Benfordov zakon: integracija statističke teorije i računovodstvene prakse

Jovana Mileusnić, Anja Kulagić

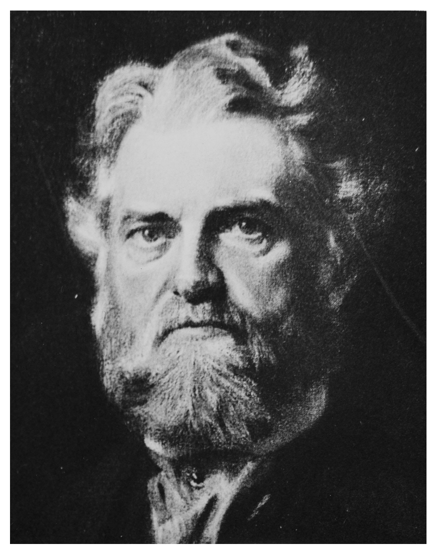
Moć broja jedan

Benfordov zakon, koji opisuje distribuciju prvih cifara u skupu brojeva, postao je popularna metoda u različitim granama istraživanja. Prema datom zakonu, nije jednako verovatno da će se cifre od 1 do 9 naći na prvom mestu u broju, već se pojavljuju u određenim proporcijama koje su definirane logaritamskom raspodelom. U daljem radu definisaćemo fenomen i opisati put njegovog razvoja, da bismo potom istražili primenu zakona vodeće cifre u različitim oblastima, uključujući finansijsku analizu, forenziku, prirodne i društvene nauke. Odgovrićemo na pitanje ***kako nam brojevi mogu otkriti istinu***. Osim toga, analiziramo i neke od kritika i ograničenja ovog zakona.

Razvoj Benfordovog zakona

Fenomen i zakon prve cifre prvi put je zapažen još 1881. godine od strane Sajmona Njukoma[1]. Tada, svoj istraživački rad započeo je objašnjenjem: „Ideja da se deset cifara ne pojavljuje sa jednakom frekvencijom mora biti očigledna svakome ko učestalo koristi logaritamske tablice[2]. Lako se može uočiti da se prve stranice troše mnogo brže od poslednjih“. (Newcomb (1881)) Ipak, sam rad prošao je otpuno nezapaženo. Naime, ideja da postoji značajna disproporcija u pojavama čije veličine počiju brojem 1 u odnosu na pojave čije vrednosti počinju, na primer, počinju brojem 9 delovala je potpuno kontraintuitivno. Takođe, smatra se da je nedostatak jasnog matematičkog dokaza i objašnjenje ujedno uticao da fenomen ne bude prepoznat i opšte prihvaćen.

Slika 1: Simon Newcomb



Članak je bio zaboravljen skoro šest decenija, dok fenomen nije (re)otkriven od strane Frenka Benforda[3], israživača i fizičara u Dženeral elektriku (engl. General Electric). Interesantno je da je Frenk, takođe, bio inspirisan

1. Sajmon Njukom (1835-1905), kanadsko-američki astronom i matematičar
2. U 20. veku logaritamske tablice korišćene su kao alat za množenje dva broja pronalaženjem logaritama željenih brojeva u tablicama, zatim sabiranja dva logaritma, a nakon toga anti-logaritmovanjem zbira u cilju pronalaženja proizvoda dva inicijalna broja
3. Frank Albert Benford (1883-1948), američki elektroinženjer i fizičar

neobičnošću koju je prethodno bio primetio Njukom u vezi korišćenja logaritamskim tablica, iako nije bio upoznat sa njegovim radom. Narednih nekoliko godina posvetio je prikupljanju preko 20 000 podataka, "*koliko su vreme i energija dozvoljavali*", iz različitih izvora kao što su površine reka, sportska statistika, atomske težine elemenata i brojevi koji se pojavljuju u časopisu Riders Dajdžest (engl. Reader's Digest). Tada verovatno niko nije mogao ni da predvidi širinu primene ovakvog fenomena, te da razume i objasni strast i istrajnost Benforda u ovom istraživanju.

Slika 2: Frank Benford

Jednom prilikom je i rekao: „*Naravno da niko nije zainteresovan za stanje logaritamskih tablica kao takvima, međutim, ovo pitanje dobija na značaju ukoliko uzmemo u obzir da su logaritamske tablice bile jedne od ključnih alata u stvaranju naše naučne, indženjerske i opšte literature*". (Benford (1938)) Rezultati višegodišnjeg istraživanja prikazani u Tabeli 1 potvrđuju sumnju da je malo verovatno da inženjeri i fizičari imaju preferencije ka logaritmima brojeva koji počinju sa jedan. Da bi dokazao zakon prve cifre, Benford je morao da izvrši preko 20 000 proračuna potpuno ručno. Shodno tome, interesatno je primetiti da rezultati prikazani u tabeli nisu najprecizniji, te da bi zbir svih proseka bio jedna verovatnoći sigurnog ishoda, preciznije jednak jedinici, tek kada bi se decimalni zapis pomerio dva mesta ulevo. Kao rezultat sprovedenog eksperimenta, Benford je uočio sledeći obrazac: verovatnoća da će se cifra jedan pojaviti na prvom mestu bliska je logaritmu broja 2, dok se čini da je relativna zastupljenost brojeva koji počinju cifrom 2 bliska logaritmu broja $\frac{3}{2}$. Ovakav šablon se nastavlja sve do cifre 9, ukazujući na monotono opadanje od $log(1+\frac{1}{1}) \approx 0.301$ do $log(1+\frac{1}{9}) \approx 0.046$.

