Pokroky matematiky, fyziky a astronomie

Jiří Dvořák

Proč jsou logaritmické tabulky nejohmatanější na začátku?

Pokroky matematiky, fyziky a astronomie, Vol. 64 (2019), No. 1, 14–28

Persistent URL: `http://dml.cz/dmlcz/147692`

Proč jsou logaritmické tabulky nejohmatanější na začátku?

Jiří Dvořák

Abstrakt. Z dnešního pohledu jsou logaritmické tabulky něco jako film pro pamětníky. Člověk si je nechává na poličce možná ze sentimentu, možná „pro všechny případy“, ale do ruky je vezme vlastně jen při úklidu. Přesto otázka v názvu tohoto příspěvku motivovala vznik zajímavého kousku matematiky s užitečnými aplikacemi, relevantními i v dnešní době. Tento článek podává stručný přehled této problematiky.

1. Ohmatanost logaritmických tabulek

Americký astronom a matematik Simon Newcomb v roce 1881 publikoval článek [21], v němž upozornil na skutečnost, že první stránky logaritmických tabulek v univerzitní knihovně jsou zřetelně opotřebovanější a špinavější než stránky na konci. Z toho usoudil, že uživatelé těchto tabulek – návštěvníci knihovny, vědci a studenti přírodovědných i společenských oborů – se při své práci častěji setkávají s čísly začínajícími číslicí 1 nebo 2, jejichž logaritmy jsou uvedeny v přední části tabulek, než s těmi, které začínají číslicí 8 nebo 9.

Na první pohled se zdá přirozené předpokládat, že první platná číslice čísel, s nimiž se lidé setkávají, bude se stejnou pravděpodobností jednička, dvojka, . . . , devítka. Newcombovo tvrzení je ale v rozporu s touto intuitivní představou.

Nedůvěřivému čtenáři rovnou nabídneme příklad. V ČR bylo k 1. 1. 2018 celkem 6 254 obcí s kladným počtem obyvatel a, poněkud překvapivě, čtyři obce s nulovým [6]. Tyto z následujícího zkoumání vyřadíme. U obcí s kladným počtem obyvatel si zapíšeme první platnou číslici tohoto počtu. Četnosti výskytu číslic 1, 2, . . . , 9 na prvním platném místě jsou po řadě 1 821, 1 150, 796, 593, 511, 420, 365, 310, 288. Tento příklad snad přesvědčí i skeptického čtenáře, že v některých souborech čísel se na prvním platném místě vyskytují nízké číslice častěji než vysoké.

Ve 30. letech 20. století si Frank Benford povšiml stejného nerovnoměrného opotřebení stránek logaritmických tabulek a zřejmě bez znalosti [21] vydal v roce 1938 vlastní článek [2]. Díky tomu začal být fakt, že v mnoha případech nejsou první platné číslice rozděleny rovnoměrně, označován jako Benfordův zákon, v originále *Benford's law*. Tento pojem by v některých situacích bylo vhodnější přeložit jako Benfordovo rozdělení, v angličtině jsou však tyto pojmy obvykle rozlišitelné jen podle kontextu a ještě k tomu obtížně vzhledem k volnému vyjadřování mnoha autorů. Další obvyklá označení jsou například *first-digit law* nebo *significant digit law*. V tomto článku budeme používat pojem *Benfordův zákon* pro označení fenoménu nerovnoměrného výskytu prvních platných číslic a termín *Benfordovo rozdělení* pro konkrétní pravděpodobnostní rozdělení popsané v definici 2.3.

RNDr. Jiří Dvořák, Ph.D., Katedra pravděpodobnosti a matematické statistiky MFF UK, Sokolovská 83, 186 75 Praha 8, e-mail: `dvorak@karlin.mff.cuni.cz`

Výborný přehled problematiky poskytuje například starší článek [26], z novějších pak stojí za pozornost článek [11]. Dodejme, že v české literatuře se tomuto tématu věnoval už Pavel Kantorek [18]. Dalším zdrojem může být práce [9], z níž zde z velké části čerpáme.

2. Benfordův zákon a Benfordovo rozdělení

V celém článku budeme pracovat pouze s kladnými čísly. V následujících kapitolách, s výjimkou kapitoly 7, si budeme představovat, že zkoumané hodnoty jsou náhodné a pocházejí z nějakého pravděpodobnostního rozdělení. Před dalším výkladem definujme několik užitečných pojmů.

Definice 2.1. *Mantisová funkce (při vyjádření čísel v desítkové soustavě) je funkce* $m\colon (0, \infty) \to [1, 10)$ *taková, že pro každé* $x \in (0, \infty)$ *platí* $x = m(x) \cdot 10^n$ *pro nějaké* $n \in \mathbb{Z}$*. Hodnotě* $m(x)$ *říkáme mantisa čísla* x*.*

Takové číslo $m(x)$ je v intervalu $[1, 10)$ jediné, a proto je definice korektní.

Definice 2.2. $D_k\colon (0, \infty) \to \{1, 2, \ldots, 9\}$ *je pro přirozené číslo* k *funkce určující* k*-tou platnou číslici argumentu při vyjádření v desítkové soustavě.*

Například tedy platí $D_1(\pi) = D_1(10\pi) = 3$, $D_2(\pi) = 1$, $m(100\pi) = \pi$ a podobně.

Definice 2.3. *Nechť* X *je náhodná veličina s kladnými hodnotami. Její mantisa má Benfordovo rozdělení, pokud platí*

$$\mathbb{P}(m(X) < t) = \log_{10} t, \quad t \in [1, 10]. \tag{1}$$

V tuto chvíli tedy můžeme formulovat Benfordův zákon tak, že v některých přirozeně se vyskytujících souborech číselných údajů je rozdělení jejich mantisy Benfordovo. Samozřejmě jde spíše o pozorování, ne zákon, budeme se však držet tohoto tradičního označení.

Simon Newcomb i Frank Benford dospěli k vyjádření, které odpovídá vzorci (1), každý však jinou cestou. Newcombovy úvahy v [21] jdou popsat takto: každé kladné reálné číslo x lze zapsat ve tvaru $x = 10^s$ pro nějaké $s \in \mathbb{R}$. Protože celá část čísla s neovlivní první platnou číslici $D_1(x)$ ani mantisu $m(x)$, stačí uvažovat pouze $s \bmod 1$. Po krátké úvaze dochází k závěru, že v případě „v přírodě se vyskytujících čísel“ má $s \bmod 1$ rovnoměrné rozdělení na intervalu $(0, 1)$.

