

INFORMAČNÍ BULLETIN



České statistické společnosti

Ročník 32, číslo 4, prosinec 2021

Obsah

Vědecké a odborné články

Ondřej Vencálek

Úloha o rozdělení sázky – tradiční a bayesovský pohled 3

Zprávy a informace

Jan Kalina

Knižní recenze: Myšlení rychlé a pomalé 19

Jiří Rybička

Mezinárodní konference v projektu IOSCS 21

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 $\text{\LuaHB\TeX}$ s písmy balíku $\text{\Csf}$ onts.

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.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.

ÚLOHA O ROZDĚLENÍ SÁZKY – TRADIČNÍ A BAYESOVSKÝ POHLED

PROBLEM OF DIVISION OF THE STAKES FROM TRADITIONAL AND BAYESIAN POINT OF VIEW

Ondřej Vencálek

Adresa: Katedra matematické analýzy a aplikací matematiky, Přírodovědecká fakulta Univerzity Palackého v Olomouci

E-mail: ondrej.vencalek@upol.cz

Abstrakt: Úloha o rozdělení sázky má důležité místo v historii rozvoje teorie pravděpodobnosti. Její klasickou podobu a řešení, které v 17. století navrhl Pascal a Fermat, zná současný český čtenář například z knihy J. Anděla Matematika náhody. Náš příspěvek se zabývá jak touto klasickou podobou, tak i situací, kdy je pravděpodobnost vítězství jednotlivých hráčů v dílčích partiích neznámá. V této situaci navrhneme úlohu řešit bayesovsky, jak to v 18. století navrhl Laplace.

Klíčová slova: bayesovský, pravděpodobnost, rozdělení sázky.

Abstract: The problem of division of the stakes has an important place in the history of the development of probability theory. Its classic form and solution, suggested by Pascal and Fermat in the 17th century, are probably known to the contemporary Czech reader from J. Anděl's book Mathematics of Chance. Our contribution deals with both this classic form and the situation when the probability of victory of individual players in sub-games is unknown. In this situation, we propose a Bayesian solution, as Laplace suggested in the 18th century.

Keywords: bayesian, probability, division of the stakes.

1. Úloha o rozdělení sázky a její role v rozvoji teorie pravděpodobnosti

Klasická úloha o rozdělení sázky zní takto: Dva hráči hrají hru, jejímž vítězem se stane ten, který dosáhne určitého předem pevně stanoveného počtu bodů v sérii partií, kde vítězství v jedné partii přinese jeden bod. Šance na získání bodu je přitom pro oba hráče stejná. Výsledky jednotlivých partií považujeme za nezávislé. Hra je přerušena v situaci, kdy jeden z hráčů vede, a už v ní nelze pokračovat. Jak spravedlivě rozdělit částku, o kterou se hraje?

1.1. Krátká historie a význam úlohy o rozdělení sázky

Této celkem nenápadné úloze, známé v anglicky psané literatuře jako „problem of points“ či „problem of division of the stakes“, ve skutečnosti náleží důležité místo v historii rozvoje teorie pravděpodobnosti. V polovině 17. století tuto úlohu úspěšně vyřešili Blaise Pascal a Pierre de Fermat. Jejich vzájemná korespondence ohledně úlohy o rozdělení sázky – jedna ze slavných korespondencí v historii matematiky – se odehrála v roce 1654. Více o této slavné historii viz např. [4], [8] či [9]. Zájemcům doporučujeme práci Karla Mačáka *Počátky počtu pravděpodobnosti* [7], obsahující mimo jiné komentovaný překlad *Pojednání o aritmetickém trojúhelníku* (*Traité du triangle arithmétique*), které Pascal sepsal právě v roce 1654 (vydáno bylo v roce 1665). V roce 1994 připomněl slavnou historii na stránkách *Informačního bulletinu ČStS* Jan Coufal [3].

Záludnost úlohy spočívá v tom, že není zřejmé, jak správně popsat základní množinu všech možností a specifikovat pravděpodobnosti těchto možností, neboli, jak správně zvolit pravděpodobnostní prostor. Úloha se tedy týká samotných základů teorie pravděpodobnosti. Proto není překvapivé, že úlohu o rozdělení sázky zmiňuje např. Anděl ve své knize *Matematika náhody* [1] hned v první kapitole věnované základním pojmům teorie pravděpodobnosti.

Méně často se zmiňuje, že přibližně sto let po Fermatovi s Pascalem úlohu úspěšně řešil Laplace, viz [6]. Jeho přístup byl přitom bayesovský. Náš článek má ambici tento přístup k řešení slavné úlohy představit českému čtenáři. Nevycházíme přitom z Laplaceova textu, ale z vlastních úvah a výpočtů, které ostatně nejsou příliš složité.

1.2. Klasické řešení úlohy o rozdělení sázky

Řešení úlohy¹ o rozdělení sázky vypadá následovně: Označme písmenem h počet bodů, který je nutný pro celkové vítězství. O vítězi rozhodne nejvýše $2h - 1$ partií (označme tento počet symbolem n). Uvažujeme všechny možné výsledky těchto n partií, byť se to může zdát zbytečné, protože možná ani tolik partií k rozhodnutí potřeba nebude. Pro jednoduchost si představme případ, kdy $h = 3$. Bude potřeba odehrát nejvýše $n = 2h - 1 = 5$ partií. Průběh všech pěti partií můžeme popsat pětici písmen složenou ze znaků A a B , kde A značí výhru prvního hráče a B značí výhru druhého hráče. Např. pětice $AAAAA$ znamená, že všech pět partií vyhrál první hráč, pětice $ABBAB$ odpovídá situaci, kdy první hráč vyhrál v první a čtvrté partii a druhý hráč ve

¹Toto řešení je připisováno Fermatovi. Pascalovo řešení uvedeme ve čtvrté kapitole.

všech ostatních. Snadno nahlédneme, že všech možných průběhů je 2^n . Tento přístup, kdy uvažujeme výsledky všech n partií, má výhodu v tom, že při rovnosti šancí na vítězství v jednotlivých partiích mají všechny možné průběhy stejnou pravděpodobnost. Výpočet pravděpodobnosti vítězství jednotlivých hráčů podmíněné určitým průběžným stavem je pak snadným cvičením.

Anděl ve své knize *Matematika náhody* uvádí situaci, kdy k celkové výhře je třeba vyhrát šest partií, tedy $h = 6$, a hra byla přerušena za stavu 5:3. Uvádí také seznam řešitelů, kteří si před Pascalem a Fermatem na úloze takřikajíc vylámali zuby – dospěli k nesprávnému řešení. Uveďme nyní správné řešení.

