Luděk Spíchal

Zipfův zákon a další mocninné zákony

Učitel matematiky, Vol. 28 (2020), No. 2, 94–109

Persistent URL: `http://dml.cz/dmlcz/148635`

ZIPFŮV ZÁKON A DALŠÍ MOCNINNÉ ZÁKONY

Luděk Spíchal

Úvod

Různé objekty podléhající zkoumání mají v řadě případů určitou typickou velikost. Měření či pozorování dávají výsledky, které se koncentrují kolem určité průměrné hodnoty s odchylkami relativně malého rozsahu. Takto můžeme uvažovat např. o měření hmotnosti zralých (životaschopných) semen určité dřeviny. Zjištěné hodnoty v daném případě oscilují pouze v omezeném rozsahu hodnot, případné extrémy prakticky vylučují životaschopnost semene. Grafickým znázorněním rozložení hodnot bude v uvedených případech zvonovitá křivka s výraznou koncentrací hodnot kolem průměru (normální rozdělení).

Na druhou stranu nalezneme příklady řady měření, která neodpovídají výše uvedenému modelu. Měřené hodnoty nejsou centrované kolem průměru a mohou zasahovat do několika číselných řádů. Příklady takových rozdělení jsou např. rozdělení počtu obyvatel v obcích a městech nebo jednotlivých slov v literárních textech, tedy rozdělení s širokým rozpětím hodnot. Zmíněná rozdělení (Paretův zákon, Zipfův zákon) jsou v literatuře dobře popsaná (např. Garbaix, 1999, Luckstead, 2014, Newman, 2004) a jako taková zasahují do řady různých oblastí přírodních i společenských disciplín.

Záměrem článku je krátce zmínit historii problematiky, ověřit platnost uvedených zákonitostí na datech z českých reálií a zasadit pojem Zipfova zákona do širších souvislostí.

Historie

Vědci se někdy věnují z pohledu běžného smrtelníka podivným oblastem výzkumu, často ovšem podivným jen na první pohled. Takto například ve 40. letech minulého století strávili výzkumníci z Wisconsinské univerzity v USA mnoho času (více než 1 rok) stříháním a tříděním jednotlivých slov v Odysseovi od Jamese Joyce (Hanley, 1937). Jednotlivá slova poté uspořádali podle četnosti výskytu. Tato primárně jazykovědná práce by patrně do matematiky nikdy nepronikla, pokud by si G. Zipf nevšiml fascinující závislosti mezi pořadím slov a jejich četností.[1]

Tab. 1: Vztah mezi pořadím a četností vybraných slov v knize Jamese Joyce Odysseus (upraveno podle Bellos, 2016).

Slovo	**Pořadí**	**Četnost**
já	10	2 653
říct	100	265
taška	1 000	26
rudooranžový	10 000	2

Pokud se pozorně podíváme do tabulky 1, všimneme si, že pro jednotlivá slova ve zmíněném textu platí přibližně následující vztah

$$\text{pořadí} \times \text{četnost} \sim 26\,500,$$

nebo po úpravě

$$\text{četnost} \sim \frac{26\,500}{\text{pořadí}}.$$

Rovnice ukazuje, že četnost výskytu daného slova v textu je nepřímo úměrná jeho pořadí. Spisovatelé, ale nejenom oni, používají slova s určitou frekvencí, řada slov (přibližně 50 %) je ve většině textů však použita pouze jednou (Bellos, 2016).[2]

[1] George Kingsley Zipf (1902–1950) byl americký filolog a lingvista. G. Zipf patrně nebyl prvním, kdo zaznamenal vztah mezi pořadím a četností, německý fyzik Felix Auerbach (1856–1933) si závislosti povšiml již v r. 1913 (Garbaix, 1999).

[2] Zmíněná studie se týkala anglicky psaného textu, v jiných jazycích se zastoupení slov použitých v textech pouze jednou nemusí nutně shodovat

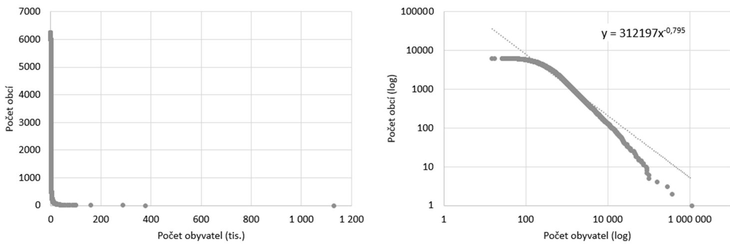
Na základě tohoto pozorování proběhly další výzkumy, které naznačují, že uvedený jev není ojedinělou kuriozitou, ale naopak poměrně často se vyskytujícím vztahem mezi pořadím a četností, který lze vyjádřit ve tvaru

$$\text{četnost} \sim \frac{k}{\text{pořadí}},$$

kde k je kladná konstanta.