Nechť tedy S je náhodná veličina s rovnoměrným rozdělením na intervalu $(0, 1)$, která odpovídá výše uvedeným hodnotám $s \bmod 1$. Pak rozdělení náhodné veličiny $Y = 10^S$ je následující:

$$\mathbb{P}(Y < y) = \mathbb{P}(10^S < y) = \mathbb{P}(S < \log_{10} y) = \log_{10} y, \quad y \in [1, 10],$$

protože $\mathbb{P}(S < s) = s$, $s \in [0, 1]$. Náhodná veličina Y má tedy Benfordovo rozdělení. Pro pozorovanou hodnotu $Y = y$ platí $y = m(y)$, protože Y nabývá pouze hodnot z intervalu $(1, 10)$. Dále je $y = m(y) = m(10^s) = m(10^{s \bmod 1}) = m(x)$. To znamená, že veličina Y popisuje mantisu původně uvažovaného čísla x. Podle Newcomba se tedy množina „všech čísel vyskytujících se v přírodě“ řídí Benfordovým rozdělením, resp. v naší terminologii se tímto rozdělením řídí jejich mantisy, ale to nebudeme v dalším textu explicitně rozlišovat.

Na rozdíl od Newcomba Benford v článku [2] svá tvrzení založil na empirických pozorováních. Shromáždil číselné údaje z různých zdrojů a oborů, například plochy povodí řek, měrné skupenské teplo chemických sloučenin, čísla vyskytující se na titulní stránce novin a další. Dohromady zpracoval více než 20 000 údajů a ukázal, že první číslice se nevyskytují všechny stejně často.

Když hledal jednoduchý vzorec, kterým by mohl popsat rozdělení prvních platných číslic ve svých datech, dospěl k výrazu

$$\mathbb{P}(D_1(X) = d) = \log_{10}\left(1 + \frac{1}{d}\right), \quad d = 1, 2, \ldots, 9. \tag{2}$$

Ten je ovšem důsledkem vztahu (1), protože pro $d = 1, 2, \ldots, 9$ je

$$\begin{aligned}\mathbb{P}(D_1(X) = d) &= \mathbb{P}(d \leqq m(X) < d+1) = \mathbb{P}(m(X) < d+1) - \mathbb{P}(m(X) < d) = \\ &= \log_{10}(d+1) - \log_{10} d = \log_{10}\left(\frac{d+1}{d}\right).\end{aligned}$$

Podobně se ukáže, že ze vztahu (1) plyne

$$\mathbb{P}(D_2(X) = d) = \sum_{k=1}^{9} \log_{10}\left(1 + \frac{1}{10k + d}\right), \quad d = 0, 1, \ldots, 9. \tag{3}$$

To také odpovídá hodnotám, které udává Newcomb [21]. Následující tabulka ukazuje číselné hodnoty pravděpodobností, že se jednotlivé číslice objeví na prvním, resp. druhém platném místě. Hodnoty jsou určeny vzorci (2) a (3) a jsou zaokrouhleny na 3 desetinná místa.

Číslice	0	1	2	3	4	5	6	7	8	9
první		0,301	0,176	0,125	0,097	0,079	0,067	0,058	0,051	0,046
druhá	0,120	0,114	0,109	0,104	0,100	0,097	0,093	0,090	0,088	0,085

Vztah (1) dokonce určuje sdružené rozdělení veličin $D_1(X), D_2(X), \ldots, D_k(X)$. Pro všechna $k \in \mathbb{N}$, $d_1 \in 1, 2, \ldots, 9$, $d_j \in 0, 1, \ldots, 9$, $j = 2, \ldots, k$, platí

$$\mathbb{P}(D_1(X) = d_1, \ldots, D_k(X) = d_k) = \log_{10}\left(1 + \frac{1}{\sum\limits_{i=1}^{k} d_i \cdot 10^{k-i}}\right). \tag{4}$$

Z toho ovšem krátkým výpočtem vyplyne, že jednotlivé číslice na sobě nejsou nezávislé. Platí například, že pravděpodobnost jevu $\{D_2(X) = 2\}$ je přibližně 0,109, ovšem podmíněná pravděpodobnost jevu $\{D_2(X) = 2\}$, když víme, že nastal jev $\{D_1(X) = 1\}$, je přibližně 0,116. Tyto dva jevy se tedy navzájem ovlivňují.

Navíc stojí za povšimnutí, že s rostoucím k se rozdělení veličiny $D_k(X)$ blíží rovnoměrnému rozdělení na množině $\{0, 1, \ldots, 9\}$, tedy $\mathbb{P}(D_k(X) = d) \to 1/10$, $k \to \infty$, $d \in \{0, 1, \ldots, 9\}$. Díky vyjádření (4) se dá totiž psát

$$\mathbb{P}(D_k(X) = d_k) = \sum_{d_1=1}^{9} \sum_{d_2=0}^{9} \cdots \sum_{d_{k-1}=0}^{9} \log_{10} \left(\frac{\sum\limits_{i=1}^{k} d_i \cdot 10^{k-i} + 1}{\sum\limits_{i=1}^{k} d_i \cdot 10^{k-i}} \right) \tag{5}$$

a dále pro $d = 0, 1, 2, \ldots, 8$, snadno odhadnout rozdíl $\mathbb{P}(D_k(X) = d) - \mathbb{P}(D_k(X) = d + 1)$ zdola nulou a shora výrazem konvergujícím k nule pro k rostoucí nade všechny meze. Podrobnosti lze najít v práci [9].

Formannův článek [11] pomocí simulací zkoumá, jak dobře odpovídají pravděpodobnostem (5) hodnoty pocházející z několika běžných pravděpodobnostních rozdělení. Zajímavým závěrem tohoto článku je, že obecně se míra shody zlepší, pokud místo hodnot náhodné veličiny X mající dané rozdělení uvažujeme podíly X/Y, kde X, Y jsou nezávislé náhodné veličiny s tímto rozdělením. Zkoumání těchto podílů je motivováno některými úvahami z Newcombova článku [21].

3. Benfordova data

V článku [2] předkládá Benford výsledky své několikaleté práce, během níž sbíral údaje „z tolika různých oblastí, kolik čas a energie dovolily“. Dohromady zpracoval více než 20 000 údajů z 20 různých oblastí, což lze ve třicátých letech 20. století považovat za úctyhodný výkon.

Benford hledal rozdělení prvních platných číslic v „přirozeně se vyskytujících“ číselných údajích a uvádí pouze relativní četnosti výskytu prvních číslic ve všech 20 zkoumaných kategoriích spolu s informací, kolik údajů bylo do jednotlivých kategorií zahrnuto. Pro každou číslici tak Benford získal 20 různých relativních četností (pro každou kategorii jednu hodnotu) a z nich vypočítal aritmetický průměr.

Na těchto průměrných hodnotách pak Benfordovi bylo nápadné, že se blíží hodnotám $p_i = \log_{10}\left(1 + \frac{1}{i}\right)$, $i = 1, \ldots, 9$, a dále se ve svém článku pokusil zdůvodnit, proč právě tyto hodnoty jsou „ty správné“.

Kdyby Benford nepracoval s průměrnými relativními četnostmi, ale místo toho by zpracoval všechny údaje najednou v jednom velkém souboru, získal by odlišné relativní četnosti. Ty se sice více liší od hodnot p_i, přesto jsou jim stále dost blízké.

