *probitová*), druhá z aproximací je *logitová* pomocí logistické funkce (8) buď s jedním parametrem b (G. Rasch, [5, 8]) nebo logistické funkce (9) se dvěma parametry a, b (A. Birnbaum, [6, 11]):

$$Ra(x) = p(x) = \frac{1}{1 + e^{-(x-b)}}; \quad b, x \in (-\infty; +\infty) \quad (8)$$

$$Bir(x) = p(x) = \frac{1}{1 + e^{-a(x-b)}}; \quad b, x \in (-\infty; +\infty); \quad a > 0. \quad (9)$$

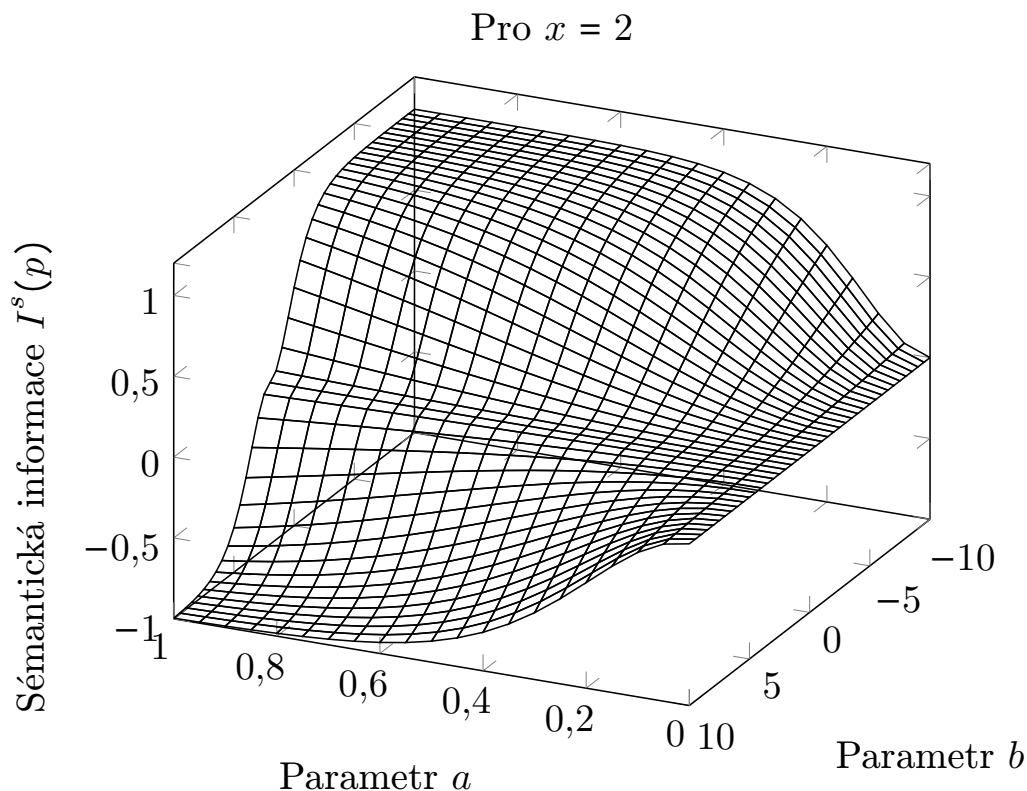
Parametr b se interpretuje jako obtížnost dotazníkové položky, parametr a jako její diskriminační schopnost. Odhady uvedených parametrů se realizují z výsledků odpovědí na dotazník na základě logistické regrese jak jsme již naznačili.

Dosadíme-li nyní pravé strany z výrazů (8) nebo (9) do některého z předchozích vztahů za pravděpodobnost správné odpovědi p_i , dostaneme hlubší souvislosti mezi neurčitostí (případně informací), i -tou dotazníkovou položkou a schopností jedince. (Podrobněji v [2].) Ještě připomeňme, že z psychologického hlediska není ještě daný problém vyřešen. Schopnost není jednorozměrná veličina, závisí ještě na obsahu položek (ten jsme ale neuvažovali). Aby bylo možné vypočítané veličiny (entropii a informaci) správně interpretovat, musí být dotazníkové položky homogenizovány vzhledem ke svému obsahu (k tomu může posloužit například faktorová analýza, která položky roztrídí z hlediska obsahu). Další důležitou veličinou pro analýzu dotazníku je čas, který k odpovědi respondent má. Víme, že existují jisté zlomové okamžiky, které mají na reakci respondenta podstatný vliv, pravděpodobnost správné odpovědi není vždy spojitou funkcí schopností či času (rozvinutější úvahy o tom jsou například v [7]).