Z $n = 2h - 1 = 11$ partií je již $5 + 3 = 8$ odehráno a zbývá dohrát tři partie (pokud budeme ve hře pokračovat, i kdyby již bylo rozhodnuto o celkovém vítězství). Možných (stejně pravděpodobných) scénářů dalšího vývoje je celkem $2^3 = 8$. Jenom jediný z nich odpovídá porážce průběžně vedoucího hráče. Je to scénář, v němž všechny tři zbývající hry vedoucí hráč prohraje. Sázka by měla být rozdělena v poměru odpovídajícím šanci jednotlivých hráčů na výhru, tedy 7:1 (šance na výhru je 7:1 z pohledu vedoucího hráče, a 1:7 z pohledu jeho soupeře).

Uveďme nyní obecné řešení. Označme z_1 počet partií, které do celkového vítězství chybí prvnímu hráči, a z_2 počet partií, které k celkovému vítězství chybí jeho soupeři. Označme dále maximální počet partií, které zbývají k rozhodnutí o celkovém vítězi, symbolem n_z (index z připomíná, že jde o zbývající partie). Tedy $n_z = z_1 + z_2 - 1$.

Celkové vítězství prvního, resp. druhého hráče, budeme označovat symboly W_1 , resp. W_2 (W jako *Win*, tedy výhra). První hráč celkově zvítězí, když jeho soupeř zaznamená nejvýše $z_2 - 1$ úspěchů (potřebuje jich totiž z_2). Pravděpodobnost, že druhý hráč vyhraje právě k z n_z partií je $\binom{n_z}{k}/2^{n_z}$. Pravděpodobnost celkového vítězství prvního hráče je tudíž

$$P(W_1) = \sum_{k=0}^{z_2-1} \frac{\binom{n_z}{k}}{2^{n_z}}.$$

Pravděpodobnost výhry druhého hráče se vypočte analogicky – ve vzorci stačí nahradit symbol z_2 symbolem z_1 . Tedy

$$P(W_2) = \sum_{k=0}^{z_1-1} \frac{\binom{n_z}{k}}{2^{n_z}}.$$

Sázka by měla být rozdělena v poměru pravděpodobností na výhru $P(W_1) : P(W_2)$, který odpovídá šanci na výhru prvního hráče (pro druhého hráče je

šance na výhru vyjádřena převrácenou hodnotou tohoto podílu, jde tedy o stejné rozdělení banku).

$$\sum_{k=0}^{z_2-1} \binom{n_z}{k} : \sum_{k=0}^{z_1-1} \binom{n_z}{k}. \quad (1)$$

Doplňme, že v situaci popsané Andělem je $z_1 = 1$, $z_2 = 3$, a tedy $n_z = z_1 + z_2 - 1 = 3$ a vztah (1) dává výše uvedený poměr 7:1.

2. Jak rozdělit sázku, když šance nejsou vyrovnané a stav je 0:0

Uvažujme nyní zobecnění úlohy o rozdělení sázky spočívající v tom, že šance na výhru v jednotlivých partiích nejsou pro oba hráče stejné. Budeme se na hru dívat optikou prvního hráče. Jeho pravděpodobnost výhry v jednotlivých partiích budeme označovat $p_1 \in (0, 1)$.

Vyšetřeme nyní závislost pravděpodobnosti celkové výhry prvního hráče $P(W_1)$ na pravděpodobnosti jeho výhry v jednotlivých partiích p_1 . Očividně $P(W_1)$ je rostoucí funkcí p_1 – čím větší je pravděpodobnost výhry v jedné partii, tím větší je pravděpodobnost celkové výhry. Tuto funkci budeme označovat w_1 , t.j.

$$P(W_1) = w_1(p_1).$$

Představme si posloupnost n nezávislých pokusů, v nichž pravděpodobnost výhry v jedné partii je p_1 (a je ve všech partiích stejná). Celkový počet vyhraných partií označme X_1 . Tato náhodná veličina má binomické rozdělení s parametry n a p_1 . K celkové výhře přitom dojde, pokud počet výher X_1 bude alespoň h . Proto

$$w_1(p_1) = P(X_1 \geq h).$$

Tato pravděpodobnost se dá vyjádřit pomocí distribuční funkce binomického rozdělení $F_{\text{Bin}(n, p_1)}$ takto:

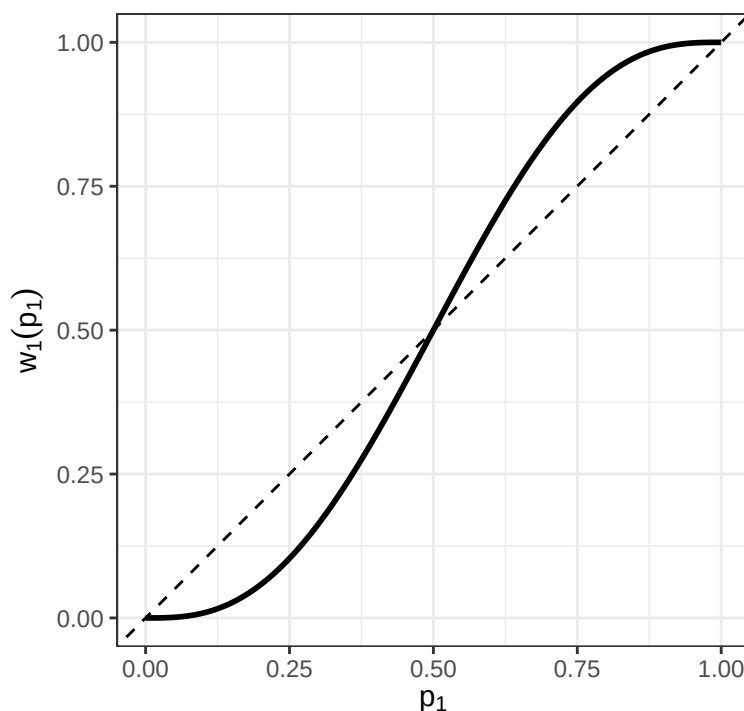
$$w_1(p_1) = 1 - P(X_1 \leq h - 1) = 1 - F_{\text{Bin}(n, p_1)}(h - 1). \quad (2)$$

Praktický výpočet vychází z tvaru pravděpodobnostní funkce binomického rozdělení

$$w_1(p_1) = 1 - \sum_{k=0}^{h-1} \binom{n}{k} p_1^k (1 - p_1)^{n-k}. \quad (3)$$

Ze vztahu (3) je patrné, že $w_1(p_1)$ lze vyjádřit jako polynomickou funkci proměnné p_1 . Například pro $h = 3$ platí²

$$\begin{aligned} w_1(p_1) &= \binom{5}{3} p_1^3 (1-p_1)^2 + \binom{5}{4} p_1^4 (1-p_1)^1 + \binom{5}{5} p_1^5 (1-p_1)^0 \\ &= 6p_1^5 - 15p_1^4 + 10p_1^3. \end{aligned} \quad (4)$$