Zvláštní je, že ve všech skupinách v Benfordově tabulce v článku [2] se relativní četnosti nasčítají přesně na 1. Díky zaokrouhlování relativních četností na tři desetinná místa by mohlo dojít k tomu, že se v některé skupině nenasčítají přesně na 1, ale v případě Benfordem předložených dat dává součet 1 všech dvacet skupin.

Upozornili na to Diaconis a Freedman v článku [8], v němž se zabývají právě otázkou chování součtu zaokrouhlených procentuálních údajů. Při použití jejich modelů vyjde pravděpodobnost, že se relativní četnosti v každé z dvaceti skupin nasčítají přesně na 1, „astronomicky malá“ (zhruba $1 : 2^{20}$).

Vyvozují z toho, že se Benfordovy četnosti neřídí žádným z jejich modelů, a to vede k podezření, že Benford s částí dat manipuloval, aby docílil lepší shody s hodnotami $\log_{10}\left(1+\frac{1}{i}\right)$. Například v prvním řádku Benfordovy tabulky (viz [2]) je podíl čísel začínajících na 7 uveden jako 5,5 % z celkového počtu 335 údajů. Snadný výpočet ukazuje, že počet čísel začínajících na 7 musel být 18 nebo 19, ale $\frac{18}{335}$ se zaokrouhlí na 5,4 % a $\frac{19}{335}$ na 5,7 %. Žádná skutečná hodnota tedy nemohla vést k údaji 5,5 %.

Ony chybně zaokrouhlené hodnoty jsou skutečně blíž požadovaným číslům než hodnoty zaokrouhlené správně, ale i neupravená data, která je možné z tabulky v článku [2] rekonstruovat postupem ukázaným v předchozím odstavci, se poměrně dobře shodují s předpokládanými četnostmi.

4. Benfordův zákon v praxi

Od vydání Benfordova článku [2] se další autoři snažili ukázat, které soubory číselných údajů se řídí Benfordovým rozdělením. Například rozsáhlé soubory čísel obsažených v účetních záznamech nebo daňových přiznáních firem a jednotlivců se obvykle Benfordovým rozdělením řídí velmi přesně, viz Nigrini [22]. To vede k zajímavým aplikacím, o kterých pojednáme níže.

Ley [20] zjistil, že posloupnost jednodenních výnosů na indexu Dow-Jones Industrial Average (DJIA) a Standard and Poor's (S&P) je v dobré shodě s Benfordovým rozdělením. Jde o řádově desetitisíce údajů z let 1900–1993 (DJIA), resp. 1926–1993 (S&P). Z podobné oblasti uveďme ještě například Gilesův článek [12], který ukazuje, že vítězné nabídky v aukcích na serveru eBay odpovídají Benfordovu rozdělení.

Varian [30] popisuje, jak se koncem 60. let zabýval vývojem počítačových modelů předpovídajících ekonomický vývoj v Sanfranciské zátoce do roku 1990. Výstupy srovnával s Benfordovým rozdělením na základě úvahy, že když se vstupní data modelu shodují s Benfordovým rozdělením, měly by se jím řídit i výstupy, pokud je model „rozumný“.

Po rozboru výstupů svého modelu dospěl Varian k tomu, že „data jsou v poměrně dobré shodě s Benfordovým rozdělením“. Konkrétně nejlepší shodu vykazovala ta část modelu, jejíž vstupní data pocházela ze sčítání lidu – to je obecně považováno za přesnější než jiné zdroje údajů.

Benfordovým rozdělením se také poměrně přesně řídí i údaje o odběru elektřiny jednotlivými spotřebiteli na Šalamounových ostrovech, jak uvádí Raimi [26]. Jeho data pochází z října 1969 a zajímavý je jejich původ. Údaje nezískával autor sám, ale už zpracované mu byly zaslány ředitelem energetické společnosti ve městě Honiara na Šalamounových ostrovech, který byl zaujat problematikou Benfordova zákona poté, co si přečetl populárně laděný Raimiho článek [25] v časopisu Scientific American.

Příklady podobných výsledků lze nalézt nejen v ekonomii, ale třeba i ve fyzice. Burke a Kincanon [5] se rozhodli prozkoumat, zda se fyzikální konstanty řídí Benfordovým rozdělením. Vzali konstanty uvedené na vnitřním přebalu úvodní vysokoškolské učebnice fyziky a spočítali četnosti výskytu prvních platných číslic při vyjádření čísel v jednotkách SI (pro číslice 1 až 9 postupně 8, 2, 1, 0, 2, 3, 0, 2 a 2) a v britských jednotkách (7, 2, 3, 1, 1, 3, 0, 1, 2). Pozoruhodné je, že při tak malém rozsahu souboru (pouze 20 údajů) je vidět vůbec nějaká shoda s Benfordovým rozdělením. Při vyjádření čísel v soustavě SI a v britských jednotkách má jednička zdaleka nejvyšší četnost výskytu na prvním platném místě. Pro ostatní číslice je však případný trend zcela překryt šumem.

Buck, Merchant a Perez [4] vyvinuli metodu, jak počítat poločas rozpadu nestabilních nuklidů podléhajících alfa rozpadu a při porovnávání vypočtených hodnot s experimentálně zjištěnými údaji si všimli, že rozdělení prvních platných číslic je nerovnoměrné. V době vzniku práce [4] bylo známo 477 nuklidů podléhajících alfa rozpadu. Jejich poločasy rozpadu pokrývají interval od zhruba 10^{-6} sekundy až po 10^{15} let. Při podrobnějším rozboru Buck a kol. zjistili, že četnosti výskytu jednotlivých číslic na prvním platném místě vypočtených i naměřených hodnot se poměrně dobře shoduje s četnostmi předpovězenými Benfordovým rozdělením. U vypočtených hodnot zkoumali také rozdělení druhých platných číslic a i zde data vykázala vysokou míru shody s Benfordovým rozdělením. Naměřené hodnoty jsou často zaznamenány pouze s přesností na jednu platnou cifru a výsledky získané rozborem druhých platných číslic by nebyly směrodatné.

Na druhou stranu ne v každé sadě přirozeně se vyskytujících údajů je rozdělení prvních platných číslic Benfordovo. Raimi [26] uvádí pěkný příklad, byť dnes již také z kategorie pro pamětníky: telefonní seznam. U čísel na stránce vybrané z místního telefonního seznamu bude na prvním platném místě výrazně převažovat jedna číslice, ostatní budou zastoupeny minimálně.

5. Pokusy o vysvětlení

V průběhu let se objevilo mnoho pokusů o vysvětlení Benfordova zákona, většinou čistě matematické povahy. Často se autor snažil ukázat, že množina kladných reálných (případně přirozených) čísel splňuje (1) a z toho učinit závěr, že fenomén nerovnoměrného rozdělení první číslice je jednoduše vlastností používaného číselného systému.