Následující Obr. 4 ukazuje dvourozměrný graf sémantické informace dotazníkové položky při hodnotě parametru $b = 0$ a v závislosti na úrovni respondenta x při různých hodnotách parametru a . Závislost není lineární, ale s růstem úrovně respondenta x a při konstantním parametru b roste sémantická informace.

Obtížnost i -té položky se v klasické testové teorii určuje jako $o_{1i} = 1 - p_i$. Podle Birnbauma ji ale musíme uvažovat jako funkci položkových parametrů a, b a úrovně schopnosti jedince x . Sémantizovaná informace nám také umožňuje definovat obtížnost výrazem

$$o_{2i} = \begin{cases} 1 - H(p)/2 & \text{pro } 0 \leq p \leq 0,5, \\ H(p)/2 & \text{pro } 0,5 \leq p \leq 1. \end{cases} \quad (10)$$



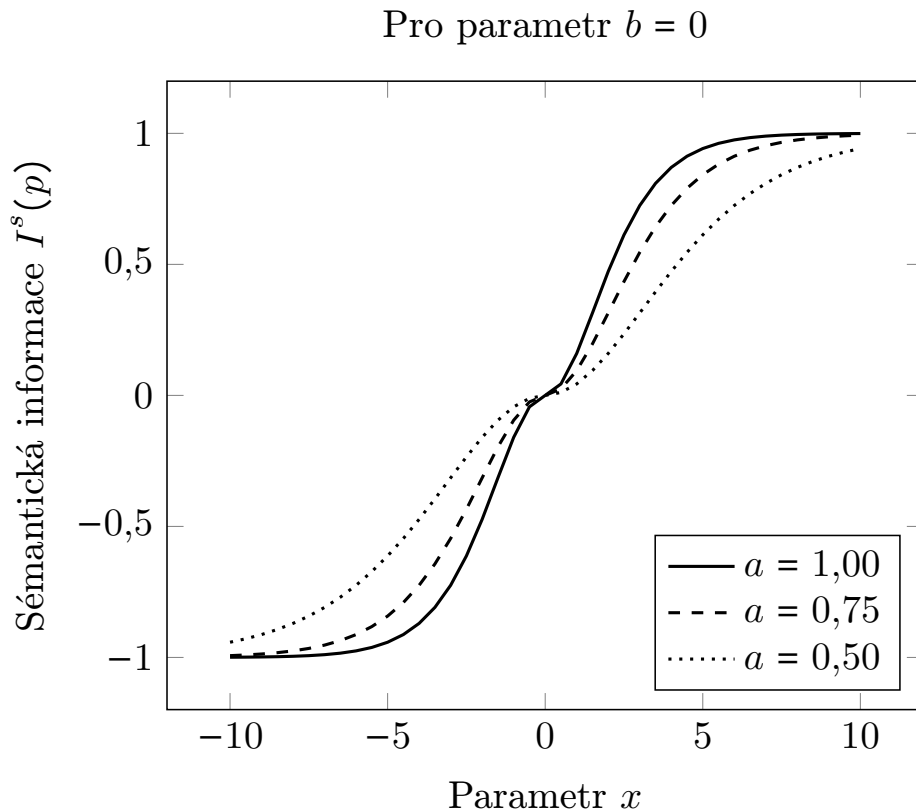
Obrázek 3: Graf sémantické informace položky, odvozený z (3) dosazením za p z pravé strany vztahu (9) pro $x = 2$ v závislosti na parametrech a, b .

Do vztahu (10) je možno za $H(p)$ dosazovat buď výraz z (1b), (4) nebo (5a). Pokaždé ovšem dostaneme jiné hodnoty, platí však ve všech případech, že s rostoucím p klesá obtížnost, měřená na stupnici od 0 do 1.

Viděli jsme, že je více možností jak odhadovat prostřednictvím pravděpodobnosti množství informace, potřebné k určitému rozhodnutí a jaká je případně jeho obtížnost. V následujícím odstavci si ukážeme jiné metodické východisko k podobným odhadům.