Zipfův zákon a rozdělení obyvatelstva v obcích České republiky

V průběhu 20. století se významným způsobem změnilo rozložení světové populace mezi městy a venkovskými oblastmi. Zatímco v r. 1960 se poměr odhadoval na 33/67 ve prospěch venkovských oblastí, v r. 2010 již byla převaha na straně měst vyjádřená poměrem 52/48 (Luckstead, 2014).[3]



Obr. 1: Rozdělení obyvatelstva v obcích ČR (stav k 1. lednu 2018)

Na stránkách Ministerstva vnitra ČR je možné nalézt statistiku týkající se obcí v České republice.[4] Obrázek 1 (vlevo) znázorňuje rozdělení obyvatel v obcích ČR. Na levém okraji grafu

s uvedenou hodnotou.

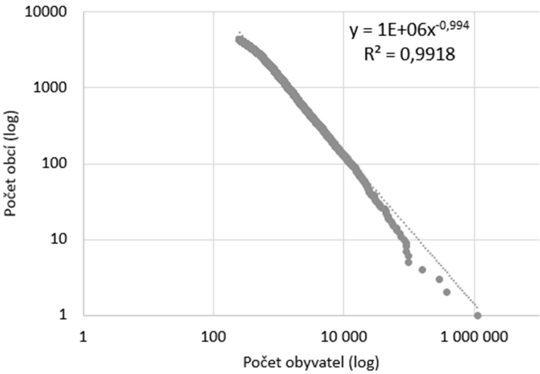
[3]Zmíněný poměr je značně závislý na sledované oblasti. Zatímco ve Spojených státech dosáhl v r. 2010 dokonce hodnoty 82/18 ve prospěch městských aglomerací, v Indii byl ve stejném období 31/69.

[4]Statistika zachycuje celkem 6 257 obcí, stav k 1. lednu 2018, dostupné online z `http://www.mvcr.cz/clanek/statistiky-pocty-obyvatel-v-obcich.aspx`.

zaznamenáme vysoký počet obcí čítajících stovky nebo jednotky tisíc obyvatel. Sestrojený graf tvarem zjevně odpovídá určitému typu mocninné funkce. Mnohem zajímavější obraz rozložení dat ovšem získáme, pokud graf zobrazíme v logaritmickém měřítku (přesněji dvojitém logaritmickém, tj. na obou osách). V tomto případě mají data poměrně překvapivě v podstatné části grafu tvar blížící se přímce (obr. 1, vpravo), od které se odchylují v levé části, která odpovídá menším obcím (cca do 250 obyvatel).[5] Vložením spojnice trendu (MS Excel, zvolena mocninná regrese) určíme rovnici regrese (obr. 1, vpravo)

$$y = 312\,197 \cdot x^{-0{,}795}.$$

Tvar závislosti je zjevně ovlivněn skutečností, že v ČR je velký počet relativně malých obcí.



Obr. 2: Rozdělení obyvatelstva v obcích ČR s alespoň 250 obyvateli (stav k 1. lednu 2018)

Graf na obrázku 2 znázorňuje data odpovídající obcím o určité minimální velikosti. Pokud jsou takto zpracovány obce o velikosti alespoň 250 obyvatel (celkem 4 323 obcí v ČR), pak je hodnota

[5] Statistika obcí MV ČR zachycuje samosprávné celky, ke kterým je dále přidruženo cca 12 tisíc místních částí, z nichž ve zhruba 700 nejmenších sídlech žije méně než 5 obyvatel [17]

koeficientu $\alpha = 0{,}994$.[6] Regrese vyjádřená rovnicí

$$y = 10^6 \cdot x^{-0{,}994}$$

po logaritmování

$$\ln y = 13{,}82 - 0{,}994 \ln x$$

ukazuje, že směrnice je velmi blízko hodnotě -1.