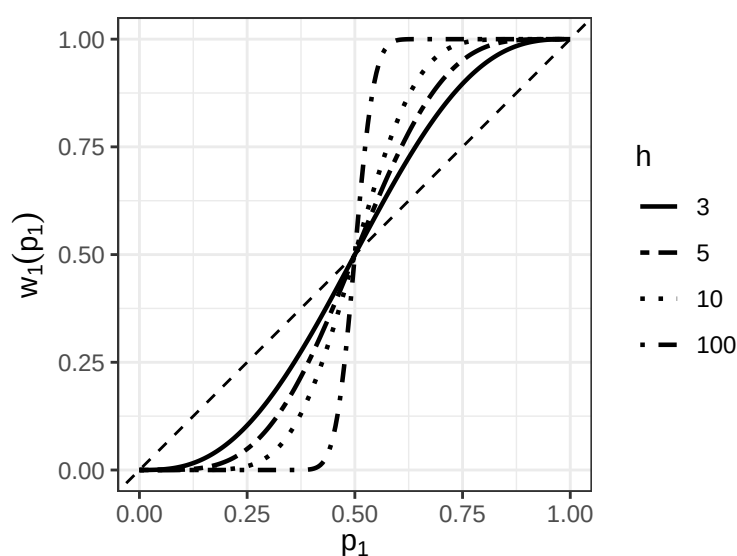
Obrázek 1: Závislost pravděpodobnosti celkové výhry $w_1(p_1)$ na pravděpodobnosti výhry v jedné partii p_1 při hře na tři vítězství.

Obr. 1 znázorňuje závislost pravděpodobnosti celkového vítězství $w_1(p_1)$ na pravděpodobnosti výhry v jedné partii p_1 v situaci, kdy k celkovému vítězství stačí tři vyhrané partie ($h = 3$). V obrázku je pro snazší orientaci

²S úpravami nám může pomoci např. software Mathematica:
`1-CDF[BinomialDistribution[5,p],2]`
`Expand[FunctionExpand[%]]`

dokreslena diagonála $w_1(p_1) = p_1$ odpovídající situaci, kdy se hraje jediná partie ($h = 1$). Je vidět, že (pro $h = 3$) je-li např. $p_1 = 0,25$, je $w_1(p_1) < 0,25$, tedy pokud má hráč 25% pravděpodobnost výhry v jedné partii, je pravděpodobnost, že celkově zvítězí, menší než 25 %. Potřeba zvítězit opakovaně v několika partiích tak zvyšuje férovost hry – pravděpodobnost celkového vítězství „lepšího“ hráče (hráče s $p_1 > 0,5$) je ještě větší než pravděpodobnost jeho vítězství v jednotlivých partiích.

Čím větší je počet vítězných partií požadovaný pro celkové vítězství, tím větší je (při libovolné pevně dané hodnotě p_1) pravděpodobnost výhry lepšího z hráčů, tedy platí „čím více potřebných vítězství, tím je hra férovější“. Tato zákonitost je ilustrována na obrázku 2.



Obrázek 2: Srovnání závislosti pravděpodobnosti celkové výhry $w_1(p_1)$ na pravděpodobnosti výhry v jedné partii p_1 pro různé počty vítězných partií požadovaných pro celkové vítězství.

3. Když šance nejsou vyrovnané a hra je rozehraná

Úvahy a výpočty z předešlého odstavce lze snadno upravit pro situaci, kdy hra je již rozehraná. Stejně jako v kapitole 1 budeme počty zbývajících výher

pro jednotlivé hráče označovat z_1 a z_2 , a maximální počet partií potřebných k rozhodnutí n_z .

První hráč zvítězí, bude-li počet partií, které vyhraje, alespoň z_1 , tj.

$$\begin{aligned} w_1(p_1; z_1, z_2) &= 1 - P(X_1 \leq z_1 - 1) = 1 - F_{\text{Bin}(n_z, p_1)}(z_1 - 1) \\ &= 1 - \sum_{k=0}^{z_1-1} \binom{n_z}{k} p_1^k (1-p_1)^{n_z-k}. \end{aligned} \quad (5)$$

V tomto vzorci jsme zdůraznili, že pravděpodobnost vítězství prvního hráče není pouze funkcí pravděpodobnosti p_1 , ale závisí také na průběžném stavu hry, resp. na počtu partií, které k celkovému vítězství potřebují jednotliví hráči, tedy na z_1 a z_2 . Vzorce (2) a (3) odvozené v kapitole 2 jsou vlastně speciálním případem tohoto vzorce, kde $z_1 = z_2 = h$, a tedy $n_z = n$.

Uvažujme speciální případ, kdy zatím všechny hry vyhrál první hráč a zbývá mu poslední vítězná partie, tedy stav je $(h-1) : 0$. V tomto případě $z_1 = 1$, $z_2 = h$, $n_z = z_1 + z_2 - 1 = h$. Proto pravděpodobnost celkového vítězství prvního hráče je dle vzorce (5)

$$w_1(p_1; 1, h) = 1 - (1 - p_1)^h.$$

Pokud p_1 je blízké nule, může být tato pravděpodobnost stále dosti malá. Např. při hře na tři vítězství ($h = 3$) a průběžném vedení „outsidera“, jehož pravděpodobnost vítězství v jednotlivých partiích je $p_1 = 0,05$, je stále výrazně pravděpodobnější, že vyhraje druhý hráč, neboť pravděpodobnost celkové výhry prvního hráče je $1 - (1 - 0,05)^3 = 0,14$. Na druhou stranu, pokud je počet partií dosti velký a vedoucí outsider není příliš horší než favorit, má již dosti dobrou šanci na vítězství. Např. při $p_1 = 0,2$ a $h = 5$, je pravděpodobnost výhry prvního hráče $1 - (1 - 0,2)^5 = 0,67$.

4. Pascalovo řešení úlohy o rozdělení sázky

Pascalova úvaha vedoucí k řešení úlohy o rozdělení sázky je dnes známá jako *věta o úplné pravděpodobnosti*. Pascal si uvědomil, že po situaci, kdy k celkovému vítězství chybí prvnímu z hráčů z_1 vítězných partií a druhému hráči z_2 partií, mohou následovat pouze dvě (vzájemně se vylučující) situace:

- první zvítězí a bude mu zbývat $z_1 - 1$ vítězných partií, přičemž druhému stále bude chybět z_2 partií k vítězství,
- druhý hráč zvítězí a bude mu zbývat $z_2 - 1$ vítězných partií, zatímco prvnímu stále bude chybět z_1 partií.

Jelikož první z uvedených možností nastane s pravděpodobností p_1 a druhá s pravděpodobností $1 - p_1$, musí pro pravděpodobnost celkového vítězství prvního hráče $w_1(p_1; z_1, z_2)$ dle věty o úplné pravděpodobnosti platit:

$$w_1(p_1; z_1, z_2) = w_1(p_1; z_1 - 1, z_2) \cdot p_1 + w_1(p_1; z_1, z_2 - 1) \cdot (1 - p_1). \quad (6)$$

Přitom zřejmě

- $w_1(p_1; 0, z_2) = 1$ pro libovolné $z_2 > 0$,
- $w_1(p_1; z_1, 0) = 0$ pro libovolné $z_1 > 0$.