Jednostavnije rečeno, Benfordov zakon nalaže da, ukoliko poređamo podatke u rastući niz, oni će otprilike pratiti geometrijski niz („otprilike" zbog činjenice da je moguće naići na dva identična podatka u Benfordovom setu). Ono što se stavlja kao uslov jeste da će data aproksimacija biti približnija na skupu višecifrenih brojeva. Dok za očekivanu frekvenciju cifara na prvom mestu važi da formira monotono opadajući logaritamski niz, verovatnoća da se cifre od 1 do 9 nađu na drugoj poziciji prati približno uniformnu raspodelu. Kod treće i četvrte cifre, odstupanje empirijskih proporcija od očekivane uniformne raspodele su zanemarljivo. (Tabela **??**).

Ono što je važnije, Benford je uspeo da empirijski dokaže važnost i univerzalnost pomenutog fenomena u svom radu pod nazivom „Zakon brojeva sa anomalijom", objavljen 1938. godine u naučnom časopisu objavljenom od strane američkog filozofskog društva. Zašto je baš ovakav naslov izabrao za svoj rad? Tokom svog eksperimenta, Benford je zaključio da je uočeni logaritamski šablon uočljiviji na nizu nasumičnih, takozvanih „brojeva bez logičke veze" kao što su recimo arapski projevi pronađeni u nekoliko uzastopnih članaka na naslovnoj strani lokalnih novina (ne računajući datum). Njukom ih je još nazivao „prirodnih brojevima" zato što brojevi koji koriste naučnici nisu ni čisto slučajni brojevi, ali ni čisto deterministički. Takođe, zanimljivo je da, uprkos činjenici da mnoge pojave uključene u

Tabela 1: Procentualna zastupljenost prirodnih brojeva od 1 do 9 kao prve cifre u posmatranom broju u skupu od 20 229 opservacija

Grupa	Pojava	Broj	Prva cifra									Diff
			1	2	3	4	5	6	7	8	9	
A	Rivers, Area	335	31.0	16.4	10.7	11.3	7.2	8.6	5.5	4.2	5.1	9.8
B	Population	3259	33.9	20.4	14.2	8.1	7.2	6.2	4.1	3.7	2.2	16.6
C	Constants	104	41.3	14.4	4.8	8.6	10.6	5.8	1.0	2.9	10.6	34.9
D	Newspapers	100	30.0	18.0	12.0	10.0	8.0	6.0	6.0	5.0	5.0	2.8
E	Spec. Heat	1389	24.0	18.4	16.2	14.6	10.6	4.1	3.2	4.8	4.1	24.2
F	Pressure	703	29.6	18.3	12.8	9.8	8.3	6.4	5.7	4.4	4.7	3.2
G	H.P. Lost	690	30.0	18.4	11.9	10.8	8.1	7.0	5.1	5.1	3.6	4.8
H	Mol. Wgt.	1800	26.7	25.2	15.4	10.8	6.7	5.1	4.1	2.8	3.2	23.2
I	Drainage	159	27.1	23.9	13.8	12.6	8.2	5.0	5.0	2.5	1.9	21.6
J	Atomic Wgt.	91	47.2	18.7	5.5	4.4	6.6	4.4	3.3	4.4	5.5	35.4
K	n^{-1}, $\sqrt{n}, ...$	5000	25.7	20.3	9.7	6.8	6.6	6.8	7.2	8.0	8.9	22.8
L	Design	560	26.8	14.8	14.3	7.5	8.3	8.4	7.0	7.3	5.6	16.6
M	Readers' Digest	308	33.4	18.5	12.4	7.5	7.1	6.5	5.5	4.9	4.2	8.4
N	Cost Data	741	32.4	18.8	10.1	10.1	9.8	5.5	4.7	5.5	3.1	12.4
O	X-Ray Volts	707	27.9	17.5	14.4	9.0	8.1	7.4	5.1	5.8	4.9	7.4
P	Am. League	1458	32.7	17.6	12.6	9.8	7.4	6.4	4.9	5.6	3.0	6.6
Q	Black Body	1165	31.0	17.3	14.1	8.7	6.6	7.0	5.2	4.7	5.4	7.2
R	Adresses	342	28.9	19.2	12.6	8.8	8.5	6.4	5.6	5.0	5.5	5.4
S	$n^1, n^2, ..., n!$	900	25.3	16.0	12.0	10.0	8.5	8.8	6.8	7.1	5.5	13.8
T	Death Rate	418	27.0	18.6	15.7	9.4	6.7	6.5	7.2	4.8	4.1	11.2
	Prosek	1011	30.6	18.5	12.4	9.4	8.0	6.4	5.1	4.9	4.7	–
	Moguća greška	–	±0.8	±0.4	±0.4	±0.3	±0.2	±0.2	±0.2	±0.2	±0.3	–