Uvedené vyjádření můžeme současně chápat jako formulaci Zipfova zákona: *Pokud znázorníme logaritmus počtu obcí (uspořádaných podle velikosti) proti logaritmu počtu obyvatel (velikost obce), pak získáme přímku se směrnicí blížící se číslu 1.*[7] Pravděpodobnost, že velikost města $\tilde{S}$ je větší než nějaké S, je úměrná hodnotě $1/S$

$$P(\tilde{S} > S) = \frac{k}{S^\alpha},$$

kde $\alpha \simeq 1$, k je nějaká konstanta (Gabaix, 1999).[8]

Zipfův zákon v české literatuře

Zjištění učiněná G. Zipfem můžeme rovněž ověřit na dostupných zdrojích elektronických knih. Např. na stránkách Městské knihovny v Praze je k volnému stažení řada knih, jimž buď vypršela autorská práva, nebo své svolení k volnému šíření poskytli sami

[6] Garbaix (1999) konstatuje, že je věcí pohledu, jakou velikost obce (studie se týkala stavu obvyklého v USA) považovat za hraniční mezi obcí (town) a městem (city).

[7] Uvedenou hodnotu směrnice uvádí např. Garbaix (2012) pro velká americká města. Na druhou stranu někteří autoři, např. Eeckhout (2004), upozorňují, že s rostoucí velikostí vzorku hodnota směrnice klesá. Navíc, v případě malých vzorků nelze výsledky považovat za objektivní.

[8] Zipfův zákon byl testován např. na jazykovém korpusu Brown, který obsahuje zhruba jeden milion slov tvořících vzorky 500 textů z různých oblastí americké angličtiny a popisující jazykový trend v období vzniku korpusu (r. 1961). Provedená studie ukázala, že se rozdělení slov velmi dobře shoduje s Zipfovým zákonem. Nejčastěji zastoupené slovo „the“ se objevuje s frekvencí přibližně 7 %, druhé v pořadí „of“ má frekvenci zhruba 3,5 %.

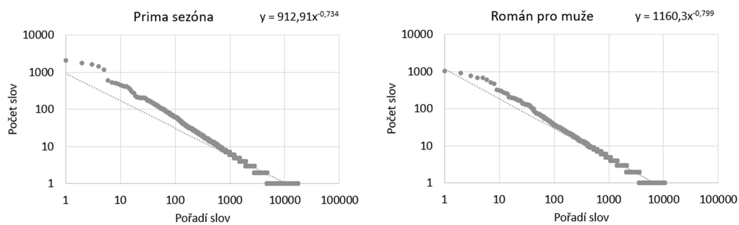
Naopak ve vzorcích textů schizofreniků, malých dětí nebo ve vojenské komunikaci jsou hodnoty exponentu odlišné od čísla 1 (Manin, 2008).

autoři.[9] Pro potřeby článku byly analyzovány dvě knihy: *Prima sezóna* Josefa Škvoreckého a *Román pro muže* Michala Viewegha. V tabulce 2 jsou uvedeny základní statistické informace týkající se celkového počtu slov, celkového počtu různých slov (včetně slov tvořených jednou hláskou), v tabulce 3 pak 10 nejfrekventovanějších slov obou textů.

Tab. 2: Analýza počtu slov *Prima sezóny* Josefa Škvoreckého (vlevo) a *Románu pro muže* Michala Viewegha (vpravo)

	Prima sezóna	**Román pro muže**
Celkový počet slov	60 779	38 091
Celkový počet různých slov	18 927	10 386
Slova použitá pouze jednou (%)	12 801 (≈ 67,6 %)	6 759 (≈ 65,1 %)

Na obrázku 3 jsou znázorněná rozdělení slov podle pořadí v logaritmickém měřítku. Nalezené hodnoty exponentu α jsou výrazně menší než číslo 1, a nevykazují tak dobrou shodu s Zipfovým zákonem popsaným na textech v anglickém jazyce.