Cílem je na $\mathbb{N}$, resp. na $\mathbb{R}^+$ zavést pravděpodobnostní míru P, která bude určovat rozdělení první platné číslice, případně mantisy. Typicky je prvním krokem určit pravděpodobnost, že přirozené číslo n má první platnou číslici 1, tj. patří do množiny $\{D_1 = 1\} = \{k \in \mathbb{N}, D_1(k) = 1\}$. Nechť $\alpha(n) = 1$, pokud $D_1(n) = 1$, a $\alpha(n) = 0$ jinak. Pak by se zdálo přirozené definovat

$$P(D_1 = 1) = \lim_{n\to\infty} \frac{1}{n} \sum_{k=1}^{n} \alpha(k).$$

Tato limita ale neexistuje, hodnoty výrazu oscilují mezi přibližně $\dfrac{1}{9}$ pro $n = 10^k, k \in \mathbb{N}$, a $\dfrac{5}{9}$ pro $n = 2 \cdot 10^k$, $k \in \mathbb{N}$.

To vedlo k použití různých zobecněných sčítacích metod, které měly množině $\{D_1 = 1\}$ přiřadit „správnou“ pravděpodobnost $\log_{10} 2$. V tomto ohledu je zajímavá Flehingerova práce [10]. Používá iterace Cesàrovy sčítací metody

$$\begin{aligned}\alpha_1(k) &= \frac{1}{k}\sum_{i=1}^{k}\alpha(i),\\ \alpha_t(k) &= \frac{1}{k}\sum_{i=1}^{k}\alpha_{t-1}(i)\end{aligned}$$

a dokáže, že ačkoliv funkce $\alpha_t(k)$ jsou pořád oscilující, celý iterační proces konverguje ke kýžené hodnotě ve smyslu

$$\lim_{t\to\infty}\liminf_{k\to\infty}\alpha_t(k) = \lim_{t\to\infty}\limsup_{k\to\infty}\alpha_t(k) = \log_{10} 2.$$

Žádný z těchto postupů nevede k uspokojivému výsledku i přesto, že uvažované sčítací metody jsou regulární (pro konvergentní řady dávají jako součet limitu částečných součtů) a množinám typu $\{D_1 = 1\}$ přiřazují požadovanou pravděpodobnost. Jak totiž píše Raimi [26], existuje mnoho regulárních sčítacích metod, které jim naopak přiřazují jiné pravděpodobnosti, a nelze *a priori* rozhodnout, která sčítací metoda je „správná“.

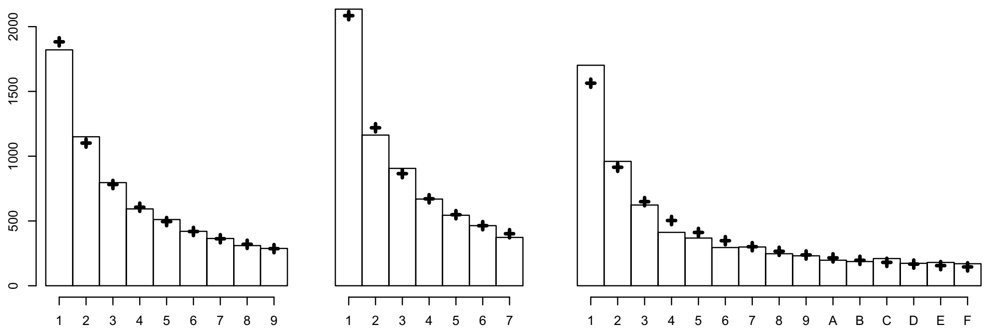
Ve spojitém případě, kdy byla místo přirozených čísel uvažována kladná reálná čísla, byly použity různé integrační metody, metody Fourierovy analýzy i teorie Banachových měr. Žádný postup však nevyústil v zavedení požadované pravděpodobnosti na $\mathbb{R}^+$ ve smyslu σ-aditivní množinové funkce. Raimi [26] podává obsáhlý přehled výsledků v diskrétním i spojitém případě.

V souvislosti s Benfordovým zákonem jsou často zmiňovány dvě další hypotézy: *invariance vzhledem ke změně základu* a *invariance vzhledem ke změně měřítka*. První z nich říká, že by se data, která se řídí Benfordovým rozdělením, měla řídit jeho obdobou i při vyjádření čísel v soustavě o libovolném jiném základu, nejen v desítkové soustavě. V soustavě o základu b je potom obdobou vztahu (1) vztah

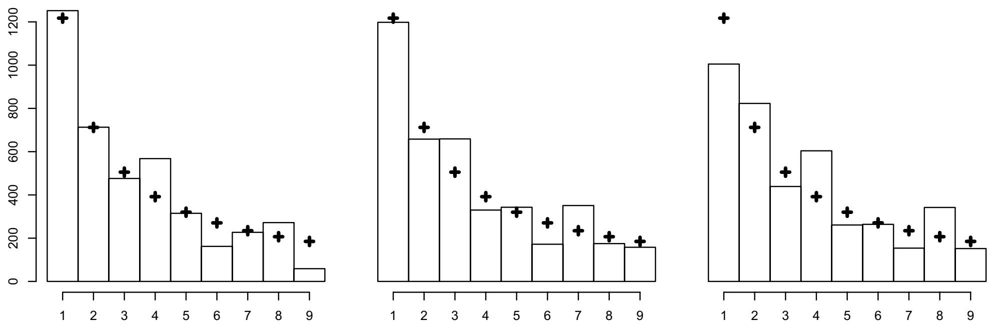
$$\mathbb{P}(m_b(X) < t) = \log_b t, \quad t \in [1, b], \tag{6}$$

kde symbol $m_b(x)$ označuje mantisu čísla x při vyjádření v soustavě o základu b. Jako ilustraci jsme převedli počty obyvatel v jednotlivých obcích České republiky do osmičkové, resp. šestnáctkové soustavy. Četnosti výskytu prvních platných číslic v tomto souboru dat pak ukazují histogramy na obrázku 1. Míra shody s pravděpodobnostmi určenými vztahem (6) je až překvapivá. Aparát potřebný ke studiu invariance vzhledem ke změně základu je poměrně technický a v tomto článku se pro přehlednost už nebudeme tímto tématem dále zabývat. Podrobnosti je možné nalézt například v článku [15] nebo práci [9].