Napomena: Tabela 1 je sastavljena uglavnom od četvorocifrenih, petorocifrenih i šestocifrenih brojeva.
Izvor: Benford (1938)

Tabela 2: Očekivana raspodela frekvencija cifara definisana Benfordovim zakonom

Cifra	Prva	Druga	Treća	Četvrta
0		.11968	.10178	.10018
1	.30103	.11389	.10138	.10014
2	.17609	.10882	.10097	.10010
3	.12494	.10433	.10057	.10006
4	.09691	.10031	.10018	.10002
5	.07918	.09668	.09979	.09998
6	.06695	.09337	.09940	.09994
7	.05799	.09035	.09902	.09990
8	.05115	.08757	.09864	.09986
9	.04576	.08500	.09827	.09982

Izvor: Nigrini (1996)

Benfordovo istraživanje, pojedinačno ne prate zakon prve cifre, „ono što je najbliže zakonu jeste unija svih table“ (Pogledati Raimi (1976)). Takođe, nasumično uzorkovanje iz svake od tabela proizvešće raspodelu frekvencija prve cifre na način koji približno prati Benfordov zakon. Benford svoj rad zaključuje sledećim zapažanjem: „*Analogija je potpuna i čovek dolaži u iskušenje da pomisli da skala 1,2, 3, … nije prirodna skala; međutim, ako uključimo Ojlerov broj e kao bazu prirodnog logaritma, prirodna broji na sledeći način:* $e^0, e^x, e^{2x}, e^{3x}, \ldots$ *i gradi i funkcioniše u skladu sa tim*“. (Benford (1938)) ”Ne pamtim puno iz vemena kada je otac pisao o brojevima sa anomalijom osim da je istragu počeo nakon što je primetio da su mu logaritmaske tablice bile prljavije na stranicama sa manjim ciframa. Takođe se sećam da je uživao dok je sa prijateljima pričao o svom pronalasku”, izjavio je Harry Benford, sin Frenka Benforda.

Tako je opisani fenomen postao Benfordov zakon, dok u literaturi možemo naići na još nekoliko alternativnih naziva kao što je zakon prve cifre ili zakon značajne (vodeće) cifre. U znak sećanja na njegovog pionira, povremeno se naziva i „Njukom-Benfordov zakon“. U knjizi „Benfordov zakon: teorija i primena“, unuk Frenka Benforda rekao je: „To što je „Zakon brojeva sa anomalijom“ mog dede postao poznat pod nazivom „Benfordov zakon“ umesto „Njukomov zakon“ je istorijska slučajnost. Ne žalim se, naravno, ali potomci Sajmona Njukoma imaju puna prava na nezadovoljstvo.“ (Miller (2015))

Sada, kada smo se upoznali sa osnovnom idejom, možemo da damo i formalnu definiciju Benfordovog zakona:

Definicia 1. ***(Benfordov zakon vodeće cifre**[4]**)*** *Kažemo da skup podataka zadovoljava Benfordov zakon ako je verovatnoća pojave cifre d kao prve cifre približno jednaka* $log(1+\frac{1}{d})$.

Navedena formula u skladu je sa idejom da je distribucija vodećih cifara pristrasna prema nižim ciframa, te da se manje cifre poput 1, 2 ili 3 pojavljuju mnogo češće nego cifre višeg reda.

Navedenu pravilnost možemo uopštiti, te definisati verovatnoće i na ostale pozicije brojeva od interesa kroz tzv. ***Opšti zakon cifre od interesa***. Zakon je formalno definisan na sledeći način (Hill (1995a)):

$$P(D_1 = d_1, ..., D_k = d_k) = log_{10}[1 + (\sum_{i=1}^{k} d_1 10^{k-i})^{-1}] \tag{1}$$

gde je $k \in N$, $d_1 \in 1, 2, ..., 9$, $d_j \in 0, 1, 2, ..., 9$ i $j = 2, ..., k$.

Intuitivno, možemo primetiti da posmatrane cifre nisu nezavisne, te da važi:

$$P(D_2 = d_2 | D_1 = d_1) = \frac{log(1 + \frac{1}{d_1 d_2})}{log(1 + \frac{1}{d_1})} \tag{2}$$

Dakle, verovatnoću pojave dvojke kao druge cifre bismo dobili na sledeći način (Jednačina 3):

$$\begin{aligned} P(D_2 = 2) = \sum_{d_1=1}^{9} log(1 + \frac{1}{d_1 d_2}) = \\ = log(1 + \frac{1}{12}) + log(1 + \frac{1}{22}) + log(1 + \frac{1}{32}) \\ + log(1 + \frac{1}{42}) + log(1 + \frac{1}{52}) + log(1 + \frac{1}{62}) \\ + log(1 + \frac{1}{72}) + log(1 + \frac{1}{82}) + log(1 + \frac{1}{92}) \\ = 0.10882 \end{aligned} \tag{3}$$