Obr. 3: Rozdělení slov podle pořadí absolutní četnosti výskytu v obou textech (log-log měřítko)

Nabízí se přirozená otázka, zda mohou zjištěné odchylky od teoretických hodnot typických pro anglicky psané texty souviset s jazykem obou textů? Česky psané texty mají vskutku jisté specifické rysy, které je odlišují od textů anglických. Vzhledem k provedeným analýzám je patrně nejpodstatnější odlišností oheb-

[9] `https://www.mlp.cz/cz/`

nost českých slov. Analytické programy celkem pochopitelně nezohledňují různé tvary stejného slova, a dochází tak jednak k navýšení celkového počtu slov použitých v textu, jednak k navýšení počtu slov s nízkou frekvencí výskytu (tab. 2).[10]

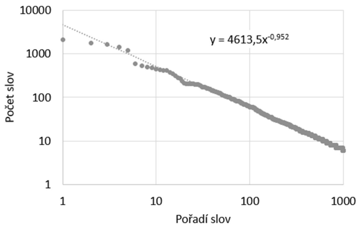
Tab. 3: Absolutní četnost prvních deseti nejfrekventovanějších slov obou textů

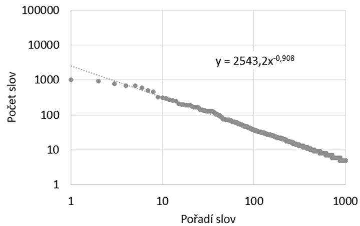
Pořadí frekvence výskytu slov	**Prima sezóna**		**Román pro muže**	
	slovo	**frekvence**	**slovo**	**frekvence**
1.	a	2 088	se	1 011
2.	jsem	1 763	a	910
3.	se	1 617	to	779
4.	to	1 419	je	671
5.	na	1 180	na	668
6.	že	600	v	593
7.	ale	532	si	497
8.	já	498	že	462
9.	je	484	ale	325
10.	tak	446	s	307

Pokud tedy uvážíme, že se významné části obou grafů na obrázku 3 tvarem blíží přímce, a současně vyloučíme okrajové hodnoty odpovídající málo frekventovaným slovům v posuzovaných textech, pak můžeme sledovat značný posun v hodnotě koeficientu α. Obrázek 4 znázorňuje rozdělení prvního tisíce slov obou analyzovaných textů. Prvních tisíc slov s největší frekvencí tvoří v případě *Prima sezóny* cca 61 % z celkového počtu slov, v případě *Románu pro muže* pak cca 64 % slov. Hodnoty koeficientu α

[10] Frekvenční analýza textu byla provedena online na adrese `https://wordcounter.net/`, zde umístěný program lze používat bezplatně. Pro daný text je možné určit nejen celkový počet slov a počet unikátních slov v textu, ale také počet vět tvořících text, průměrnou délku slova, průměrný počet slov ve větách atd. Text je možné psát přímo do textového pole, popř. kopírovat a vkládat z jiného programu. Kromě počítání slov a znaků obsahuje rovněž jednoduchou kontrolu gramatiky, při psaní textu pomáhá zlepšit výběr slov a styl psaní. Programů pracujících na podobném principu je online k dispozici celá řada.

se v obou případech zřetelně přiblížily hodnotě předpokládané Zipfovým zákonem.





Obr. 4: Zastoupení prvního tisíce slov s největší absolutní četností (log-log měřítko) v *Prima sezóně* (vlevo) a *Románu pro muže* (vpravo)

Co stojí za Zipfovým zákonem?

Zipfův zákon byl popsán jako zákonitost popisující vztah mezi pořadím a počtem slov v literárních textech. Uspokojivé teoretické vysvětlení zákona není v současnosti k dispozici a do jisté míry je sporná samotná otázka, zda zákon odhaluje něco zajímavého o struktuře přirozených jazyků (Manin, 2008). Na druhou stranu by neměl být opomenut rozsáhlý soubor provedených a v literatuře popsaných pozorování (např. Pinto, 2012).

Li (1992) provedl test, ve kterém ukázal, že rozdělení slov vytvořených z náhodně generovaných písmen rovněž odpovídá Zipfovu zákonu. Kratší slova se v náhodně generovaných souborech objevovala častěji než delší, z čehož Li usoudil, že vztah frekvence a pořadí je důsledkem rozložení slov podle jejich délky. Přestože zastoupení slov s různou délkou není v přirozených jazycích přesně exponenciální, vývoj lidských jazyků i v tomto případě přirozeně směřoval k obdobnému rozdělení. I když je tedy možné vytvořit mnohem větší počet dlouhých slov než krátkých, faktické zastoupení slov v přirozených jazycích ovlivňuje mimo jiné jejich praktická vyslovitelnost.