Uvažujme soubor číselných údajů, vybíraných náhodně z nějakého pravděpodobnostního rozdělení μ. Invariance vzhledem ke změně měřítka je požadavek, aby μ nezáviselo na tom, v jakých jednotkách jsou údaje vyjádřeny, tedy aby se neměnilo po přenásobení hodnot libovolnou kladnou konstantou. Řada autorů se snažila dokázat, že invariance vzhledem ke změně měřítka implikuje Benfordovo rozdělení mantis. Pro ilustraci uvažme celkovou výši škod při jednotlivých dopravních nehodách na dálnicích v České republice v roce 2018 [24]. Četnosti prvních platných číslic v tomto souboru



Obr. 1. Histogram prvních platných číslic počtů obyvatel v jednotlivých obcích České republiky ke dni 1. 1. 2018. Vlevo: při vyjádření v desítkové soustavě; uprostřed: při vyjádření v osmičkové soustavě; vpravo: při vyjádření v šestnáctkové soustavě. Svislá osa je shodná pro všechny obrázky. Křížky znázorňují očekávané hodnoty odpovídající Benfordovu rozdělení mantis při vyjádření v soustavě o daném základu



Obr. 2. Histogram prvních platných číslic v souboru hodnot udávajících celkovou výši škody při dopravních nehodách na dálnicích v České republice v roce 2018. Vlevo: při vyjádření v českých korunách; uprostřed: při vyjádření v eurech; vpravo: při vyjádření v amerických dolarech. Pro konverzi měn byl použit kurz vyhlášený Českou národní bankou dne 18. 1. 2019. Svislá osa je shodná pro všechny obrázky. Křížky znázorňují očekávané hodnoty odpovídající Benfordovu rozdělení mantis při vyjádření v desítkové soustavě

dat při vyjádření v českých korunách, eurech a amerických dolarech ukazují histogramy na obrázku 2. Opět je míra shody s Benfordovým rozdělením vysoká.

V souvislosti s invariancí vzhledem ke změně měřítka například Pinkham [23] uvažuje hypotetický „soubor všech čísel“ ve smyslu hodnot vyhledávaných v logaritmických tabulkách a distribuční funkci F určující jejich rozdělení. Za předpokladu invariance vzhledem ke změně měřítka a spojitosti F (aby se žádné konkrétní číslo z uvažovaného souboru všech čísel vyhledávaných v logaritmických tabulkách neobjevovalo s kladnou pravděpodobností) pak ukáže, že rozdělení mantis je Benfordovo.

Problém ovšem je, že na $(0, \infty)$ neexistuje borelovská pravděpodobnostní míra invariantní vzhledem ke změně měřítka. Poznamenejme, že borelovská je taková míra, zhruba řečeno, která dovede měřit otevřené intervaly, jejich konečná či spočetná sjednocení a doplňky takových množin. Pokud by μ byla borelovská pravděpodobnostní míra na $(0, \infty)$ invariantní vzhledem ke změně měřítka, musela by mimo jiné splňovat $\mu((0, 1)) = \mu((0, s))$, $\forall s \in (0, \infty)$. Pak by ale funkce $F(x) = \mu((0, x))$, $x > 0$, byla konstantní. Protože je ale μ pravděpodobnostní míra, požadujeme $\lim_{x \to 0+} F(x) = 0$ a $\lim_{x \to \infty} F(x) = 1$, a to je spor. Způsob, jak zdůvodnit platnost Benfordova zákona, je třeba hledat jinde.

6. Moderní pohled

Protože už Newcomb formuloval svá tvrzení jako pravděpodobnostní problém – „jaká je pravděpodobnost, že vybrané číslo bude mít na prvním platném místě číslici d“ – snažil se v devadesátých letech dvacátého století Theodore Hill zasadit Benfordův zákon do rámce moderní teorie pravděpodobnosti, viz například práce [14], [15]. Výchozím bodem Hillových úvah je názor, že pracovat s borelovskými mírami je pro daný účel nevhodné, viz předchozí odstavec. Místo toho pracuje s mírami, které dovedou měřit množiny typu $\bigcup_{n=-\infty}^{\infty} [a, b) \cdot 10^n$, $1 \leqq a < b < 10$, jejich konečná či spočetná sjednocení a doplňky takových množin (tento systém množin označíme $\mathcal{M}$). To stačí pro práci s mantisami, proto budeme takovým mírám říkat mantisové. Nepožaduje však, aby tyto míry dovedly měřit otevřené intervaly, čímž mnoho problémů odstraní. Tyto míry tedy nemusí být borelovské.

Definice 6.1. *Pravděpodobnostní mantisová míra P je invariantní vzhledem ke změně měřítka, pokud $P(S) = P(sS)$ pro všechna $s > 0$ a každou $S \in \mathcal{M}$, kde $sS = \{su, u \in S\}$.*

Pravděpodobnostní míra P je rozdělením náhodné veličiny X, pokud platí $\mathbb{P}(X \in S) = P(S)$ pro všechny přípustné množiny S. Pojmy „pravděpodobnostní míra“ a „rozdělení“ pro nás tedy splývají a předchozí definice umožňuje mluvit jak o mírách, tak rozděleních invariantních vzhledem ke změně měřítka. Vlastnost invariance vzhledem ke změně měřítka dokonce charakterizuje Benfordovo rozdělení (1), je to jediné rozdělení na mantisových množinách s takovou vlastností. Důkaz následující věty je možné nalézt v [15].

Věta 6.2. *Pravděpodobnostní mantisová míra P je invariantní vzhledem ke změně měřítka právě tehdy, když*

$$P\left(\bigcup_{n=-\infty}^{\infty} [1, t) \cdot 10^n\right) = \log_{10} t \quad \textit{pro všechna } t \in [1, 10). \tag{7}$$

Tato věta ukazuje jasnou souvislost Benfordova zákona s vlastností invariance vzhledem ke změně měřítka, ale nevysvětluje, proč se některé soubory čísel řídí Benfordovým rozdělením a jiné ne. V případě údajů sesbíraných Frankem Benfordem vykazovaly některé skupiny velmi dobrou shodu s Benfordovým rozdělením a některé, jako třeba odmocniny přirozených čísel, v podstatě žádnou. Nejblíže Benfordovu rozdělení byl ale souhrn všech 20 229 údajů z dvaceti různých skupin. Spíše než to, že by všechny údaje pocházely ze stejného rozdělení „všech konstant světa“ (Pinkham [23]), je vhodné předpokládat, že údaje pocházejí z různých rozdělení. Frank Benford v článku [2] uvádí, že se skutečně snažil „sesbírat údaje z co nejvíce různých oblastí“.

Tady přicházejí ke slovu *náhodné pravděpodobnostní míry*, viz např. [17]. Ty právě umožňují studovat situace, kdy ve zkoumaném vzorku nepocházejí všechny hodnoty $X_1, \ldots, X_n$ ze stejného rozdělení P, ale potenciálně každá z jiného rozdělení P_i, $i = 1, \ldots, n$. V tomto kontextu nepovažujeme rozdělení P_i za předem daná, ale naopak za náhodná. Jejich náhodnost je pak popsána právě náhodným rozdělením (náhodnou pravděpodobnostní mírou) $\mathcal{P}$. Potom pro množinu S je $\mathcal{P}(S)$ náhodná veličina. Podrobnosti je možné nalézt v [9]. Pro další výklad nám bude stačit vědět, že stejně jako pro náhodnou veličinu X můžeme definovat její střední hodnotu $\mathbb{E}X$ (pokud je $\mathbb{E}|X| < \infty$), můžeme pro náhodnou pravděpodobnostní míru $\mathcal{P}$ definovat její *očekávané rozdělení* $\mathbb{E}\mathcal{P}$, které popisuje průměrné chování náhodné míry $\mathcal{P}$.