Nyní již stačí tuto diferenční soustavu rovnic vyřešit – vyjádřit $w_1(p_1; z_1, z_2)$.

5. Když pravděpodobnost výhry neznáme – Bayesovský přístup

V předchozích částech jsme předpokládali, že známe pravděpodobnost výhry jednotlivých hráčů v jednotlivých partiích, tedy p_1 a $p_2 = 1 - p_1$ jsou známé konstanty. V klasické podobě úlohy popsané v kapitole 1 přitom jde o „férovou“ hru, kdy $p_1 = p_2 = 0,5$. Pokládat hodnoty p_1 a p_2 za známé je realistické, například když jde o házení (dobře vyváženou) kostkou nebo jinou hazardní hru, kdy p_1 umíme vypočítat, jako je třeba ruleta ($p_1 = 18/37$ z pohledu sázejícího na černou/červenou; u americké rulety, která má zdvojenou nulu, je $p_1 = 18/38$). Jak ale rozdělit výhru mezi dva tenisty, když první z hráčů vede 2:0 na sety (při hře na tři vítězné sety), ale jeho dosavadní zápasová bilance je pasivní – všechny čtyři předešlé vzájemné duely prohrál? Otázka také může znít, jak rozdělit svou sázku na vítězství těchto hráčů? Parametr p_1 je v tomto případě neznámý. Považujeme za přirozené o něm uvažovat bayesovsky, tedy jako o náhodné veličině.

5.1. Volba apriorního rozdělení hodnoty p_1

Ještě předtím než hráči začnou hrát a než uvidíme, jak se kterému z nich daří, máme nějakou více či méně vágní představu o pravděpodobnosti výhry prvního hráče v jednotlivých partiích p_1 . Tuto představu můžeme vyjádřit pomocí hustoty $f_{p_1}: (0, 1) \rightarrow \mathbb{R}_0^+$.

Za apriorní rozdělení hodnot p_1 můžeme vzít například beta rozdělení. Připomeňme, že toto rozdělení je dáno hustotou

$$f_{p_1}(p_1) = \frac{1}{B(\alpha, \beta)} \cdot p_1^{\alpha-1} (1 - p_1)^{\beta-1}, \quad p_1 \in (0, 1), \quad (7)$$

kde α a β jsou parametry rozdělení (jde o kladná reálná čísla) a $B(\alpha, \beta)$ je hodnota tzv. beta funkce v bodech α, β (viz kapitola 1.3.1 v [2]). Hodnota $B(\alpha, \beta) \in (0, \infty)$ je vlastně jen normalizační konstanta – kladné reálné číslo, jímž násobíme výraz $p_1^{\alpha-1}(1-p_1)^{\beta-1}$, aby šlo o hustotu (integrál hustoty přes celé $\mathbb{R}$ musí být roven 1).

Parametry apriorního rozdělení α a β volíme podle typu hry, jakou spolu hráči hrají, a také podle toho, co o nich víme. Pokud nemáme důvod žádného z hráčů favorizovat, položíme $\alpha = \beta$ (dále budeme mluvit jen o parametru α). Čím větší hodnotu α v takovém případě zvolíme, tím menší rozptyl bude rozdělení mít, neboť platí

$$\text{var}(p_1) = \frac{\alpha\beta}{(\alpha+\beta)^2(\alpha+\beta+1)} = \left|_{\alpha=\beta} \frac{\alpha^2}{(2\alpha)^2(2\alpha+1)} = \frac{1}{4(2\alpha+1)}.$$

Rozptyl přitom můžeme chápat jako míru nejistoty – čím větší α , tím menší rozptyl a tím menší nejistota. Míra apriorní nejistoty ohledně hodnoty p_1 (vyjádřená, jak jsme uvedli, pomocí parametrů apriorního rozdělení) má podstatný vliv na odhad celkových šancí na výhru:

- Uvažujeme nejprve hru, kde výsledek téměř nezávisí na schopnostech hráčů, ale závisí hlavně na „štěstí“. Takovou hrou je například hra, v níž o výsledku rozhoduje hod mincí. Nevíme-li nic o minci, se kterou se hází, bude rozumné uvažovat apriorní rozdělení s hustotou symetrickou kolem hodnoty 0,5 a s malým rozptylem, který odpovídá (malé) nejistotě ohledně „férovosti“ mince. Dojde-li ve hře např. ke stavu 3:0, naši představu o hodnotě parametru p_1 to sice ovlivní (začínáme se lehce klonit k možnosti, že mince je „vychýlená“), ovšem zřejmě nijak výrazně, neboť i při hodnotě $p_1 = 0,5$ může k danému stavu celkem běžně dojít.
- V jiném typu her výsledek hodně závisí na schopnostech hráčů. Jde např. o různá sportovní klání. Pokud na začátku o kvalitách (schopnostech) soupeřů nic nevíme, budeme opět volit apriorní rozdělení symetrické kolem hodnoty 0,5, neboť nemáme důvod žádného ze soupeřů favorizovat. Bude však rozumné volit apriorní rozdělení vágní, tj. s velkým rozptylem. Dojde-li pak ve hře ke stavu 3:0, bude to celkem jasný signál, že jeden ze soupeřů je lepší a naši představu o hodnotě parametru p_1 to značně ovlivní.

Zatím jsme probrali případy, kdy nemáme důvod dopředu některého z hráčů favorizovat. Někdy však favorita známe. Jde o situaci, kdy výsledek hry

záleží dosti na schopnostech hráčů, a my známe dosavadní vzájemnou bilanci hráčů. První hráč v minulosti vyhrál v_1 partií a jen v_2 partií prohrál ($v_1 > v_2$). Pak rozumnou volbou apriorního rozdělení je $\text{beta}(1 + v_1, 1 + v_2)$. Toto rozdělení je vlastně aposteriorním rozdělením v situaci, kdy apriorní rozdělení je rovnoměrné, tj. $\text{beta}(1,1)$, po odehrání $v_1 + v_2$ partií s v_1 vítězstvími prvního hráče.