Garbaix (1999) zmiňuje dvě podstatné vlastnosti charakterizující Zipfův zákon v souvislosti s rozdělením obyvatel ve městech a obcích.

1. Skutečnost, že Zipfův zákon je zákonem mocninným, můžeme odvodit z jednoduchého fyzikálního principu, kterým je nezávislost (invariance) na měřítku (škále, číselné soustavě).[11] Pokud růst populace probíhá na všech úrovních stejnou měrou, pak musí být výsledná distribuce nezávislá (invariantní) k měřítku.
2. Zipfův zákon je mocninovým zákonem s exponentem 1. Uvažujme konstantní průměrnou velikost populace měst a předpokládejme dále, že populace určitého města se může v každém časovém okamžiku buď zdvojnásobit, nebo zmenšit na polovinu. Pravděpodobnost, že dojde ke zdvojnásobení populace města, je $\frac{1}{3}$ a pravděpodobnost, že se populace zmenší na polovinu, je $\frac{2}{3}$, očekávaný růst je

$$\frac{1}{3} \cdot 2 + \frac{2}{3} \cdot \frac{1}{2} - 1 = 0.$$

   Pokud chceme ukázat, že počet měst dané velikosti může být konstatní, označme velikost určitého města jako S. Snadno lze nahlédnout, že počet měst velikosti $2S$ by měl být poloviční oproti počtu měst velikosti S a počet měst o velikosti $S/2$ by měl být naopak dvojnásobný. Uvedené vyjádření odpovídá formulaci Zipfova zákona, kdy počet měst dané velikosti je úměrný převrácené hodnotě velikosti města.[12]

Příjmení obyvatel České republiky

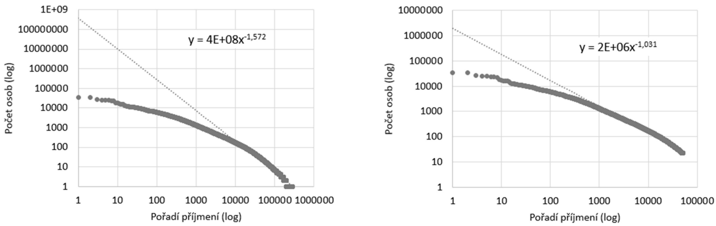
V České republice (dále jen ČR) na konci roku 2016 žilo cca 10 270 000 obyvatel, kteří byli nositeli cca 280 tis. příjmení.[13] Patrně asi nikoho příliš nepřekvapí, že největší zastoupení mělo příjmení Nováková/Novák. Nositelek a nositelů tohoto příjmení žilo v ČR tehdy celkem 34 812, resp. 33 881 [18, 19]. Na opačné straně škály pak statistici zaznamenali téměř 87 tis. příjmení s pouhým

[11] Tato nezávislost je charakteristickou vlastností datových (číselných) souborů řídících se Benfordovým zákonem.

[12] Předpokládáme, že časový rámec, ve kterém probíhají zmíněné změny, je dostatečný k tomu, aby rozdělení populace konvergovalo k Zipfovu zákonu.

[13] Přechýlená příjmení jsou uvažována jako samostatné položky.

jediným nositelem (Casti, 1995). Obrázek 5 znázorňuje rozdělení příjmení obyvatel ČR, kde po vložení spojnice trendu odečteme hodnotu koeficientu $\alpha \simeq 1{,}572$. Hodnota koeficientu $\alpha \gg 1$, rozdělení příjmení v daném případě neodpovídá Zipfovu zákonu.



Obr. 5: Příjmení obyvatel České republiky v r. 2016 (vlevo všechna příjmení, vpravo prvních 51 tis. nejčastějších příjmení, log-log měřítko)

Jestliže podle výše naznačené úvahy (rozdělení obyvatelstva v obcích) budeme sledovat rozdělení příjmení vzhledem k určitému pořadí, pak zaznamenáme přiblížení hodnoty koeficientu α hodnotě 1. V případě 51 tis. nejčastějších příjmení obyvatel (více než 90 % obyvatel ČR) je hodnota koeficientu $\alpha \simeq 1{,}031$ (obr. 5).