4. Nula ne može biti prva cifra, dok se kod negativnih projeva, negativan predznak zanemaruje.

Možemo primetiti da zakon cifre od interesa kaže da je nula najverovatnija druga cifra (Tabela **??**). Interesantno je napomenuti da, iako Njukomb nije precizno definisao ove formule, mogli bismo pretpostaviti da ih je bio svestan, a imajući u vidu da je već u svom pionirskom radu jasno naznačio verovatnoću javljanja cifre od interesa na prvom i/ili mestu.
Značenje pojma prva cifra može delovati sasvim jasno i intuitivno, no zbog izuzetaka koji postoje kod brojeva manjih od nula, kao i brojeva koji se nalaze u intervalu $0 < x < 1$, definišemo pojam vodeće cifre kroz Jednačinu 4, a koristeći se formom standarne (naučne) notacije[5]:

$$\text{Standardni zapis} = a \times 10^n \ (\text{gde } 1 \leqslant a < 10, \text{i n} \in \mathbb{Z}) \tag{4}$$

Na primer, 5,600,000 možemo zapisati kao 5.6 x 10^6 u formi naučne notacije. Ukoliko bismo želeli da saznamo prve dve cifre datog broja, zapisali bismo ga kao 56 x 10^5. Da zaključimo, Benfordov zakon ne pravi razliku između 30 i 30 000, kako obe veličine počinju brojem 3, dok su ostale cifre 0.
Prepoznavanje koji od skupova podataka će najverovatnije pratiti Benfordov zakon, a koji neće, može biti od presudnog značaja. Takođe, posmatrani set podataka mora da opisuje slične podatke, dok same vrednosti ne smeju imati gornju i/ili donju granicu. U Tabeli 3 možemo da vidimo značajno odstupanje raspodele frekvencija prve cifre od one koja je opisana kroz logaritamski obarazac, a usled postojanja gornje granice. Razlika između očekivane i empirijske raspodele frekvencije ne može se opisati kao značajna u slučaju druge cifre. Ovaj zaključak može svedočiti obuhvatnijoj i manje osetljivoj primeni datog fenomena u slučaju druge u odnosu na prvu cifru. (Pogledati Pericchi and Torres (2011))

Tabela 3: Distribucija prve i druge cifre bez i sa prisustvom gore granice

	0	**1**	**2**	**3**	**4**	**5**	**6**	**7**	**8**	**9**
BL1		0.301	0.176	0.125	0.097	0.079	0.067	0.058	0.0512	0.046
$BL1_{800}$		0.330	0.193	0.137	0.106	0.087	0.073	0.064	0.006	0.005
BL2	0.120	0.114	0.109	0.104	0.100	0.097	0.093	0.090	0.088	0.085
$BL2_{800}$	0.121	0.114	0.109	0.104	0.100	0.097	0.093	0.090	0.087	0.085

Izvor: Pericchi and Torres (2011)

Iako važi preduslov da podaci ne smeju biti previše ograničeni kako bi pratili Benfordov zakon, on takođe zahteva da podaci ne budu ni potpuno slučajni. Možda nam ova pretpostavka može pomoći u davanju odgovora na pitanje kako i zašto nam pozavanje Benfordovog zakona ne može povećati šansu za dobitak na lutriji, budući da svaki mogući ishod ima jednake šanse da se dogodi. Dr Nigrini je primetio: „*U lutriji je samo otrebno da neko jednostavno izvuče kuglice iz tegle. Kuglice zapravo i nisu brojevi. One su označene nizom brojeva, ali se te oznake veoma lako mogu promeniti u imena životinja. Brojevi su uniformno raspoređeni što znači da svaki od njih ima jednaku šansu da bude izvučen, a Benfordov zakon se ne primenjuje na uniformne raspodele*“. (Browne (1998)) Takođe, empirijski je dokazano da je Benfordova raspodela uočljivija kada podaci u sebi sadrže brojeve različitih redova veličina, preciznije, ukoliko je to skup jednocifrenih, dvocifrenih, trocifrenih, četvorocifrenih... brojeva.

5. Često se koristi u oblasti nauke, astronomije i fizike, odnosno naukama koje rade sa veoma malim ili veoma velikim brojevima.

(***The Spread Theory***, pogledati Hill (1995b), Sambridge, Tkalčić, and Jackson (2010), Raimi (1969)) Ipak, moramo imati u vidu da je Benfordov zakon delimično ograničen zato što „preferira" veće brojeve (barem četvorocifrene) kako bi osnovne pretpostavke i kalkulacije bile ispunjene. Dalje, i kod brojeva koji imaju od četiri cifre može važiti Benfordov zakona, ali uz određenu pristrasnot ka ciframa nižeg reda. (Nigrini (2012)) Najzad, da bi se mogao uočiti obrazac koji je opisan Benforovim zakonom, potrebno je da set sadrži najmanje 100 opservacija. (Više o ovoj temi u nastavku: Figure **??**)