Definice 6.3. *Očekávané rozdělení náhodné pravděpodobnostní mantisové míry $\mathcal{P}$ je pravděpodobnostní míra $\mathbb{E}\mathcal{P}$ taková, že*

$$(\mathbb{E}\mathcal{P})(S) = \mathbb{E}(\mathcal{P}(S)) \quad \forall S \in \mathcal{M}.$$

Následující definice formalizuje vlastnost číselných souborů, u kterých rozdělení mantisy nezávisí na použitých jednotkách.

Definice 6.4. *Posloupnost náhodných veličin $X_1, X_2, \ldots$ má hodnoty mantisy nezávislé na měřítku, pokud*

$$\frac{|\#\{i \leqq n\colon X_i \in S\} - \#\{i \leqq n\colon X_i \in sS\}|}{n} \xrightarrow{n\to\infty} 0 \text{ s pravděpodobností } 1$$

pro všechna $s > 0$ a všechna $S \in \mathcal{M}$.

Nyní zformulujeme limitní větu o chování prvních platných číslic. Nejzajímavější část věty říká následující: pokud jsou náhodně vybírána pravděpodobnostní rozdělení a z nich se berou náhodné veličiny tak, že celý proces nezvýhodňuje nějakou volbu jednotek, pak rozdělení mantis v celém souboru bude konvergovat k Benfordovu rozdělení. Tím tato věta pomáhá vysvětlit nebo předpovídat shodu souborů číselných údajů s Benfordovým rozdělením.

Věta 6.5. *Nechť $\mathcal{P}$ je náhodná pravděpodobnostní mantisová míra. Následující tvrzení jsou ekvivalentní:*

(i) *$\mathbb{E}\mathcal{P}$ je invariantní vzhledem ke změně měřítka,*

(ii) *$\mathbb{E}[\mathcal{P}(D_t)] = \log_{10} t$ pro všechna $t \in [1, 10)$, kde $D_t = \bigcup_{n=-\infty}^{\infty} [1, t) \cdot 10^n$ je množina kladných čísel, jejichž mantisa je v intervalu $[1, t)$,*

(iii) *posloupnost $X_1, X_2, \ldots$ generovaná $\mathcal{P}$ má hodnoty mantisy nezávislé na měřítku,*

(iv) *posloupnost $X_1, X_2, \ldots$ generovaná $\mathcal{P}$ splňuje*

$$\frac{\#\{i \leqq n\colon m(X_i) \in [1, t)\}}{n} \overset{n\to\infty}{\longrightarrow} \log_{10} t \quad \textit{s pravděpodobností } 1$$

pro všechna $t \in [1, 10)$.

Důkaz této věty je možné najít v článku [14] nebo práci [9]. Klíčovým bodem této věty je fakt, že nevyžaduje, aby konkrétní realizace $\mathcal{P}_i$ byly invariantní vzhledem ke změně měřítka (často se dokonce stává, že žádná realizace $\mathcal{P}_i$ nemá požadovanou vlastnost), ale pouze aby celý proces sbírání údajů *v průměru* nezvýhodňoval nějakou volbu jednotek.

Věta 6.5 pomáhá vysvětlit, proč se některé soubory číselných údajů řídí Benfordovým rozdělením lépe než jiné. Například na čísla vypsaná Frankem Benfordem z titulních stran místních novin (viz [2]) se dá dívat tak, že pocházejí z různých rozdělení, která spolu navzájem nesouvisí a tedy dohromady nezvýhodňují jednu volbu jednotek před ostatními. V kontextu uvedené věty to znamená, že rozdělení mantis v tomto výběru se bude asymptoticky blížit Benfordovu rozdělení.

Pokud naopak rozdělení, z něhož zkoumané údaje pochází, preferuje konkrétní volbu jednotek, *nebude* se rozdělení mantis blížit Benfordovu. Přínos věty 6.5 spočívá právě v tom, že poskytuje kritérium pro rozhodování, zda se zkoumaný soubor bude či nebude řídit Benfordovým rozdělením.

7. Silně benfordovské posloupnosti

Některé číselné posloupnosti mají tu vlastnost, že se chování mantisy jejich prvků dá dobře popsat Benfordovým rozdělením, pokud se bere v úvahu velmi dlouhý úsek těchto posloupností. V takovém případě se dají označit jako *benfordovské.* Tuto problematiku přehledně shrnuje např. Raimi [26]. Zdůrazňujeme, že v této kapitole se budeme zabývat pouze nenáhodnými posloupnostmi a roli pravděpodobnosti určitého jevu zde přebírá relativní četnost výskytu tohoto jevu.

Definice 7.1. *Posloupnost $\{b_n\}$ reálných čísel z intervalu $[0, 1)$ se nazývá rovnoměrně distribuovaná na $[0, 1)$, pokud pro každý interval $[a, b) \subset [0, 1)$ platí*

$$\lim_{k\to\infty} \frac{1}{k} \sum_{n=1}^{k} \beta(n) = b - a,$$

kde $\beta(n) = 1$, pokud $b_n \in [a, b)$, a $\beta(n) = 0$ jinak.

Definice 7.2. *Nechť $\{a_n\}$ je posloupnost kladných reálných čísel a nechť $b_n = (\log_{10} a_n) \bmod 1$. Pokud je $\{b_n\}$ rovnoměrně distribuovaná na $[0, 1)$, nazývá se $\{a_n\}$ silně benfordovská posloupnost.*

Předchozí definice odpovídá Newcombovým úvahám popsaným v kapitole 2, tedy že v případě „v přírodě se vyskytujících čísel“ $Y = 10^S$ má $S \bmod 1$ rovnoměrné rozdělení na intervalu $[0, 1]$. Silně benfordovská posloupnost $\{a_n\}$ se řídí Benfordovým rozdělením v tom smyslu, že

$$\lim_{k\to\infty} \frac{1}{k} \sum_{n=1}^{k} \alpha_p(n) = \log_{10} p, \quad p \in [1, 10),$$

kde $\alpha_p(n) = 1$, pokud $a_n \in \{x \in \mathbb{R}^+, m(x) < p\} = \bigcup_{n=-\infty}^{\infty} [10^n, p \cdot 10^n)$, a $\alpha_p(n) =$ $= 0$ jinak. Jinými slovy, v dlouhých úsecích posloupnosti bude rozdělení mantis blízké Benfordovu.