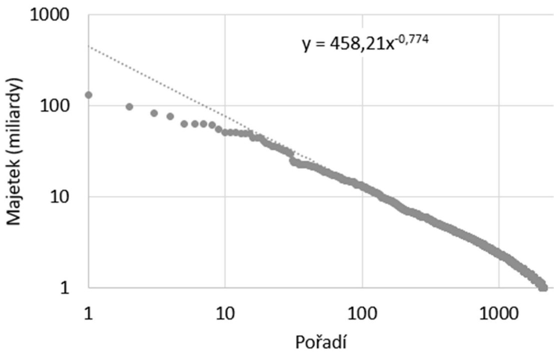
Další mocninné zákony

Zipfův zákon patří ke skupině zákonitostí, které se řídí vztahem nepřímé úměrnosti

$$y \sim \frac{k}{x^a},$$

kde k a a jsou konstanty. Často se takové zákonitosti objevují v přírodních vědách, např. síla zemětřesení je nepřímo úměrná počtu zemětřesení dané velikosti nebo velikost měsíčního kráteru je nepřímo úměrná počtu kráterů dané velikosti (Newman, 2004). Zaznamenat je můžeme také v některých oblastech ekonomiky, např. počet zaměstnanců ve firmách je nepřímo úměrný počtu firem dané velikosti (Stanley, 1995).

Časopis Forbes pravidelně uveřejňuje přehled nejbohatších obyvatel planety. Jejich počet a majetek se v průběhu času samozřejmě mění, na jaře 2019 napočítali redaktoři časopisu 2 153 dolarových miliardářů s celkovým majetkem téměř 9 tis. miliard dolarů (obr. 6).[14]



Obr. 6: Dolaroví miliardáři (log-log měřítko)

V. Pareto[15] přišel na počátku 20. století s teorií, že rozdělení majetku mezi lidmi se řídí vztahem

$$\text{majetek} \sim \frac{k}{\text{pořadí}^a},$$

kde k a a jsou konstanty (Paretův zákon). Nejbohatší člověk je výrazně bohatší než druhý v pořadí, který je bohatší (ne o tolik) než třetí v pořadí atd. Pareto ukázal, že princip platí pro data z mnoha dalších oblastí a období. Taková distribuce bohatství v sobě samozřejmě nese značnou nerovnost, kdy malá skupina lidí vlastní obrovský majetek. Pareto tak podle pozorování v Itálii na začátku 20. století např. uvedl, že cca 20 % obyvatel vlastní cca 80 % pozemků (pravidlo 80/20). Paretovo rozdělení majetku je

[14] Graf je sestaven podle údajů platných k 8. 2. 2019. Nejbohatším člověkem byl Jeff Bezos, zakladatel internetového e-shopu Amazon, s odhadovaným majetkem ve výši 131 miliard dolarů [20].

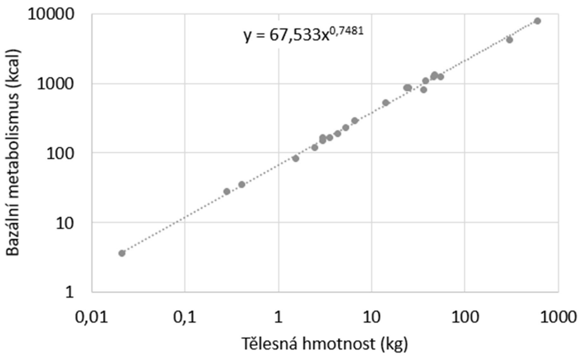
[15] Vilfredo Pareto (1848–1923) byl italský ekonom, matematik, statistik, sociolog a politolog.

tedy, podobně jako Zipfův zákon, popisující rozdělení slov v textech podle délky, mocninným zákonem.

V 30. letech 20. století si M. Kleiber[16] všiml, že velikost bazálního metabolismu teplokrevných živočichů je závislá na tělesné velikosti.[17] Vycházel z nezávislých studií, které sledovaly velikost metabolismu různých druhů živočichů včetně člověka. Ukázal např. že kočka, jejíž hmotnost je zhruba 100krát větší, než je hmotnost myši, spotřebuje za stejnou dobu cca 32násobek energie spotřebované myší. Na obrázku 7 je graficky znázorněna závislost velikosti bazálního metabolismu na tělesné hmotnosti teplokrevných organismů.[18] Použití logaritmického měřítka ukazuje, že regresním grafem je přímka s rovnicí

$$\ln y = \ln 67{,}533 + 0{,}7481 \ln x,$$

jejíž směrnice $k \approx 0,75$.