Oblasti primene Benfordovog zakona

Nakon što je sproveo eksperiment sa svojim studentima matematika na Tehnološkom institut u Džordžiji, Dr Hill[6] je u jednom od svojih intervjua istakao: *„Istina je da većina ljudi ne zna da na ubedljiv način lažira sopstvene podatke zato što nisu dobro upoznati sa dobijenim rezultatima ovog fenomena". Njegov ekperiment bazirao se na ideji da studenti ili bacaju 200 puta novčić i zabeleže dobijene rezultate, ili da se pretvaraju da su bacali novčić i da lažiraju 200 rezultata. On je uspeo da prepozna lažne rezultate tako što je proverio da li rezultati studenata sadrže nizove od šest uzastopnih glava ili pisama, jer „većina prevaranata izbegava da beleži duge rezultate nizova pisama ili glava, verujući da su takvi rezultati nemogući"*. (Browne (1998))

Pilot rezultati dobijeni na popisu iz 1990. godine od ukupno 3141 država, pokazuju da je statistika stanovništva u skladu sa Benfordovim zakonom, uprkos činjenici da je oko 1 000 od 19 000 brojeva jednocifreno i dvocifreno. (Nigrini (1999)) Podaci pokrivaju veliki raspon brojeva i uglavnom su rezultat egzogenih procesa i spoljnih efekata. Dakle, pretpostavke Benfordovog zakona su ispunjene. Pored toga, otkrili su da je zakon važio i za vrednosti projektovane za 1991. i 1992. godinu. Otuda, ako skup podataka prati Benfordov zakon, to će važiti i za modele izvedene iz prvobitnog niza. Ideja za pomenutu analogiju došla je od Varian (1972), dok Hil opusuje ovaj fenomen kao *„Benford-in-Benford-out"*. (Hill (1995c) and Jamain (2001)) Hil zaključuje: *„Ovo bi mogao biti test za validnost modela koji ima zadatak da predviđa buduće vrednosti posmatrane pojave – ako dobijene vrednosti ne prate barem približno Benfordov zakon, revidirajte model"*. (Matthews (1999))

Još jedan zanimljiv slučaj primene Benfordovog zakona jeste studija čiji je cilj bio da ispita validnost makroekonomskih podataka dostavljenih od strane grčke ambasade Evrostatu u Luksemburgu. Kako je u datom skupu podataka raspodela frekvencija cifara od interesa značajno odstupala od logaritamskog šablona koji smo definisali, dobijeni rezultati sugerisali su na moguće prisustvo kreativnog računovodstva u izveštajima grčke vlade. (Rauch et al. (2011)) Benfordov zakon može biti od koristi i za otkrivanje nedoslednosti u izbornim ishodima, što je takođe atraktivno polje primene. (Za više o ovoj temi: *"Election Forensics: Vote Counts and Benford's Law"* (Mebane Jr (2006)))

Pored navedenih primera, Malkom Sambridž u svom radu *"Benford's Law in the Natural Sciences"*

6. Teodor Preston Hil (1943 -), američki matematičar specijalizovan u oblasti teorije verovatnoće

uvodi skupove podataka sa lognormalnom distribucijom koja opisuje mnoge prirodne pojave (Tabela 4).

Tabela 4: Distribucija prve cifre izražene u procentima za različite skupove podataka u fizici

	Distrbucija prvih cifara								
	1	2	3	4	5	6	7	8	9
P_d	30.1	17.6	12.49	9.69	7.92	6.69	5.8	5.12	4.58
Geomagnetno polje	28.9	17.7	13.3	9.4	8.1	6.9	6.1	5.1	4.5
Geomagnetni pokreti	32.3	19.4	13.9	11.8	5.3	4.3	3.2	5.4	4.3
Seizmičke brzine talasa ispod J3 Pacifika	30.0	17.6	13.3	9.8	7.9	6.4	5.6	4.89	4.47
Zemljina gravitacija	33.0	16.6	11.2	8.5	7.5	6.7	5.94	5.57	5.03
Masa egzoplaneta	33.9	15.4	10.7	9.2	6.23	9.47	5.98	4.48	4.48
Grekvencija rotacije pulsara	33.9	20.7	12.7	7.6	5.3	5.00	4.94	4.67	4.88
Teleskop Fermi za detektovanje gama-zraka	30.3	17.9	13.00	9.9	7.6	6.96	5.23	5.23	2.72
Dubina zemljotresa	31.6	16.9	14.00	8.69	6.98	7.42	5.27	4.58	4.36
S-A seizmograf	28.4	15.7	12.5	9.6	8.97	7.37	6.52	6.04	4.93
Emisije gasova steklene bašte u zemljama	29.9	17.9	11.4	7.6	9.2	8.15	5.97	4.89	4.89
Globalne temp anomalije 1880–2008	27.7	19.4	12.7	12.1	8.9	5.4	6.61	4.32	2.81
Osnovne konstante u fizici	34.00	18.4	9.2	8.28	8.58	7.36	3.37	5.21	5.52
Globalni slučajevi zaraznih bolesti	33.7	16.7	13.2	10.7	7.3	5.4	4.56	5.07	3.34
Geometrijski niz	29.8	17.4	13.00	10.00	7.8	6.6	5.8	5.00	4.6
Fibonačijevi brojevi	30	17.7	12.5	9.6	8.00	6.7	5.7	5.3	4.5
Kombinovano	30.9	17.4	13.2	9.00	7.6	6.4	5.7	4.8	5.00