4. Neurčitost modelovaná fuzzy množinami a intuicionistickými fuzzy množinami

Ohodnocení volby určité varianty pravděpodobností je někdy velmi obtížné nebo není vůbec možné. Například rozhodnutí, zda je pacient vyléčený nebo nevy léčený záleží jen na úvaze odborníka, podobně rozhodování o tom, že určitý žák je nadaný nebo nenadaný se nedá statisticky rozhodnout (i když se takové pokusy realizují). Místo objektivizovaných statistických šetření se musí uskutečňovat expertní odhady. I v situacích, kdy nelze od sebe přesně „oddělit“ některé varianty se uplatňuje teorie fuzzy množin. [10] Fuzzy



Obrázek 4: Dvourozměrný graf sémantické informace dotazníkové položky ze vztahu (3) po dosazení za p z pravé strany výrazu (9) při hodnotě parametru $b = 0$ a v závislosti na úrovni respondenta x při různých hodnotách parametru a : plná linka je pro parametr $a = 1$, přerušovaná linka je pro parametr $a = 0,75$ a tečkovaná linka je pro parametr $a = 0,5$.

množina reprezentuje sémantiku pojmu rozhodnutí. Odpovídá na to, jak „silně“ jsou v rozhodnutí (které je fuzzy množinou) zastoupeny určité prvky ze souboru Ω .

Uvažujme pouze konečnou množinu prvků (možností, vlastností) $\Omega = \{\alpha_1, \alpha_2, \dots, \alpha_n\}$. Jestliže každému prvku α_i přiřadíme míru $\mu_i = \mu(\alpha_i)$, $\mu_i \in \langle 0; 1 \rangle$, vyjadřující, že α_i je součástí rozhodnutí $\underline{A}$, je $\underline{A}$ fuzzy množinou nad Ω . Čím bližší je hodnota μ_i jedné, tím spíše je v rozhodnutí $\underline{A}$ prvek α_i , hodnota $\mu_i = 0$ znamená, že v rozhodnutí $\underline{A}$ prvek α_i není. Fuzzy množinu můžeme zapisovat ve tvaru ([10])

$$\underline{A} = \{\alpha_1/\mu_1, \alpha_2/\mu_2, \dots, \alpha_n/\mu_n\}. \quad (11)$$

Hodnoty μ_i nevznikají obecně jako pravděpodobnosti, představují expertní rozhodnutí o míře zastoupení α_i v rozhodnutí $\underline{A}$. Proto nemusí obecně

platit, že $\sum \mu_i = 1$. (Pro další úvahy prvek α_i , jehož $\mu_i = 0$ do seznamu Ω zařazovat nebudeme.)

Neurčitost, o které budeme zde uvažovat, má svůj původ v nejistotě odhadu, které prvky α_i do $\underline{A}$ patří. Soustředíme se nyní na určitou vlastnost $\alpha \in \Omega$. Jaká je neurčitost zastoupení $\alpha \in \Omega$ ve fuzzy množině $\underline{A}$?

Je možné vyjít z modelu rozhodování charakterizovaném fuzzy množinou $\underline{B}$, kde $\beta = \Omega - \{\alpha\}$

$$\underline{B} = \{\alpha/\mu(\alpha); \beta/\mu(\beta)\}, \quad \mu(\alpha) + \mu(\beta) \leq 1 \quad (12)$$

kde míry věrohodnosti $\mu(\alpha), \mu(\beta) \in \langle 0, 1 \rangle$ jsou také subjektivním nebo objektivizovaným odhadem volby příslušné varianty. Neurčitost pro fuzzy množinu (12) můžeme při nezávislých volbách variant odhadnout ze vztahu (13) (viz [2])

$$\begin{aligned} H(\underline{B}) &= 2 - [\max\{\mu(\alpha), 1 - \mu(\alpha)\} + \max\{\mu(\beta), 1 - \mu(\beta)\}] \\ &= \min\{\mu(\alpha), 1 - \mu(\alpha)\} + \min\{\mu(\beta), 1 - \mu(\beta)\} \text{ [lit]}; \\ &0 \leq H(\underline{B}) \leq 1. \end{aligned} \quad (13)$$

Informaci $I(\alpha)$ pro volbu varianty α i její sémantizovanou hodnotu $I^s(\alpha)$ určíme podobně jako pro klasickou pravděpodobnostní neurčitost $H(p)$ z následujících vztahů:

$$I(\alpha) = 1 - H(\underline{B}); \quad I^s(\alpha) = \begin{cases} -I(\alpha); & 0 \leq \mu(\alpha) \leq 0,5, \\ I(\alpha); & 0,5 \leq \mu(\alpha) \leq 1. \end{cases} \quad (14)$$

Pak ovšem je $-1 \leq I^s(\alpha) \leq 1$. Zápornou hodnotu informace můžeme interpretovat jako míru dezinformace nebo pomýlenosti apod. vzhledem k volbě varianty α . Jednotky informace odpovídají jednotkám neurčitosti ([2]).