5.2. Rozdělení hodnoty p_1 a rozdělení hodnoty w_1

Naše přesvědčení se týká hodnoty p_1 vyjadřující pravděpodobnost výhry v jedné partii, nikoliv v celé sérii. S touto pravděpodobností se zřejmě intuice vypořádává lépe než s pravděpodobností celkového vítězství, která, jak jsme ukázali, dosti závisí na počtu výher nutných k celkovému vítězství. Ze vztahu (2) však plyne, že w_1 je rostoucí funkcí parametru p_1 , a můžeme proto vypočítat hustotu veličiny w_1 z hustoty veličiny p_1 . Využijeme přitom následující vztah, viz Věta 3.5 v [2]:

$$f_{w_1}(y) = \frac{f_{p_1}(w_1^{-1}(y))}{w_1'(w_1^{-1}(y))}, \quad y \in (0, 1), \quad (8)$$

kde f_{w_1} je hustota veličiny w_1 (pravděpodobnost celkového vítězství), f_{p_1} je hustota veličiny p_1 (pravděpodobnost vítězství v jedné partii) a w_1^{-1} je inverzní zobrazení k zobrazení, které hodnotě p_1 přiřazuje hodnotu w_1 dle vzorce (2).

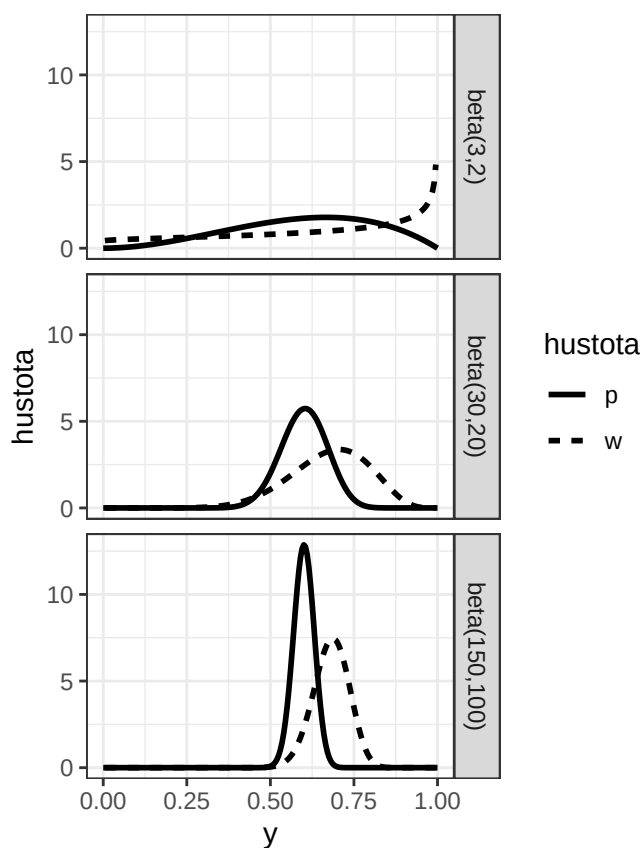
Drobnou vadou na kráse (či právě naopak krásou?) je skutečnost, že inverzní funkci často neumíme explicitně vyjádřit, neboť k jejímu vyjádření by bylo potřeba nalezení kořene příslušného polynomu (vyššího stupně). Například pro výše zmíněnou situaci hry na tři vítězství ($h = 3$) je $w_1(p_1)$ polynomem 5. stupně, viz (4). Derivace této funkce má tvar

$$w_1'(p_1) = 30p_1^2(1 - p_1)^2,$$

a proto při hustotě proměnné p_1 odpovídající beta rozdělení s parametry α a β můžeme hustotu veličiny w_1 v libovolném bodě $y \in (0, 1)$ vyjádřit následovně:

$$f_{w_1}(y) = \frac{(w_1^{-1}(y))^{\alpha-1}(1 - w_1^{-1}(y))^{\beta-1}}{B(\alpha, \beta) \cdot 30(w_1^{-1}(y))^2(1 - w_1^{-1}(y))^2}.$$

Pro pevně dané $y \in (0, 1)$ vypočteme hodnotu $w_1^{-1}(y)$ jako hodnotu $p_1 \in (0, 1)$, pro niž platí $6p_1^5 - 15p_1^4 + 10p_1^3 = y$ v intervalu $(0, 1)$.



Obrázek 3: Hustota náhodné veličiny p_1 (plná linka) a jí odpovídající hustota náhodné veličiny w_1 (prerušovaná linka) ve třech různých situacích, kdy rozdělení veličiny p_1 je beta rozdělení s parametry $\alpha = 3$, $\beta = 2$ (nahore), $\alpha = 30$, $\beta = 20$ (uprostřed) a $\alpha = 150$, $\beta = 100$ (dole).

Souvislost hustoty f_{p_1} a f_{w_1} je ukázána na obrázku 3. Ten se opět týká hry na tři vítězství (k výpočtům jsou použity výše uvedené vzorce), přičemž parametr p_1 má beta rozdělení. Jsou přitom uvažovány tři různé dvojice parametrů α a β , přičemž jejich poměr je vždy 3:2 – jde o dvojice (3, 2), (30, 20) a (150, 100). Uvažujeme tak různou míru nejistoty ohledně rozdělení parametru p_1 . Pro dvojici (3, 2) je nejistota největší, naopak pro dvojici (150, 100) je nejmenší.

Ve všech případech vidíme, že hustota f_{w_1} je oproti hustotě f_{p_1} více koncentrována u okraje intervalu $(0,1)$. To odpovídá našemu pozorování účinnému v kapitole 2, že s rostoucím počtem partií potřebým k celkovému vítězství se hra stává férovější, neboť ten, kdo je mírně favorizován v jedné partii, je (při nutnosti vyhrát více partií k dosažení celkového vítězství) ještě více favorizován na celkové vítězství.

5.3. Jak v bayesovském přístupu rozdělit sázku

V klasickém přístupu bylo celkem zřejmé, jak rozdělit sázku. Když v situaci uvažované v knize [1] může nastat 8 různých stejně pravděpodobných scénářů, z nichž jen jeden vede k vítězství druhého hráče, dělí se sázka v poměru 7:1. Jde tedy o poměr pravděpodobností vítězství, tedy $P(W_1)$ ku $P(W_2)$. Stejný poměr využíváme i v situaci, kdy p_1 nemusí být roven $1/2$, ale jde o libovolnou hodnotu z intervalu $(0,1)$, viz druhá a třetí kapitola tohoto textu. Jak ale postupovat v situaci, kdy ohledně hodnoty parametru p_1 panuje nejistota? V úvahách nám pomůže následující jednoduchý příklad:

Představme si, že máme k dispozici tři hrací kostky – dvě klasické a třetí, která má na třech stěnách vždy jednu tečku a na zbylých třech není nic. Náhodně vybereme jednu kostku a hodíme. Vyhráváme, padne-li „jednička“. Pravděpodobnost výhry je tedy buď $1/6$ (pokud jsme vybrali jednu z klasických kostek), nebo $1/2$ (pokud jsme vybrali speciální kostku). Pravděpodobnost výhry snadno vyjádříme dle věty o úplné pravděpodobnosti takto:

$$P(W_1) = \frac{1}{6} \cdot \frac{2}{3} + \frac{1}{2} \cdot \frac{1}{3}.$$

Výraz na pravé straně této rovnosti přitom můžeme také chápat jako střední hodnotu náhodné veličiny w_1 , která může nabývat hodnoty $1/6$ s pravděpodobností $2/3$, a hodnoty $1/2$ s pravděpodobností $1/3$, neboť pro tuto veličinu platí

$$Ew_1 = \frac{1}{6} \cdot \frac{2}{3} + \frac{1}{2} \cdot \frac{1}{3}.$$

Proto v bayesovském přístupu k úloze o rozdělení sázky dělíme sázku v poměru Ew_1 ku $(1 - Ew_1)$, kde

$$Ew_1 = \int_0^1 y f_{w_1}(y) dy = \int_0^1 w_1(p_1) f_{p_1}(p_1) dp_1 = E(w_1(p_1)).$$

Podotkněme, že spravedlivé dělení sázky odpovídá nejen poměru *pravděpodobností vítězství* jednotlivých hráčů, ale také poměru *očekávaných výher*

jednotlivých hráčů, neboť očekávanou výhru vypočteme (v klasickém případě) jako součin pravděpodobnosti vítězství a částky, o kterou se hraje (v případě prohry je velikost výhry nulová). V bayesovském případě očekávanou výhru vypočteme jako součin Ew_1 a částky, o kterou se hraje. V obou případech pak *poměr* očekávaných výher obou hráčů nezávisí na částce, o kterou se hraje.

Postup si opět ukážeme pro hru na tři vítězství. S využitím vztahu (4) dostáváme

$$Ew_1 = 6Ep_1^5 - 15Ep_1^4 + 10Ep_1^3. \quad (9)$$

Nyní stačí využít vztahu pro obecné momenty beta rozdělení (viz [5]):

$$EX^k = \frac{\alpha + k - 1}{\alpha + \beta + k - 1} EX^{k-1} = \prod_{i=0}^{k-1} \frac{\alpha + i}{\alpha + \beta + i}. \quad (10)$$

Využitím vztahů (9) a (10) pak dostáváme výsledný poměr

$$\frac{Ew_1}{1 - Ew_1} = \frac{(\alpha + 2)(\alpha + 1)\alpha}{(\beta + 2)(\beta + 1)\beta} \cdot \frac{\alpha^2 + 5\alpha\beta + 7\alpha + 10\beta^2 + 25\beta + 12}{10\alpha^2 + 5\alpha\beta + 25\alpha + \beta^2 + 7\beta + 12}.$$

5.4. Numerická ilustrace

Výsledky uvedené v této kapitole použijeme k výpočtu poměru, v jakém má být v několika konkrétních situacích rozdělena sázka. Budeme uvažovat zatím nedohranou hru na tři vítězství pro všechny možné průběžné stavy od 0:0 až po 2:2 a různá apriorní rozdělení. Budeme přitom uvažovat situace, kdy žádného z hráčů dopředu nefavorizujeme (rozdělení $\text{beta}(1,1)$ a $\text{beta}(4,4)$), kdy je jeden z hráčů mírným favoritem (rozdělení $\text{beta}(3,2)$ a $\text{beta}(2,3)$), a kdy je jeden hráč favorizován celkem jasně ($\text{beta}(5,2)$ a $\text{beta}(2,5)$).

Je třeba ještě rozmyslet, jakým způsobem budeme nakládat s novou informací získanou v průběhu hry. Ta samozřejmě mění naši představu o rozdělení parametrů beta rozdělení: výhra prvního hráče v první partii posune naši představu o hodnotě p_1 směrem k větším hodnotám, zatímco jeho eventuální prohra tuto představu posune směrem k menším hodnotám. Stav 1:0 (resp. 0:1) však zároveň přibližuje (vzdaluje) prvního hráče k (od) celkovému vítězství. I kdybychom zůstali u apriorního rozdělení veličiny p_1 , bude každá partie měnit poměr spravedlivého rozdělení sázky. Pokud bychom např. u tenisových utkání znali dosavadní počet vítězství a proher hráčů ve vzájemných zápasech, nikoliv však detailní informaci o poměru setů v těchto utkáních, je kombinování apriorní informace (na úrovni zápasů) a aktuální

informace (na úrovni setů) problematické. V této situaci může být opodstatněné zůstat u apriorního rozdělení a zohledňovat aktuální stav pouze jako rozdílnou „vzdálenost“ hráčů od celkového vítězství.

Výsledky výpočtů jsou uvedeny v tabulce 1 na následující straně. Pro každé apriorní rozdělení a každý průběžný stav je uveden poměr rozdělení sázky, když rozdělení p_1 průběžně aktualizujeme. V závorce je pak vždy uvedeno, jaký by tento poměr byl, kdybychom zůstali u apriorního rozdělení.

Několik pozorování z tabulky:

- Pro symetrická apriorní rozdělení: Za nerozhodného stavu je rozdělení sázky vždy 1:1. Všimněme si symetrie: za stavu 0:2 rozdělujeme v poměru 1:19, za stavu 2:0 v poměru 19:1. Rozhodně stojí za povšimnutí, že pokud bychom svou velmi vágní představu neaktualizovali, byl by poměr za výše uvedených stavů „jen“ 1:3, resp. 3:1. Stojí také za povšimnutí, že např. vedení 2:0 mění při vágním apriorním rozdělení $\text{beta}(1,1)$ poměr rozdělení sázky (19:1) výrazněji než při méně vágním apriorním rozdělení $\text{beta}(4,4)$, kdy je poměr 10:1. Pozorované dvě výhry totiž výrazněji mění naši představu o rozdělení parametru p_1 .
- Pro nesymetrická apriorní rozdělení: Za nerozhodného stavu je sázka rozdělena „nerovně“ ve prospěch toho, kdo je favorizován apriorním rozdělením. Pokud svou představu neaktualizujeme, nerovnoměrnost poměru rozdělení při nerozhodných stavech 0:0, 1:1 a 2:2 postupně klesá až k poměru určeném parametrem apriorního rozdělení, např. pro $\text{beta}(3,2)$ jsou tyto poměry postupně 9:5 (tj. cca 1,8 : 1), 22:13 (tj. cca 1,7 : 1) a 3:2 (tj. 1,5 : 1). Pokud svou představu o rozdělení p_1 průběžně aktualizujeme, pak nerozhodný stav posouvá naši představu směrem k vyrovnanosti hráčů (p_1 posouvá směrem k hodnotě 1/2). Proto rozdělení za stavu 2:2 je blíže k poměru 1:1 než kdybychom svou představu neaktualizovali – poměr je 5:4 pro $\text{beta}(3,2)$ rozdělení, což je blíže k 1:1 než poměr 3:2.

Zajímavých pozorování je možno z tabulky jistě učinit více. Ponecháváme tato pozorování na čtenářích.