Izvor: Sambridge, Tkalčić, and Jackson (2010)

U njegovoj knjizi *"Benfords Law Applications for Forensic Accounting, Auditing, and Fraud Detection"* (Nigrini (2012)), Mark Nigrini naglašava: „Najznačajnija dostignuća u razvoju i popularizaciji fenomena desila su se u akademskom krugu. Online baza (Berger, Hill, and Rogers (2009)) sadrži oko 750 obavljenih radova na temu Benfordovog zakona. Pretpostavljam da će do kraja 2015. godine ova brojka doći i do 1 500. U 1975. godini bilo je samo 50 radova na datu temu, dok je 2000-ih bilo oko 150“. Dugo je nedostatak matematički rizogoroznog dokaza mučio naučnike. Veruje se da se do adekvatnog dokaz došlo tek 1995. godine od strane T. P. Hilla, koji se prisetio rezultata Benforda i činjenice da je unija svih tabela najbliža šablonu koji nalaže Benfordov zakon. (Hill (1995a), Hill (1995c)) Značajan broj radova i istraživača dali su svoj doprinos da pomenuti fenomen dobije i svoje matematičko uporište. (Matthews (1999), Pinkham (1961), Berger and Hill (2011), Kossovsky (2006)). Neke od zanimljivijih osobina fenomena jeste ideja da je Benfordov zakon invarijantan na promenu logaritmaske baze (engl. *base invariance*), kao skaliranje podataka nenultom konstantom (engl. *scale invariance*).

Finansija i Računovodstvo

Mnoge studije su potvrdile njegovu primenu u različitim oblastima, uključujući forenziku i finansijsko izveštavanje. Kao što smo videli, Benfordov zakon može otkriti prevaru u finansijskom izveštavanju, budući da manipulisani finansijski izveštaji najčešće ne poštuju prirodnu distribuciju prvih cifara koju predviđa ovaj zakon. U ranim dvadesetim godinama 20. veka, pretpostavljamo da ni sam Benford nije mogao da vidim praktičnu primenu sopstvenog rada. Zaključci i rezultati njegovih numeričkih analiza, za koje je bilo potrebno nekoliko godine, nisu bili primenljivi u tim vremenima. Tek 1988. godine, Benfordov zakon bio je prvi put korišćen u istraživanju Čarsla Karslava. Kroz svoju studiju, Karslav je primetio da mnoge kompanije koje su objavile pozitivan (negativan) finansijski rezultat imaju višu (nižu) frekvenciju nule (devetke) na mestu druge cifre u profitu. Ovo zapažanje bio je dovoljno dobar dokaz za psotojanje sumnje da je došlo do finansijskih pronevera, što je ubrzo i dokazano. (Carslaw (1988))

Kroz ispitivanje potencijalnih obasti u kojima bi zakon vodeće cifre imao značajnu primenu, Mark Nigrini imao je značajan uticaj da dalji razvoj Benfordovog zakona, posebno kroz predstavljanje različitih statističkih testova kojima bi se dokazala signifikantna razlika između empirijski i teorijski očekivane raspodele. „On je kroz svoju doktorsku disertaciju ponudio moguće načine upotrebe Benfordovog zakona za otkrivanje utaje poreza, nakon čega je sam stupio u kontakt sa njim što je rezultovalo saradnjom koja je bila od obostrane koristi.“ Dalje, u jednom od intervjua, Robert Barton se priseća: „Naša kancelarija rešavala je sedam različitih slučajeva finansijske prevare koje smo ubrzo i dokazali, a zatim sto te slučajeve iskoristili kao test kompjuterskog programa koji je ponudio dr Nigrini. Program je ispravo uočio svih sedam slučajeva kao vrlo verovatnih prevara“. Ispitivanje povraćaja poreza predsednika Klintona u radu *"A taxpayer compliance application of Benford's law"* (Nigrini (1996)) bio je jedno od prvih i ispraćenijih primena Benfordovog zakona radi ispitivanja validnosti podataka, a kroz upotrebu softverskog rešenja koji je ponudio Nigrini. Analiza je pokazala da, iako postoje određene cifre čija raspodela frekvenija odstupa od očekivane, odstupanje nije značajno, te da ne ukazuje na postojanje prevare. Bez obzira na ishod, ovaj slučaj imao je ogroman doprinos na širenje i otkrvanje daljih oblasti primena opisanog fenomena. Naime, zaključeno je da Benfordov zakon ima dobre potencijale da postane zvaničan detektor za utaju poreza, kao što je prvobitno i bilo prepoznato u istraživačkom radu „Primena Benfordovog zakona pri detektovanju utaje poreza“. Dalje, 1996. godine Eduardo Lej utvrdio je da su jednodevni prinosi na glavne berzanske indekse kao što su S&P 500 i insutrijski indeks Dau Džouns prilično u skladu sa Benfordovim zakonom. (Pogledati Ley (1996)) Kako bi ilustrovao Benfordov zakon, dr Mark J. Nigrini je ponudio sledeći primer: „Ako pretpostavimo da je prosek cena akcija Dau Džonsa indeksa 1 000 dolara, naša prva cifra bi bila 1. Da bismo došli do prosečne cene akcije na nivou gde je prva cifra 2, vrednost mora da se uveća za 100 posto, tj. da se duplira. Recimo da cena indeksa raste po stopi od oko 20 posto godišnje. To znači da bi bilo potrebno čak pet godina da data veličina dostigne vrednost koje će počinjati sa cifrom 2. Ali ukoliko prepostavimo da inicijalna vrednost počinje brojem 5, potrebno je samo povećanje od 20 posto da bi se sa 5 000 došlo na 6 000, što je ekvivalentno periodu od godinu dana. Kada indeks dostigne vrednost od 9 000, potrebno je povećanje od 11 procenata i samo sedam meseci da dostigne 10 000 maraka, koja počine brojem 1. U tom trenutku počinjete ispočetka sa prvom cifrom koja je sada 1. Još jednom morate udvostručiti broj, sa 10 000 na 20 000, pre nego što dostinete 2 kao prvu cifru. Kao što vidite, broj 1 preovladava na svakom koraku progresije.“ (Browne (1998))