Například posloupnost přirozených čísel $\mathbb{N}$ není silně benfordovská posloupnost, protože uvedená limita neexistuje (viz kapitolu 5), ale geometrické posloupnosti $\{ar^n\}$ jsou, pokud r není racionální mocninou čísla 10. Pokud $a_n = ar^n$, je

$$b_n = (\log_{10} a_n) \bmod 1 = (\log_{10} a + n \log_{10} r) \bmod 1.$$

Přitom platí, že aritmetické posloupnosti s iracionální diferencí jsou (bráno mod 1) rovnoměrně distribuované na intervalu $[0, 1)$, viz Hardy, Wright [13]. Právě požadavek, aby diference v posloupnosti $\{b_n\}$ byla iracionální, tedy $\log_{10} r \notin \mathbb{Q}$, vede k tomu, že posloupnost $\{ar^n\}$ je silně benfordovská, právě když r není racionální mocninou 10.

Geometrické posloupnosti ale nejsou jedinými silně benfordovskými posloupnostmi. Tuto vlastnost mají i tzv. asymptoticky geometrické posloupnosti.

Definice 7.3. *Posloupnost $\{a_n\}$ se nazývá asymptoticky geometrická, pokud existuje geometrická posloupnost $\{ar^n\}$ taková, že*

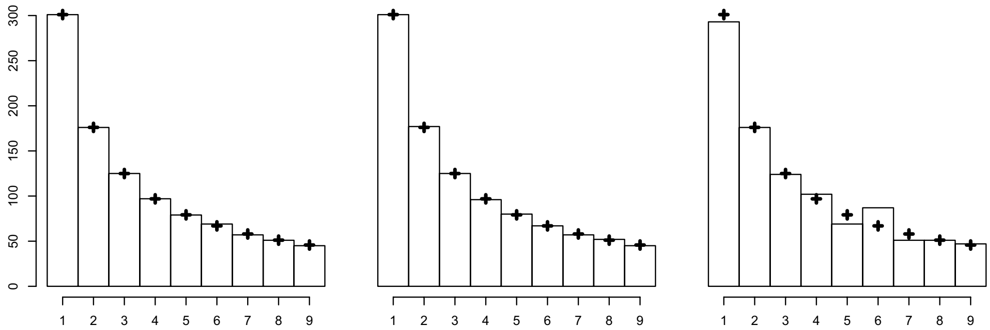
$$\lim_{n\to\infty} \frac{a_n}{ar^n} = 1.$$

V tomto případě $\log_{10} a_n - n \log_{10} r$ konverguje k nule a $\log_{10} a_n \bmod 1$ je stejně rovnoměrně distribuovaná jako $n \log_{10} r \bmod 1$. Jinými slovy, pokud $\log_{10} r \notin \mathbb{Q}$, je asymptoticky geometrická posloupnost silně benfordovská.

Jako příklad asymptoticky geometrické posloupnosti Raimi [26] uvádí Fibonacciho posloupnost. V jejím případě je limitním poměrem r z definice 7.3 zlatý řez, $r =$ $= \frac{1+\sqrt{5}}{2}$. Navíc $\log_{10} \frac{1+\sqrt{5}}{2}$ je iracionální číslo, a proto je Fibonacciho posloupnost silně benfordovská. Dalším příkladem silně benfordovské posloupnosti je $\{n!\}$, jak ukazuje Diaconis [7].

Histogramy četností výskytu prvních číslic mezi prvními 1 000 hodnotami posloupností 2^n, Fibonacciho posloupnosti a posloupnosti $n!$ jsou na obrázku 3. Shoda s Benfordovým rozdělením je velmi vysoká. Pokud bychom uvažovali delší úsek těchto posloupností, byla by shoda ještě lepší a vizuálně by zaznamenané hodnoty nešly odlišit od očekávaných. Proto zde volíme spíše krátký počáteční úsek uvedených posloupností.

Existuje také pojem slabě benfordovské posloupnosti. Ten je založen na využití jiné sčítací metody, než byla použita v definici 7.1. Podrobnosti je možné nalézt například v článku [26] nebo práci [9].



Obr. 3. Histogram prvních platných číslic v souboru prvních 1 000 hodnot vybraných posloupností. Vlevo: geometrická posloupnost 2^n; uprostřed: Fibonacciho posloupnost; vpravo: posloupnost $n!$. Svislá osa je shodná pro všechny obrázky. Křížky znázorňují očekávané hodnoty odpovídající Benfordovu rozdělení mantis při vyjádření v desítkové soustavě

8. Aplikace Benfordova zákona

V literatuře se objevuje několik návrhů na využití Benfordova zákona v praxi. Například Schatte [28] uvádí, že během dlouhých počítačových výpočtů v aritmetice s pohyblivou čárkou (tzv. *floating-point* aritmetika) mají mantisy zpracovávaných čísel přibližně Benfordovo rozdělení. Tuto informaci je možné využít při analýze zaokrouhlovacích chyb nebo při návrhu rychlejších algoritmů.

Knuth [19] předkládá určité výpočty, kterými zdůvodňuje předpoklad, že rozdělení mantisy ve vstupních datech jeho počítačových výpočtů je Benfordovo. Kromě jiného pak srovnává podle různých kritérií výhodnost používání dvojkové a šestnáctkové soustavy (s ohledem na předpokládané rozdělení mantis vstupních dat).

Z hlediska přesnosti výpočtů je vhodnější použití binární aritmetiky, horní odhad relativní chyby vzniklé zaokrouhlením je totiž poloviční než při použití šestnáctkové soustavy.

Naproti tomu výpočty prováděné v hexadecimální aritmetice probíhají o něco rychleji, protože není třeba tak často čísla normalizovat (posouvat desetinnou čárku). Cenou za rychlost je tedy nižší přesnost a naopak.

Techniky založené na Benfordovu zákonu je možné využít i při odhalování nesrovnalostí v účetních záznamech, například daňových úniků nebo nesprávného proplácení faktur. Jednu takovou metodu předkládá Nigrini v článku [22]. Vychází z pozorování, že se soubor čísel z rozsáhlých účetních záznamů (zpracovaných bez chyb a neoprávněné manipulace s daty) velmi dobře řídí Benfordovým rozdělením. To odpovídá výsledkům uvedeným ve větě 6.5, protože účetní záznamy lze považovat za směs údajů pocházejících z různých rozdělení. Pokud prověřovaný soubor vykazuje velké odchylky od předpokládaného rozdělení mantis, je označen za podezřelý a Nigrini doporučuje podrobit dotyčnou společnost auditu, který podezření potvrdí nebo vyvrátí. Může totiž jít o náhodnou odchylku, byť je její pravděpodobnost malá.

Při rozhodování, zda se konkrétní (napozorovaná) data řídí Benfordovým rozdělením, se velmi dobře uplatní Pearsonův χ^2 test dobré shody, viz například knížku Anděl [1]. Tento test se obecně používá k testování hypotézy, že vektor četností výskytu určitých jevů má multinomické rozdělení s předpokládanými parametry. Zde jde o četnosti výskytu prvních platných číslic a pravděpodobnosti jsou určeny vzorcem (2).