6. Závěr

Článek připomíná slavnou úlohu o rozdělení sázky, která se dotýká samotných základů teorie pravděpodobnosti. Připomněli jsme její řešení, zobecnili je na situaci, kdy pravděpodobnost výhry obou hráčů v jednotlivých partiích

Stav	Parametry			Rozdělení					
	n_z	z_1	z_2	beta(1,1)	beta(4,4)	beta(3,2)	beta(2,3)	beta(5,2)	beta(2,5)
0:0	5	3	3	1:1 (1:1)	1:1 (1:1)	9:5 (9:5)	5:9 (5:9)	53:13 (53:13)	13:53 (13:53)
0:1	4	3	2	1:4 (2:3)	3:8 (23:43)	5:9 (1:1)	1:5 (17:53)	35:31 (2:1)	29:301 (5:37)
1:0	4	2	3	4:1 (3:2)	8:3 (3:2)	5:1 (53:17)	9:5 (1:1)	301:29 (37:5)	31:35 (1:2)
0:2	3	3	1	1:19 (1:3)	1:10 (1:5)	5:37 (2:5)	1:20 (4:31)	7:26 (5:7)	4:161 (1:20)
1:1	3	2	2	1:1 (1:1)	1:1 (1:1)	25:17 (22:13)	17:25 (13:22)	119:46 (65:19)	46:119 (19:65)
2:0	3	1	3	19:1 (3:1)	10:1 (5:1)	20:1 (31:4)	37:5 (5:2)	161:4 (20:1)	26:7 (7:5)
1:2	2	2	1	1:4 (1:2)	5:17 (5:13)	5:13 (2:3)	1:5 (1:4)	21:34 (15:13)	6:29 (3:25)
2:1	2	1	2	4:1 (2:1)	17:5 (13:5)	5:1 (4:1)	13:5 (3:2)	29:6 (25:3)	34:21 (13:15)
2:2	1	1	1	1:1 (1:1)	1:1 (1:1)	5:4 (3:2)	4:5 (2:3)	7:4 (5:2)	4:7 (2:5)

Tabulka 1: Poměr, v němž bude sázka rozdělena při hře na 3 vítězství, různých průběžných stavech a různých apriorních rozděleních. Poměry v závorce jsou počítány za předpokladu, že apriorní rozdělení průběžně neaktualizujeme.

není stejná (hra není „férová“). Ukázali jsme také, jak rostoucí počet partií potřebných k celkovému vítězství zvyšuje šanci favorita na vítězství. Poté jsme ukázali, jak k úloze přistoupit v situaci, kdy pravděpodobnost vítězství hráčů v jednotlivých partiích je neznámá. Přirozený se nám jeví bayesovský přístup k tomuto problému. Ukázali jsme, jak rozdělení pravděpodobnosti výhry v jedné partii transformovat na rozdělení pravděpodobnosti celkové výhry. Nakonec jsme vypočítali, jak při tomto přístupu rozdělit sázku.

Věříme, že zvědavý čtenář může mít mnoho dotazů a zlepšujících nápadů. Pokud se takový čtenář najde, může volně na tento text navázat. Budeme rádi, když své nápady zveřejní na stránkách našeho bulletinu.

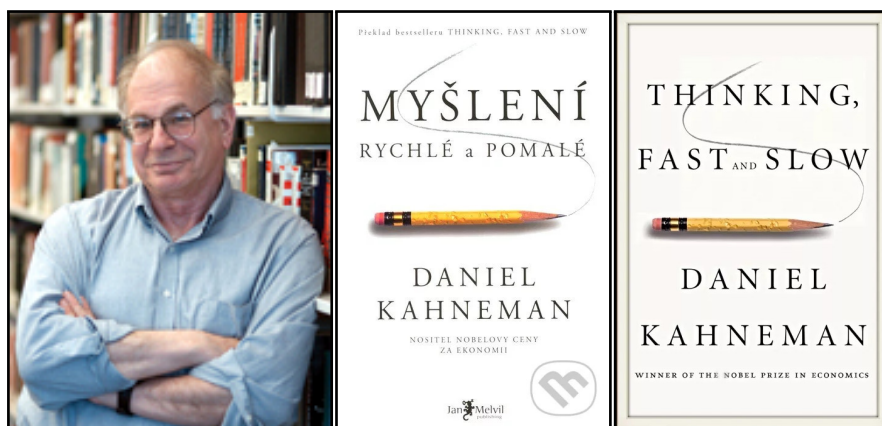
Literatura

- [1] Anděl, Jiří. *Matematika náhody*. Vydání čtvrté. Praha: MatfyzPress, nakladatelství Matematicko-fyzikální fakulty Univerzity Karlovy, 2020. Odborná edice Matfyzpress. ISBN 978-80-7378-420-1. cit. 4, 14
- [2] Anděl, Jiří. *Základy matematické statistiky*. Vydání třetí. Praha: MatfyzPress, nakladatelství Matematicko-fyzikální fakulty Univerzity Karlovy, 2011. Odborná edice Matfyzpress. ISBN 978-80-7378-162-0. cit. 11, 12
- [3] Coufal, Jan. Alea iacta est aneb půl tisíciletí od vytištění úlohy rytíře de Méré. *Informační bulletin ČStS*, 5(1 a 2). cit. 4
- [4] Devlin, Keith. *The unfinished game: Pascal, Fermat, and the seventeenth-century letter that made the world modern*. Basic Books, 2010. cit. 4
- [5] Johnson, Norman Lloyd; Kotz, Samuel; Balakrishnan, Narayanaswamy. *Continuous univariate distributions*. 2nd ed. New York: Wiley, c1994-c1995. ISBN 978-0-471-58494-0. cit. 15
- [6] Laplace, Pierre-Simon (1774). *Mémoire sur la probabilité des causes par les évènements*. English translation by Stephen M. Stigler, appearing in *Statistical Science*, 1(3), 364–378. cit. 4
- [7] Mačák, Karel. *Počátky počtu pravděpodobnosti*. Praha: Prometheus, 1997. Dějiny matematiky. ISBN 80-7196-089-6. cit. 4
- [8] Mlodinow, Leonard. *Život je jen náhoda: jak náhoda ovlivňuje naše životy*. V Praze: Slovart, 2009. ISBN 9788073912598. cit. 4
- [9] Saxl, Ivan. *Vybrané kapitoly z historie teorie pravděpodobnosti a matematické statistiky*. Učební text 2003. cit. 4

KNIŽNÍ RECENZE: MYŠLENÍ RYCHLÉ A POMALÉ BOOK REVIEW: THINKING, FAST AND SLOW

Jan Kalina

E-mail: kalina@cs.cas.cz



Knihu napsal Daniel Kahneman, vyšla 2011, v češtině 2012 v nakladatelství Jan Melvil Publishing, 542 stran, recenzi napsal Jan Kalina (Ústav informatiky AV ČR) v březnu 2021.