Obr. 7: Kleiberův zákon (podle Kleiber, 1947; log-log měřítko)

[16]Max Kleiber (1893–1976) byl švýcarský biolog.

[17]Bazální metabolismus (bazální metabolický výdej) je množství energie vydané v klidovém stavu v teplotně neutrálním prostředí na lačno.

[18]Jednotlivé body reprezentují myš, krysu, morče, králíka, kočku, psa, šimpanze, kozu, ovci, krávu. Některé druhy jsou vyjádřeny více body, jeden z bodů znázorňuje ženu s tělesnou hmotností cca 55 kg.

Kleiberův zákon říká, že velikost bazálního metabolismu je úměrná tělesné hmotnosti podle vztahu

$$\text{bazální metabolismus} \sim c\,(\text{hmotnost})^{\frac{3}{4}}\,,$$

kde c je kladná konstanta ($c \approx 70$, Kleiber, 1947; Bellos, 2016). V novějších pracích (např. Prothero, 1984) autoři nepovažovali uvedenou hodnotu exponentu za odpovídající, neboť dle jejich názoru by velikost bazálního metabolismu měla být úměrná velikosti povrchu organismu. Vycházeli přitom z faktu, že velikost povrchu P je (u živočichů s geometricky podobným tvarem těla) úměrná hmotnosti m podle vztahu

$$P \sim m^{\frac{2}{3}}.$$

Protože dále platí, že většina tepla produkovaného bazálním metabolismem M uniká z těla přes povrch, pak by rovněž velikost metabolismu měla být obdobně úměrná hmotnosti

$$M \sim m^{\frac{2}{3}}.$$

Provedená měření (Prothero, 1984) ukázala, že reálná hodnota exponentu je skutečně blíže hodnotě $\frac{2}{3}$ než $\frac{3}{4}$.

Kleiberův zákon se řadí mezi vztahy vyjádřené přímou úměrností, zapisované ve tvaru

$$y \sim kx^a.$$

Velmi podobnou hodnotu exponentu a zaznamenáme také např. ve vztahu, který popisuje závislost plochy křídla S na hmotnosti letce m (Dvořák, 2015)

$$S \sim m^{0{,}71\,(\text{až } 0{,}78)} \text{ pro ptáky a } S \sim m^{0{,}64} \text{ pro netopýry.}$$

Závěr

Mocninné zákony (Paretův, Zipfův) nám říkají, že velikost určitého jevu je nepřímo úměrná jeho frekvenci. Zipfův zákon ukazuje, že na první pohled nahodilé jevy (např. vztah mezi pořadím a četností slov v literatuře nebo rozdělení obyvatelstva v obcích různé velikosti) se mohou řídit určitými pravidly, a lze je tedy matematicky zkoumat. Uvedené zákony jako Zipfův či Paretův nejsou ovšem přijímané bezvýhradně. Někteří autoři je považují za pouhé statistické fenomény a argumentují rovněž absencí silných důkazů platnosti zmíněných zákonů. Další naopak doporučují zkoumat, jak dobře zkoumaná data vyhovují či nevyhovují určitému rozdělení místo jejich odmítání (Pinto, 2012).

V biologických vědách se řada mocninných vztahů shrnuje pod názvem růstové alometrie (např. Kleiberův zákon). Růstové alometrie se v té či oné míře týkají v podstatě všech tělesných orgánů u všech druhů organismů, kde popisují diferencovaný růst tělesných částí organismů. Při zvětšování či zmenšování tělesných rozměrů se velikost orgánů a tělních částí nemění přímo úměrně změně velikosti těla. Odhaduje se například, že při zvětšení tělesných rozměrů jelena o 10 % se může váha paroží zvětšit až o 20 % (Flégr, 2005). Jedním z prvních, kdo se těmito vztahy systematicky zabýval, byl skotský biolog a matemamatik D'Arcy Thompson.[19]

Rozdělení hodnot se ovšem jen zřídka řídí určitým konkrétním mocninným zákonem v celém rozsahu. V reálných podmínkách se obvykle rozdělení odchyluje od mocninného zákona pod určitou minimální hodnotou. V příkladech popisujících rozdělení slov v literárních textech nebo příjmení obyvatel v České republice jsme takové hodnoty zaznamenali. Po jejich vyloučení se hodnoty značně přiblížily Zipfovu zákonu. Na druhou stranu podobné manipulace s datovými soubory celkem samozřejmě vyvolávají nutnost určení vhodné hraniční hodnoty a měly by se týkat pouze malé části dat tvořících v rozdělení hodnot pravý nebo levý ocas (Newman, 2004).