Pro fuzzy množinu $\underline{B}$ nemusí obecně platit následující vztah (15), který odpovídá základní vlastnosti pravděpodobnosti pro dvě vzájemně se vylučující varianty α, β (viz [2, 4])

$$\mu(\alpha) + \mu(\beta) = 1 \quad (15)$$

Protože jsou však v našem modelu volby variant na sobě závislé (popíšeme situaci volby jen právě jedné z variant), musíme předpokládat platnost vztahu (15). Pak neurčitost pro fuzzy množinu (12) za podmínky (15) můžeme odhadnout např. ze vztahu (16), který odpovídá (1a):

$$H(\underline{B}) = -\mu(\alpha) \cdot \log_2 \mu(\alpha) - (1 - \mu(\alpha)) \log_2 (1 - \mu(\alpha)) \text{ [bit]} \quad (16)$$

Pak je také $0 \leq H(\underline{B}) \leq 1$ a informaci i její sémantizovanou podobu můžeme stanovit podobně jako v předchozím případě.

Máme však více možností pro odhad neurčitosti $H(\underline{B})$ tak, aby byly splněny intuitivní odpovídající podmínkám (2) a (15) (viz také [2]). Jednoduchý vztah pro neurčitost $H(\underline{B})$, předpokládáme-li platnost podmínky (15), můžeme vyjádřit i ve tvaru (porovnejte se vztahem (4)):

$$H(\underline{B}) = 2 \cdot \min\{\mu(\alpha), 1 - \mu(\alpha)\} \text{ [lit]}; \quad 0 \leq H(\underline{B}) \leq 1. \quad (17)$$

Poznámky:

1. Analogicky s neurčitostí, definovanou pro diskrétní rozdělení podle (5), je možné určit i neurčitost fuzzy množiny (11) s odpovídající interpretací ve tvaru $H(\underline{A}) = 1 - \sqrt{\frac{4}{n} \sum_1^n (\mu_i - 0,5)^2}$ [lit], $0 \leq H(\underline{A}) \leq 1$, $H(\underline{A}) = 0$ když pro každé i je μ_i buď rovno 0 nebo 1, $H(\underline{A}) = 1$ když všechna $\mu_i = 0,5$.

2. V práci [2] jsou uvedeny další možnosti volby definic neurčitosti fuzzy množin. Jeden z příkladů vychází z definice funkce $h(p_i)$ podle (6a). Pro fuzzy množinu (11) definujeme neurčitost $H(\underline{A})$ následujícím vztahem

$$H(\underline{A}) = \frac{1}{n} \sum_1^n h(\mu_i) \text{ [lit]}.$$

Pak je zřejmě $0 \leq H(\underline{A}) \leq 1$. Nejmenší hodnoty $H(\underline{A})$ dosahuje, když všechna $\mu_i = 0$ nebo 1 a největší hodnoty, když všechna $\mu_i = 0,5$. Má i další vhodné vlastnosti. ([2])

3. Shannonova metoda posuzování neurčitosti se může uplatnit i pro fuzzy množiny, jen je třeba míry věrohodnosti $\mu(\alpha_i)$ pro fuzzy množinu $\underline{A}$ normovat. Odpovídající míra neurčitosti je pak tvaru

$$H(\underline{A}) = -\frac{1}{n} \sum_1^n \frac{\mu_i}{\sum_1^n \mu_i} \log_2 \frac{\mu_i}{\sum_1^n \mu_i} \text{ [bit]}, \quad 0 \leq H(\underline{A}) \leq 1.$$

Interpretace předchozí neurčitosti je ale obtížnější než ta, která odpovídá vztahu založeném na pravděpodobnostním rozdělení. (Mezi pravděpodobnostním rozdělením a fuzzy množinou je často velký rozdíl i v interpretaci zdánlivě odpovídajících vztahů. Předchozí vztah lze jako míru neurčitosti s dobrou interpretací užít jen pro fuzzy množinu $\underline{B}$.)

Chceme-li přizpůsobit model nejjednoduššího rozhodování i situaci, kdy nemusí být v tomto procesu vybrána žádná z nabídnutých možností (když např. rozhodovatel (respondent) má odpor vůči volbě některé z variant nebo neví jak se má rozhodnout), pak je vhodné předpokládat, že

$$\mu(\alpha) + \mu(\beta) \leq 1. \quad (18)$$

V tom případě lze pro charakterizaci rozhodování mezi dvěma variantami α, β využít intuicionistickou fuzzy množinu (IFS) $\mathcal{F}$ v následujícím tvaru ([1])

$$\mathcal{F} = \{\alpha/(\mu(\alpha), \nu(\alpha)); \beta/(\mu(\beta), \nu(\beta))\} \quad (1)$$

kde $\mu(\alpha), \nu(\alpha) \in \langle 0; 1 \rangle$, také $\mu(\beta), \nu(\beta) \in \langle 0; 1 \rangle$ a $\mu(\alpha) + \nu(\alpha) \leq 1$, $\mu(\beta) + \nu(\beta) \leq 1$.