Knihu hodnotím jako mimořádně zdařilé dílo o rozhodování za nejistoty a o nedokonalosti lidské intuice. Nejprve popisuje dva základní systémy lidského myšlení (kap. 1). Ty tvoří psychologické pozadí pro kapitoly 2 a 3 věnované rozhodování za neurčitosti; statistikům budou jistě připadat tyto kapitoly za nejhodnotnější. Dále se kniha věnuje ekonomickému rozhodování (kap. 4) a teorii štěstí (kap. 5); právě za ekonomické aplikace poznatků o rozhodování získal Daniel Kahneman Nobelovu cenu za ekonomii v roce 2002. Kniha je hned od první stránky čtivá a přináší řadu převratných myšlenek.

Čím kniha zaujme **statistickou** komunitu? Mě především zaujaly některé úlohy, v nichž se má čtenář rychle zorientovat a intuitivně (tedy co nejrychleji) si tipnout výsledek. V některých takových hádankách jsem se nechal také natchytat. Autor zde argumentuje, že není dobré se spoléhat na svou statistickou intuici; tím ovšem má na mysli pravděpodobnostní intuici nebo schopnosti odhadnout pravděpodobnosti některých jevů kolem nás. Tyto příklady, v nichž jde třeba o to bez počítání odhadnout aposteriorní pravděpodobnosti z daných apriorních pravděpodobností, pochopitelně působí uměle, protože

při reálném rozhodování nebýváme nuceni se rozhodovat naprosto zbrkle. Na druhou stranu by málokdo v reálném životě aplikoval v dané situaci Bayesův vzorec, tím spíš vzhledem k tomu, že naše logické uvažování (slovy autora) ani nebývá ochotno zpochybňovat názor, ke kterému předtím došla (třeba i jen bleskovou rychlostí) naše intuice.

Přestože lidský úsudek je v knize vykreslen jako iracionální a interpretován doslova jako absurdní, směšný či trapný, kniha vyznívá překvapivě optimisticky a naše vlastní nedokonalost je popisována s vtipem. I to přispívá ke čtivosti knihy, zrovna tak jako (často udivující) výsledky zajímavých psychologických experimentů.

K celé knize mám snad jen tyto drobné připomínky:

- Při popisu dvou systémů lidského myšlení (rychlého intuitivního a pomalého logického) je zarážející, že kniha totálně ignoruje individuální odlišnosti (snad až na drobnou zmínku na str. 148) a nezmiňuje dostupné studie, podle nichž mají některé osobnostní typy větší sklony k intuitivnímu uvažování než jiné [1].
- Na str. 220 uvádí autor patrně ve snaze maximálně přiblížit základní pravděpodobnostní pojmy laikům, že korelační koeficient r vždy nabývá hodnot z intervalu $[0, 1]$. Tím spíš může být laik zaskočen, když v příkladu na str. 281 vychází r záporné.
- Na str. 196 nemusí být jasné, co znamená, že korelace je méně než dokonalá. Nedokonalý je zde spíš překlad; zřejmě se rozumí, že nemusí platit $r = 1$ ani $r = -1$, tj. nemusí nastat situace označovaná jako *perfect correlation*.
- V obecném kontextu autor zpochybňuje experty a jejich úsudky, nicméně myslím, že se zde vyjadřuje zejména k oborům, v nichž sám pracoval (psychologie, ekonomie). Např. v medicíně se i přes pokroky v oblasti umělé inteligence uznává, že lékař má oproti algoritmům možnost využívat své tacitní znalosti. Zatímco Kahneman je ve své knize ani nezmiňuje, jde o oblíbený pojem v klinickém rozhodování popisující neformalizované znalosti expertů získané dlouhodobými zkušenostmi.

Literatura

- [1] Miková Š.: *Nejsou stejné. Jak díky Teorii typů porozumět dětem i sami sobě*. Mea Gnosis, Praha, 2018. cit. 20

MEZINÁRODNÍ KONFERENCE V PROJEKTU IOSCS INTL. CONFERENCE IN THE IOSCS PROJECT

Jiří Rybička

E-mail: rybicka@mendelu.cz

Ve dnech 25. a 26. 10. 2021 se v Brně konala „1. mezinárodní konference o otevřených programech v univerzitním informatickém vzdělávání“, která byla organizována jako jedna z aktivit projektu Erasmus+ s názvem „Innovative Open Source courses for Computer Science curriculum (IOSCS)“.

Konference probíhala v prostorách budovy Z areálu Mendelovy univerzity v Brně, která je jedním z partnerů projektu.

Cílem konference bylo nejen představit projekt a jeho dosavadní řešení, ale také seznámit účastníky s širší problematikou open source systémů a jejich využití ve vědecké, výzkumné a výukové praxi.

První příspěvek přednesl dr. Remigiusz Olejnik, koordinátor projektu ze ZUT (Zachodniopomorski Uniwersitet Techniczny ve Štětíně). Jeho obsahem byla informace o projektu, o dosud proběhlých aktivitách a výhledu na poslední rok řešení.

O jedné z aktivit – Intenzivní letní škola vzdělávání s Open Source nástroji – referoval z pohledu studenta Jan Diviš.

Poněkud šířeji byla koncipována zvaná přednáška doc. Roberta Maříka z LDF MENDELU s názvem Jupyter, aneb vědecké výpočty moderně a snadno. Ukázal v ní, jak kombinací několika open source nástrojů lze provádět poměrně náročné výpočty a transparentně pracovat s daty.

Tematicky do zaměření projektu velmi dobře zapadala i druhá zvaná přednáška předsedy České statistické společnosti a pedagoga Univerzity Palackého v Olomouci dr. Ondřeje Vencálka s názvem Daty podložené omyly. Ukázal v ní mimo jiné, jak lze efektivně zpracovávat data statistickým open source systémem R a jak je velmi důležité umět dosažené výsledky správně interpretovat.

Cílem projektu je konstrukce vzdělávacích předmětů na vysokoškolské úrovni, které používají open source nástroje. Představení těchto předmětů bylo předmětem dalšího bloku příspěvků: Probability and statistics with programming in R (dr. Aleš Kozubík, Žilinská univerzita), Mathematical Analysis supported by wxMaxima (dr. Rudolf Blaško, Žilinská univerzita), Open Source tools for text processing (doc. Jiří Rybička, Mendelova univerzita), Programming language Lua (dr. Tomáš Hála, Mendelova univerzita), Mobile Application Development (Radosław Maciaszczyk, ZUT) a Wireless Signal Processing in GNU Radio Environment (Remigiusz Olejnik, ZUT).

Konference byla určena zejména pro studenty vysokých škol, byli zde hlavně studenti z VUT, jimž se informace tohoto zaměření budou hodit nejen ve studiu, ale následně i v inženýrské praxi. Celkem bylo zaregistrováno přes sto účastníků.