Nigrini však navrhuje jednodušší postup, pomocí něhož dokáže seřadit datové soubory od nejpodezřelejších (nejmenší míra shody s Benfordovým rozdělením) k nejméně podezřelým. Nedává však žádnou kritickou hodnotu, za níž jsou data podezřelá až příliš. Jeho postup je přesto užitečný například v situaci finančního úřadu, který nemůže detailně prověřit údaje všech firem, ale tato metoda mu poskytuje možnost rozhodnout, které společnosti podrobit auditu.

Hill [14] uvádí, že díky softwaru, založenému mimo jiné na Nigriniho postupu a na jehož vývoji se sám Nigrini podílel, bylo odhaleno sedm newyorských společností, které byly později obviněny z daňových podvodů.

Dokonce i na úrovni států, nejen firem, dochází k úpravám zveřejňovaných dat. To ukazuje například článek Rauch a kol. [27] z roku 2011, který zkoumá makroekonomická data, reportovaná členskými státy EU, vztahující se ke státnímu dluhu. Podle tohoto článku vykazuje Řecko největší odchylky od Benfordova rozdělení ze všech zemí platících eurem. Ze zpětného pohledu to není vůbec překvapivé. Práce [27] dokonce vyvolala ohlas i v českém tisku, viz článek [3].

Podobné techniky zkoumání shody dat s Benfordovým rozdělením byly také použity například při hledání „zvláštností“ ve financování politických kampaní v USA [29] nebo při zjišťování, zda si tazatelé v dotazníkových šetřeních odpovědi respondentů nevymýšlejí sami [16].

Podobných aplikací je možné vymyslet celou řadu, všechny ale spoléhají na to, že kontrolovaný nezná postupy používané kontrolorem. Se vzrůstající informovaností širší veřejnosti o fenoménu nerovnoměrného rozdělení prvních platných číslic samozřejmě dojde k adaptaci „kreativních účetních“ na novou situaci a tato metoda ztratí svůj význam.

Poděkování. Autor děkuje doc. RNDr. Zdeňku Hlávkovi, Ph.D., za jeho čas a cenné připomínky při vedení bakalářské práce [9].

Literatura

[1] Anděl, J.: *Základy matematické statistiky.* MatfyzPress, Praha, 2007.

[2] Benford, F.: *The law of anomalous numbers.* Proc. Amer. Philos. Soc. 78 (1938), 551–572.

[3] Břešťan, R.: *Evropa v krizi kouzlí s čísly. Největší triky předvádějí Řekové a Rumuni.* Ekonom, 17. 10. 2011.
Dostupné z: `https://ekonom.ihned.cz/c1-53243250-evropa-kouzli-s-cisly`

[4] Buck, B., Merchant, A., Perez, M.: *An illustration of Benford's first digit law using alpha decay half lives.* Eur. J. Phys. 14 (1993), 59–63.

[5] Burke, J., Kincanon, E.: *Benford's law and physical constants: the distribution of initial digits.* Amer. J. Phys. 59 (1991), 952.

[6] Český statistický úřad: Počet obyvatel v obcích – k 1. 1. 2018. Citováno 13. 1. 2019.
Dostupné z: `https://www.czso.cz/csu/czso/pocet-obyvatel-v-obcich-see2a5tx8j`

[7] Diaconis, P.: *The distribution of leading digits and uniform distribution* mod 1. Ann. Probab. 5 (1977), 72–81.

[8] Diaconis, P., Freedman, D.: *On rounding percentages.* J. Amer. Statist. Assoc. 74 (1979), 359–364.

[9] Dvořák, J.: *Benfordovo rozdělení.* Bakalářská práce. Matematicko-fyzikální fakulta, Univerzita Karlova, Praha, 2008.
Dostupné z: `https://dspace.cuni.cz/handle/20.500.11956/17007`

[10] Flehinger, B. J.: *On the probability that a random integer has initial digit A.* Amer. Math. Monthly 73 (1966), 1056–1061.

[11] Formann, A. K.: *The Newcomb-Benford law in its relation to some common distributions.* PLoS ONE 5 (2010), e10541.

[12] Giles, D. E.: *Benford's law and naturally occurring prices in certain eBay auctions.* Appl. Econ. Lett. 14 (2007), 157–161.

[13] Hardy, G. H., Wright, E. M.: *An introduction to the theory of numbers.* 4. vyd., Oxford Univ. Press, New York, 1960.

[14] Hill, T. P.: *A statistical derivation of the significant-digit law.* Statist. Sci. 10 (1995), 354–363.

[15] Hill, T. P.: *Base-invariance implies Benford's law.* Proc. Amer. Math. Soc. 123 (1995), 887–895.

[16] Judge, G., Schechter, L.: *Detecting problems in survey data using Benford's law.* J. Hum. Resour. 44 (2009), 1–24.

[17] Kallenberg, O.: *Random measures.* Academic Press, New York, 1983.

[18] Kantorek, P.: *Benfordův zákon.* Vesmír 77 (1998), 583. Dostupné z:
`https://vesmir.cz/cz/casopis/archiv-casopisu/1998/cislo-10/benforduv-zakon.html`

[19] Knuth, D. E.: *The art of computer programming*, 2. díl. Addison-Wesley, New York, 1969.

[20] Ley, E.: *On the peculiar distribution of the U.S. stock indexes' digits.* Amer. Statist. 50 (1996), 311–313.

[21] Newcomb, S.: *Note on the frequency of use of the different digits in natural numbers.* Amer. J. Math. 4 (1881), 39–40.

[22] Nigrini, M.: *A taxpayer compliance application of Benford's law.* J. Amer. Taxation Assoc. 18 (1996), 72–91.

[23] Pinkham, R. S.: *On the distribution of first significant digits.* Ann. Math. Statist. 32 (1961), 1223–1230.

[24] Policie České republiky: *Statistiky nehodovosti.* Citováno 13. 1. 2019. Dostupné z:
`https://www.policie.cz/clanek/statistika-nehodovosti-900835.aspx`

[25] Raimi, R. A.: *The peculiar distribution of first digits.* Sci. Amer. 221 (1969), 109–120.

[26] Raimi, R. A.: *The first digit problem.* Amer. Math. Monthly 83 (1976), 521–538.

[27] Rauch, B., Göttsche, M., Brähler, G., Engel, S.: *Fact and fiction in EU-governmental economic data.* Ger. Econ. Rev. 12 (2011), 243–255.

[28] Schatte, P.: *On mantissa distributions in computing and Benford's law.* J. Inform. Process. Cybernet. 24 (1988), 443–455.

[29] Tam Cho, W. K., Gaines, B. J.: *Breaking the (Benford) law: statistical fraud detection in campaign finance.* Amer. Statist. 61 (2007), 218–223.

[30] Varian, H.: *Benford's law.* Amer. Statist. 26 (1972), 65–66.