Dakle, Benfordov zakon možemo smatrati pouzdanim instrumentom za iskazivanje sumnji o potencijalnim prevarama, proneverama, utajama poreza, kao i greškama koje prave računovođe ili komjuteri. Pomenuti pristup je široko prihvaćen u forenzičkom računovodstvu. Procedura za analizu cifara zasnovana na Benfordovom zakonu treutno je dostupna u okviru nekoliko dobro poznatih softverskih paketa za revizore kao što je ACL i CaseWare 2002. Poznato je da se tehnologije korišćene za analizu zakona prve cifre uglavnom fokusiraju na račune i transakcije na strani pasive.

Boyle (1994) je pokazao da će Benfordov zakon važiti čak i kada su elementi rezultat slučajnih promenljivih koje su podeljene/pomnožene. Ovo sugeriše da se Benfordov zakon može primeniti na računovodstvene podatke, koji su često rezultat matematičkih procedura. Durtschi, Hillison, and Pacini (2004) su ponudili smernice u kojoj situaciji analiza primenom benfordovog zakona može doneti razumne zaključke:

Tabela 5: Podacima na kojima Benfordov zakon (ne) bi trebao da važi

Kada smemo da koristimo Benfordov zakon?	**Primer**
Setovi podataka nastali kao kombinacija druga dva seta podataka	Potrživanja (cena × količina)
Podaci na nivou transakcija	Isplate, uplate troškovi
Veliki setovi podataka - više je bolje	Transakcije iz cele kalendarske godine
Setovi u kojima važi pozitivna asimetrija (asimetrija u desno)	Većina računovodstvenih podataka
Kada ne smemo da koristimo Benfordov zakon?	**Primer**
Setovi podataka koji su veštački dodeljeni	Čekovni brojevi, brojevi faktura, poštanski brojevi
Brojevi koji su proizvod ljudske odluke	Iznosi postavljeni kao limit za određene aktivnosti
Setovi podataka sa unapred definisanim minimumom i/ili maksimumom	Određene pozicije na strani aktive/pasive koje moraju da dostignu definisane minimume da bi bile zabeležene
Računi koji sadrže mnogo podataka koji su specifični za datu firmu	Računi koji su posebno kreirani da beleže povraćaj novca do 5 hiljada dinara
Računi na kojima nema zabeleženih transakcija	Prevare, mito, namešteni ugovori

Zaključak

Benfordov zakon odnosi se na fenomen koji opisuje prirodnu distribuciju prvih cifara u skupu brojeva. Može se reći da je kontraintuitivan, budući da bi inicijalna pretpostavka bila da distribucija vodećih cifara prati uniformnu raspodelu, što nije slučaj. Ova neintuitivnost je zapravo ono što čini Benfordov zakon korisnim u analizi skupa brojeva, jer otkriva skrivene uzorke i nepravilnosti koji se ne bi otkrili na intuitivan način. Shodno tome, primena ovog zakona u forenzici postaje sve popularnija, jer može biti korisna u otkrivanju prevara i nepravilnosti u finansijskim izveštajima, kao i u drugim oblastima koje zahtevaju analizu velikih skupova brojeva. Ono što čini ovu oblast još privlačnijom jeste ideja da su razni segmenti idalje nedovoljno istraženi, te postoje ogromni potencijali za dalji razvoj ovog fenomena. Istraživače i stručnjaci mogu osetiti posebno nadahnuće u oblasti razvoja statističkih testova koji će nam pomoći da otkrijemo u kojoj meri posmatrani skup podataka odstupa od teorijske raspodele, kao i da jasnije izdvojimo podskup koji će zahtevati dalju istragu. Ipak, moramo imati u vidu ograničenja koja nam stoje kao prepreka. Bilo da su podaci u skladu ili potpuno narušavaju Benfordov šablon, jedini zaključak koji sa sigurnošću možemo doneti jeste da ova tehnika analize ne dokazuje prisustvo prevare i manipulacije, već radije samo nagoveštavaju oblast koja zahteva dodatnu istragu.