Rozhodování o variantě α (a odpovídajícím způsobem i o variantě β) je tak rozděleno do 3 hledisek:

- a) posuzující míru přijetí varianty α (odhadováno $\mu(\alpha)$),
- b) posuzující míru nepřijetí varianty α (odhadováno $\nu(\alpha)$),
- c) posuzující míru nerozhodnosti pro volbu varianty α (určeno hodnotou $1 - \mu(\alpha) - \nu(\alpha)$).

Fuzzy množiny, odpovídající uvedeným typům rozhodování jsou

$$\begin{aligned} \underline{A}_1(\alpha) &= \{\alpha/\mu(\alpha); \neg\alpha/1 - \mu(\alpha)\}; \\ \underline{A}_2(\alpha) &= \{\alpha/\nu(\alpha); \neg\alpha/1 - \nu(\alpha)\}; \\ \underline{A}_3(\alpha) &= \{\alpha/\pi(\alpha); \neg\alpha/1 - \pi(\alpha)\}, \\ &\text{kde } \pi(\alpha) = 1 - \mu(\alpha) - \nu(\alpha) \end{aligned}$$

a jejich neurčitosti $H(\underline{A}_1(\alpha))$, $H(\underline{A}_2(\alpha))$, $H(\underline{A}_3(\alpha))$ určíme podle (16) nebo (17).

Pro IFS $\mathcal{F}_z$ (19) lze však také zavést neurčitost $H(\alpha)$, týkající se varianty α vztahem

$$H(\alpha) = -(\mu(\alpha) \log_2 \mu(\alpha) + \nu(\alpha) \log_2 \nu(\alpha) + \pi(\alpha) \log_2 \pi(\alpha)) \text{ [bit]} \quad (20)$$

a je pak

$$0 \leq H(\alpha) \leq \log_2 3. \quad (21)$$

Pro neurčitost $H(\alpha)$ a neurčitosti $H(\underline{A}_1(\alpha))$, $H(\underline{A}_2(\alpha))$, $H(\underline{A}_3(\alpha))$ platí nerovnosti

$$H(\underline{A}_1(\alpha)) \leq H(\alpha); \quad H(\underline{A}_2(\alpha)) \leq H(\alpha); \quad H(\underline{A}_3(\alpha)) \leq H(\alpha). \quad (22)$$

(To vyplývá z následující posloupnosti nerovností: $\log_2(\nu + \pi) \geq \log_2 \nu$; $-\log_2(\nu + \pi) \leq -\log_2 \nu$; $-\nu \log_2(\nu + \pi) \leq -\nu \log_2 \nu$.)

Nerovnost (21) nám dává možnost určit vzhledem k variantě α pro IFS $\mathcal{F}$ informaci $I(\mathcal{F})$ vztahem (23a)

$$I_\alpha(\mathcal{F}) = \log_2 3 - H(\alpha) \quad (23a)$$

nebo její normovanou podobu ve tvaru

$$I_\alpha(\mathcal{F})_{\text{norm}} = 1 - H(\alpha)/\log_2 3 \quad (23b)$$

a pak i příslušnou její sémantizaci vztahem

$$I_\alpha(\mathcal{F})_{\text{norm}}^s = \begin{cases} -I_\alpha(\mathcal{F})_{\text{norm}}; & 0 \leq \mu(\alpha) \leq 0,5, \\ I_\alpha(\mathcal{F})_{\text{norm}}; & 0,5 < \mu(\alpha) \leq 1. \end{cases} \quad (24)$$

kde je $-1 \leq I_\alpha(\mathcal{F})_{\text{norm}}^s \leq 1$.