[19]Sir D'Arcy Wentworth Thompson (1860–1948) v roce 1917 uveřejnil svá pozorování a závěry v knize *On Growth and Form*, kde na příkladech živočišných i rostlinných druhů poukazuje na souvislost mezi tělesnými parametry a rozměry tělních částí.

Mocninné zákony se často objevují v komplexních systémech, kde vliv každého jednotlivého prvku na vlastnosti systému je malý (Velarde & Robledo, 2017). Můžeme je chápat jako zajímavý nástroj pro analýzu dat (Pinto, 2012). Tato vlastnost mocninných zákonů nabízí možnost realizace matematického experimentu spočívajícího ve vyhledání vhodných dat (literární, biologické, ekonomické povahy apod.) s jejich statistickým zpracováním.

Literatura

[1] Bellos, A. (2016). *Alex za zrcadlem.* Praha: Dokořán.

[2] Gabaix, X. (1999). Zipf's law for cities: an explanation. *Quarterly Journal of Economics*, 114(3), 739–767.

[3] Luckstead, J. & Devadossb, S. (2014). Do the world's largest cities follow Zipf's and Gibrat's laws? *Economics Letters*, 125, 182–186.

[4] Newman, M. E. J. (2004). Power laws, Pareto distributions and Zipf's law. *Contemporary Physics*, 46(5), 323–351.

[5] Stanley, M., et al. (1995). Zipf Plots and the Size Distribution of Firms. *Economics Letters*, XLIX, 453–457.

[6] Velarde, C. & Robledo, A. (2017). Rank distributions: Frequency vs. Magnitude. *PLoS ONE* 12(10), nestránkováno, 13 s. Dostupné z: `https://journals.plos.org/plosone/article?id=10.1371/journal.pone.0186015`

[7] Casti, J. L. (1995). Bell Curves and Monkey languages. When do empirical relations become a law of nature? *Complexity*, 1(1), 12–15.

[8] Kleiber, M. (1947). Body size and metabolic rate. *Physiological reviews*, 27(4), 511–541.

[9] Dvořák, R. (2015). *Jak létají.* Praha: Academia.

[10] Prothero, J. (1984). Scaling of standard energy metabolism in mammals: I. Neglect of circadian rhythms. *Journal of Theoretical Biology*, 106, 1–8.

[11] Flégr, J. (2005). *Evoluční biologie*. Praha: Academia.

[12] Li, W. (1992). Random texts exhibit Zipf's-law-like word frequency distribution. *IEEE Transactions on Information Theory*, 38(6), 1 842–1 845.

[13] Manin, D. Y. (2008). Zipf's law and avoidance of excessive synonymy. *Cognitive Science*, 32, 1 075–1 098.

[14] Pinto, C. M. A., Mendes Lopes, A. & Tenreiro Machado, J. A. (2012). A review of power laws in real life phenomena. *Commun Nonlinear Sci Numer Simulat*, 17, 3 558–3 578.

[15] Eeckhout, J. (2004). Gibrat's law for (all) cities. *American Economics Review*, 94 (5), 1 429–1 451.

[16] Hanley, M. L. (1937). *Word index to James Joyce's Ulysses.* University of Wisconsin Press.

[17] Lidové noviny, 15. 2. 2019, str. 3.

[18] `http://tn.nova.cz/clanek/novaci-dal-vevodi-cesku-kolik-lidi-ma-prijmeni-jako-vy-seznam.html`

[19] `https://www.kdejsme.cz/`

[20] `https://www.forbes.com/billionaires/#7af8036251c7`

Abstract

The article deals with Zipf's law, which was described as the relationship between the order and frequency of words in literary works. After a brief review of the historical context, there are examples of both literature and available population statistics. Finally, we will show that power laws are a common phenomenon not only in literature but also in biological or economic sciences.

Luděk Spíchal
Masarykova univerzita v Brně
Ústav matematiky a statistiky
Kotlářská 267/2
611 37 Brno
Česká lesnická akademie Trutnov
Lesnická 9
541 11 Trutnov