Reference

Benford, Frank. 1938. “The law of anomalous numbers.” *Proc. Amer. Philosophical Soc.* 78:551–572.

Berger, A., T. P. Hill, and E. Rogers. 2009. “Benford Online Bibliography.” Accessed March 19, 2023. http://www.benfordonline.net.

Berger, Arno, and Theodore P Hill. 2011. “Benford’s law strikes back: No simple explanation in sight for mathematical gem.” *The Mathematical Intelligencer* 33 (1): 85.

Boyle, Jeff. 1994. “An application of Fourier series to the most significant digit problem.” *The American Mathematical Monthly* 101 (9): 879–886.

Browne, Malcolm W. 1998. “Following Benford’s law, or looking out for no. 1.” *New York Times.*

Carslaw, Charles A. P. N. 1988. “Anomalies in Income Numbers: Evidence of Goal Oriented Behavior.” *The Accounting Review* 63 (2): 321–327. ISSN: 00014826. http://www.jstor.org/stable/248109.

Durtschi, Cindy, William Hillison, and Carl Pacini. 2004. “The effective use of Benford’s law to assist in detecting fraud in accounting data.” *Journal of forensic accounting* 5 (1): 17–34.

Hill, Theodore P. 1995a. “A statistical derivation of the significant-digit law.” *Statist. Sci.* 10 (4): 354–363. ISSN: 0883-4237. http://links.jstor.org/sici?sici=0883-4237(199511)10:4⟨354:ASDOTS⟩2.0.CO;2-P&origin=MSN.

———. 1995b. “Base-invariance implies Benford’s law.” *Proc. Amer. Math. Soc.* 123 (3): 887–895. ISSN: 0002-9939. https://doi.org/10.2307/2160815. http://dx.doi.org/10.2307/2160815.

Hill, Theodore P. 1995c. “The significant-digit phenomenon.” *The American Mathematical Monthly* 102 (4): 322–327.

Jamain, Adrien. 2001. “Benford’s law.” Master’s thesis, Imperial College of London.

Kossovsky, Alex Ely. 2006. “Towards a better understanding of the leading digits phenomena.” *arXiv preprint math/0612627.*

Ley, Eduardo. 1996. "On the peculiar distribution of the US stock indexes' digits." *The American Statistician* 50 (4): 311–313. ISSN: 1532-2882. https://doi.org/10.1080/00031305.1996.10473558.

Matthews, Robert. 1999. "The power of one." *New Scientist* 163 (2194): 26–30. http://91.82.101.46/wp-content/uploads/2011/08/Benfords-Law.doc.

Mebane Jr, Walter R. 2006. "Election forensics: Vote counts and Benford's law." In *Summer Meeting of the Political Methodology Society, UC-Davis, July,* vol. 17.

Miller, Steven J., ed. 2015. *Benford's Law: Theory and Applications.* Princeton, NJ: Princeton University Press. ISBN: 978-0-691-14761-1. http://press.princeton.edu/titles/10527.html.

Newcomb, Simon. 1881. "Note on the Frequency of Use of the Different Digits in Natural Numbers." *Amer. J. Math.* 4 (1-4): 39–40. ISSN: 0002-9327. https://doi.org/10.2307/2369148. http://dx.doi.org/10.2307/2369148.

Nigrini, Mark J. 1996. "A taxpayer compliance application of Benford's law." *The Journal of the American Taxation Association* 18 (1): 72–91.

———. 2012. *Benford's Law: Applications for Forensic Accounting, Auditing, and Fraud Detection.* Wiley Corporate F&A. John Wiley & Sons. ISBN: 9781118152850. http://books.google.com/books?id=Bh5Vr_I1NZoC.

Nigrini, Mark J. 1999. "I've got your number." *Journal of Accountancy* 187 (5): 79–83.

Pericchi, Luis, and David Torres. 2011. "Quick Anomaly Detection by the Newcomb–Benford Law, with Applications to Electoral Processes Data from the USA, Puerto Rico and Venezuela." *Statistical Science* 26, no. 4 (November): 502–516. https://doi.org/10.1214/09-STS296. http://dx.doi.org/10.1214/09-STS296.

Pinkham, Roger S. 1961. "On the distribution of first significant digits." *Ann. Math. Statist.* 32:1223–1230. ISSN: 0003-4851.

Raimi, Ralph A. 1976. "The first digit problem." *The American Mathematical Monthly* 83 (7): 521–538.

Raimi, Ralph A. 1969. "The Peculiar Distribution of First Digits." *Scientific American* 221 (6): 109–120. https://doi.org/10.1038/scientificamerican1269-109. http://www.jstor.org/stable/24964397.

Rauch, Bernhard, Max Göttsche, Stefan Engel, and Gernot Brähler. 2011. "Fact and fiction in EU-governmental economic data." *German Economic Review* 12 (3): 243–255.

Sambridge, M., H. Tkalčić, and A. Jackson. 2010. "Benford's law in the natural sciences." *Geophysical research letters* 37 (22): L22301. https://doi.org/10.1029/2010GL044830.

Varian, H.R. 1972. "Benford's law." *The American Statistician* 26 (3): 65–66.