Při volbě právě jedné z možných variant α, β v našem modelu rozhodování popsaném IFS $\mathcal{F}$ musí, krom původních podmínek, také platit (18), proto

$$\nu(\alpha) + \nu(\beta) \leq 1. \quad (25)$$

Neurčitost pro IFS $\mathcal{F}$ pak můžeme za všech uvedených podmínek definovat vztahem

$$H(\mathcal{F})(\alpha) = H(\underline{A}_1(\alpha)) + H(\underline{A}_2(\alpha)) + H(\underline{A}_3(\alpha)) \quad (26)$$

a platí $0 \leq H(\mathcal{F})(\alpha) \leq 3$. Proto můžeme definovat pro IFS $\mathcal{F}$ i informaci $I(\mathcal{F})(\alpha)$ vztahem

$$I(\mathcal{F})(\alpha) = 3 - H(\mathcal{F})(\alpha)$$

nebo jako normovanou informaci

$$I(\mathcal{F})_{\text{norm}}(\alpha) = 1 - H(\mathcal{F})(\alpha)/3.$$

Uvedenou normovanou informaci můžeme také sémantizovat vzhledem k variantě α , podobně jako jsme to udělali v předchozích případech. Volíme

$$I(\mathcal{F})^s(\alpha) = \begin{cases} -I(\mathcal{F})_{\text{norm}}(\alpha); & 0 \leq \mu(\alpha) \leq 0,5, \\ I(\mathcal{F})_{\text{norm}}(\alpha); & 0,5 \leq \mu(\alpha) \leq 1, \end{cases} \quad -1 \leq I(\mathcal{F})^s(\alpha) \leq 1. \quad (27)$$

Pro variantu β můžeme odvodit odpovídající vztahy pro informaci obdobně.

Další otázkou je, jak variabilita odhadu hodnot příslušných věrohodností může ovlivnit hodnoty sémantizovaných informací. Ukážeme možný odhad chyby pro neurčitost $H(\mathcal{F}, \alpha)$, když $\mu(\alpha) = \bar{\mu}(\alpha) \pm \Delta\mu(\alpha)$, kde $\bar{\mu}(\alpha)$ je bodový odhad hodnoty $\mu(\alpha)$ z měření (např. aritmetickým průměrem z experimentálních dat) a $\Delta\mu(\alpha)$ je odhad chyby měření (např. směrodatnou odchylkou z měření):

$$\Delta H(\mathcal{F}, \alpha) \sim \frac{dH}{d\mu} \cdot \Delta\mu(\alpha) = \left(\log_2 \frac{1 - \mu(\alpha)}{\mu(\alpha)} \right) \cdot \Delta\mu(\alpha).$$

5. Závěr

Hodnoty sémantizovaných informací jsou stanovovány v závislosti na možnosti získání vstupních informací. V našem případě jde buď o odhad pravděpodobnosti p (subjektivní nebo objektivizovaný) nebo měř věrohodnosti μ . Interpretace získaných numerických hodnot je pak dána výchozími předpoklady metody i zkušeností experimentátora v určité oblasti diagnóz. Metoda se hodí *pro porovnávání* hodnot sémantizované informace různých diagnóz. Nepracuje se zde s apriorně danými a zcela exaktně definovanými jednotkami ani procedurami určení neurčitosti. Nemáme totiž v humanitních vědách jiné možnosti. Záleží pak na zkušenosti experimentátora, jakou z metod odhadu a i příslušného výpočtu bude volit a jak se naučí číselné hodnoty určené z volených vzorců interpretovat. Že lze přitom dojít i k dostatečně přesným výsledkům jsme naznačili v práci [9].

Článek ukazuje, jak je možné z kvalitativních informací zkonstruovat kvantitativně posuzovatelnou jistou vlastnost informace.

Literatura

- [1] Atanassov, K. T., *Intuitionistic Fuzzy Sets*, Fuzzy Sets and Systems 20 (1986), 87–96. *cit. 16*
- [2] Půlpán, Z., *Odhad informace z dat vágní povahy*, Academia, Praha 2012, 200 s. *cit. 7, 11, 14 a 15*
- [3] Půlpán, Z., *Mezurado bazita sur la nocio de entropio*, grkg/Humankybernetik, Band 51, Heft 2, 2010, 70–79. *cit. 6, 7*
- [4] Novák, V.: *Fuzzy Sets and Their Applications*, Adam Hilger, Bristol and Philadelphia, SNTL Praha 1986, 248 s. *cit. 14*

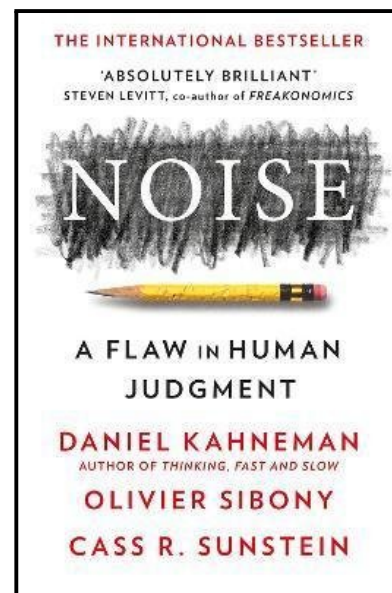
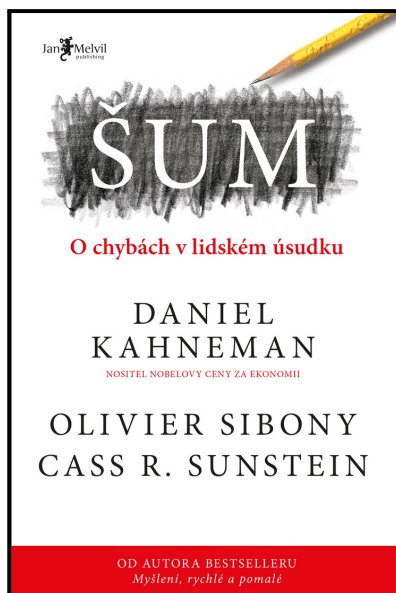
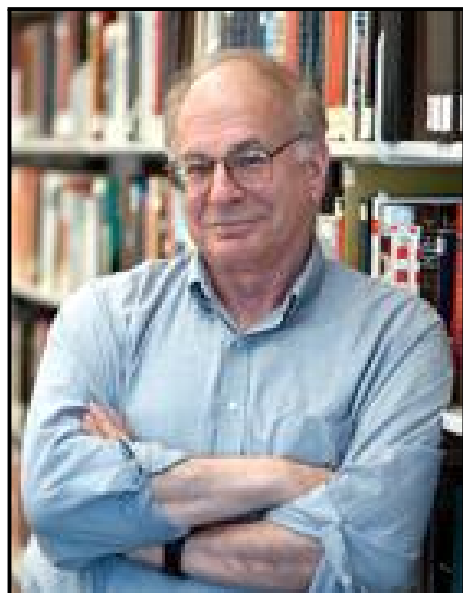
- [5] Rasch, G.: *Probabilistic models for some intelligence and attainment tests*, Copenhagen, Danish Institute for Educational Research, expanded edition, The University of Chicago Press, 1980 (1960–1980). *cit. 11*
- [6] Birnbaum, A: Some latent trait models and their use in inferring an examinee's ability, in Lord F. N. & Novick, M. R. (eds): *Statistical theories of mental test scores*, Reading, MA, Addison-Wesley (1968). *cit. 11*
- [7] Půlpán, Z.: *K problematice hledání podstatného v humanitních vědách*, Academia Praha, Studie 1/2001, 135 s. *cit. 11*
- [8] Fisher, G. H., Molenaar, I. W.: *Rasch models: foundations, recent developments and applications*, Springer-Verlag, New York 1995. *cit. 11*
- [9] Půlpán, Z., Jahodová Berková, A.: Hodnověrnost a intuitivně odhadovaná pravděpodobnost, *Informační bulletin České statistické společnosti*, roč. 30, č. 1, březen 2019. *cit. 18*
- [10] Lord, F. M.: *Applications of Item Response Theory to Practical Testing Problems*, Hillsdale, NJ, Erlbaum 1980. *cit. 12, 13*
- [11] Birnbaum, A.: *Statistical theory for the logistic mental tests models with a prior distribution of ability*, J. Math. Psychol. 6, 1969. *cit. 11*

KNIŽNÍ RECENZE: ŠUM

BOOK REVIEW: NOISE

Ondřej Vencálek

E-mail: ondrej.vencalek@upol.cz



Nová kniha Daniela Kahnemana a jeho dvou kolegů Oliviera Sibonyho a Casse Sunsteina nazvaná *Šum: o chybách v lidském úsudku* vyšla v roce 2021 a ve velmi krátké době jsme se dočkali i jejího českého překladu vydaného v nakladatelství Jan Melvil Publishing v Brně. Kniha má 407 stran. Recenzi napsal Ondřej Vencálek v únoru 2022.

Než jsem se dal do čtení tohoto nového Kahnemanova počinu, měl jsem strach, že nová kniha bude jen variací na téma slavné publikace *Myšlení: rychlé a pomalé*. Tato obava se však rozplynula hned po přečtení úvodu knihy. V něm pomocí známého příkladu¹ se čtyřmi terči autoři vysvětlují pojmy *vychýlení (bias)* a *šum*, tedy *variabilita* (měrou šumu může být třeba směrodatná odchylka). Kniha *Myšlení: rychlé a pomalé* je především o různých zkresleních, kterých se dopouštíme v našich úsudcích. Nová kniha je, jak už sám její název napovídá, věnována šumu. Jde přitom opět o úsudky, tedy o šum v našich úsudcích. Je třeba vysvětlit, co je míněno pod pojmem úsudek. Autoři tento termín chápou zejména ve významu hodnotící úsudek (např. ohodnocení trestného činu určitou sazbou trestu nebo ohodnocení studentské eseje známkou) nebo prediktivní úsudek (zahrnující nejrozumnější prognózy a odhady).

¹https://en.wikipedia.org/wiki/Bias%E2%80%93variance_tradeoff#Motivation

Tato nová Kahnemanova kniha, stejně jako ta předešlá, je jako koruna mohutného stromu. Je tvořena šesti velkými větvemi – šesti částmi, z nichž vyrůstají menší větve – jednotlivé kapitoly, kterých je celkem 28 (namátkou: kapitola 1 – *Šum v udílení trestu*, kapitola 8 – *Jak skupiny zesilují šum*, kapitola 28 – *Pravidla, nebo zásady?*). Každá kapitola-větev je ještě rozvětvena do dílčích nečíslovaných podkapitol, k listům pak můžeme přirovnat jednotlivé odstavce. Toulky touto korunou mi jako čtenáři přinesly mnoho potěšení, ale, přiznávám, taky jeden problém. Tím je příliš velké množství příliš zajímavých „listů“. Jedno přečtení knihy mi tak sice stačilo k základnímu pochopení, bohužel však ne k zapamatování všech zajímavých myšlenek a příkladů.

K některým kapitolám knihy se budu zřejmě opakovaně vracet, protože se týkají témat, o kterých (opakovaně) přemýšlím. Kapitola 12 nazvaná *V údolí normalnosti* je brilantní úvahou o „kauzálním myšlení“. Na konkrétním příkladu zde například autoři ilustrují, že „Je-li kauzalita možná, naše mysl promptně promění jakkoliv nízkou korelaci v kauzální a vysvětlující faktor.“ A uvádí nechtěné důsledky: „Kauzální myšlení a iluze chápání minulosti však přispívají k nadměrné sebedůvěře v předpovědích budoucnosti“. Záložku jsem si ponechal také u kapitolky nazvané *Kognitivní styl* (část kapitoly 18), kde autoři uvádí známé příklady testu tzv. *kognitivní reflexe*, který má odhalit, nakolik je člověk schopen „překonat první (nesprávnou) odpověď, která mu přijde na mysl“. Otázku „Když běžíte závod a předběhnete závodníka na druhém místě, na jakém místě budete?“ zařazují od nynějška do svých popularizačních přednášek. Vracet se budu i k závěrečným kapitolám knihy, které mě nutí k hlubšímu zamyšlení. A jsem docela zvědav na ohlasy této (závěrečné) části knihy.

Kniha *Šum: o chybách v lidském úsudku* rozhodně stojí za přečtení. Pro ty, kteří nemají čas či sílu pročíst celé více než 300 stránkové dílo, lze vřele doporučit alespoň *Závěr a shrnutí* v rozsahu 14 stran, kde autoři přehledně a stručně celou knihu shrnují.

SLAVNOSTNÍ SEMINÁŘ HONORARY SEMINAR

Redakce časopisu

Dílo Otakara Borůvky v kontextu současné teorie

Srdečně Vás zveme na slavnostní seminář o významu Otakara Borůvky pro současnou teorii grafů a jeho řešení minimálního problému pro světovou matematiku, který proběhne dne 30. května 2022.

Seminář pořádá Masarykova univerzita a Společnost Otakara Borůvky u příležitosti 95. výročí vyřešení minimálního problému. Uskuteční se v univerzitním kině Scala v Brně (Moravské nám. 127/3), je bezplatný a je určen především odborníkům z praxe, studentům, ale i zástupcům státní a veřejné správy.

V případě, že se rozhodnete semináře zúčastnit, prosíme o potvrzení účasti do 25. května 2022 na e-mail info@otakarboruvka.cz.



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 $\text{\TeX}$, ve formátu Lua $\text{\HB\TeX}$ s písmy balíku $\text{\Csf}$ fonts.

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: 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 MIČÁLEK, CSc., prof. Ing. Jiří MILITKÝ, CSc., doc. Ing. Iveta STANKOVIČOVÁ, PhD., Mgr. Ondřej VENCÁLEK, Ph.D.

Redaktor časopisu: Mgr. Ondřej VENCÁLEK, Ph.D., ondrej.vencalek@upol.cz.
Informace pro autory jsou na stránkách společnosti, <http://www.statapol.cz/>.

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

Toto číslo bylo vtištěno s laskavou podporou Českého statistického úřadu.