BENFORDOV ZAKON

BENFORD'S LAW

Doc. dr. sc. Omer Jukić, dipl.ing.maš.
Rudnik mrkog uglja "Kakanj", d.o.o.
Zgošćanska 17 – 72240 Kakanj

Nedžad Muhurdarević, dipl.ing.elek.
Ericsson, d.o.o.
Fra Anđela Zvizdovića 1 – 71000 Sarajevo

Ključne riječi: Newcomb, Benford, Hill, rezultanta zakona raspodjele, poslovne knjige.

REZIME

Američki astronom i matematičar dr Simon Newcomb je 1881. godine zapazio da sirovi brojčani podaci imaju tendenciju da su brojniji sa niskim početnim brojkama, nego sa visokim, neovisno od ukupne vrijednosti brojeva ili sistema mjera u kojima su veličine izražene. Američki fizičar dr Frank Benford 1938. godine ponovno otkriva ovu zakonitost i provjerava je na ogromnom skupu podataka iz svih oblasti života. Dobija vrlo dobro slaganje podataka po zakonitosti koju je utvrdio i koja će kasnije po njemu dobiti ime – Benfordov zakon. Tek 1996. godine, američki matematičar dr Theodore P. Hill je objasnio i matematski dokazao da je Benfordov zakon rezultanta svih zakona raspodjele u matematskoj statistici.

Trenutno je najubjedljivija primjena Benfordovog zakona pri reviziji poslovnih knjiga radi otkrivanja eventualnog lažiranja podataka. Na primjer, spisak izdataka u poslovnim knjigama se mora slagati s Benfordovim zakonom tim bolje što je broj stavki veći, a ako to nije slučaj postoji vrlo jaka indicija da su podaci lažni i pristupa se detaljnoj provjeri drugim metodama. Čim se počelo s eksperimentalnom primjenom Benfordovog zakona u SAD, otkrivene su prijevare u najvećem broju slučajeva gdje je to indicirao Benfordov zakon.

Key words: Newcomb, Benford, Hill, distributional law resultant, business books.

ABSTRACT

In 1881, an American astronomer and mathematician dr. Simon Newcomb, observed that raw numerical data tend to be more numerous with low starting digits than with high, independently from the total value of numbers or systems of measurement that the values are expressed in. In 1938, dr. Frank Benford, observes this law once again and tests it on a huge set of data collected from all the fields of life. He observes that the data agree with the law he established and that will be named after him later on – Benford's law. Only in 1996, an American mathematician, dr. Theodore P. Hill explained and mathematically proved that Benford's law is the resultant of all the distributional laws in mathematical statistics.

At this moment, the most convincing application of Benford's law is in the inspection of business books in order to discover possible misrepresenting of data. For example, the list of expenditures in business books must agree with Benford's law, the more if the number of items is greater, and if it is not the case, there is strong evidence that the data are false and the other methods can be used to for detailed test. As soon as the experimental application of Benford's law started in the USA, frauds were discovered in most of the cases where indicated by Benford's law.

1. UVOD

1881. godine, američki astronom i matematičar dr Simon Newcomb je zapazio da su logaritamske tablice kojima su se služili u njegovoj opservatoriji mnogo pohabanije na stranicama koje odgovaraju brojevima s početnim ciframa 1,2 itd., da bi ta istrošenost opadala kod stranica s brojevoma koji počinju s visokim ciframa kao 7, 8, 9. On je pravilno zaključio da, iz nekog nepoznatog razloga, sirovi brojčani podaci imaju tendenciju da su brojniji sa niskim početnim brojkama, nego sa visokim, neovisno od ukupne vrijednosti brojeva ili sistema mjera u kojima su veličine izražene. On je o ovome objavio i rad u *The American Journal of Mathematics* koji je ostao nezapažen.
1938. godine (57 godina kasnije, dakle), američki fizičar dr Frank Benford, iz kompanije *General Electric – Bell Laboratories*, ponovno otkriva ovu zakonitost i provjerava je na ogromnom skupu podataka iz svih oblasti života, uključujuću kućne brojeve adresa koje su objavljivane u novinskim člancima, visine planina i dužine rijeka iz geografskih atlasa, astronomske podatke, atomske težine hemijskih elemenata, statističke podatke američke baseball lige itd. Dobija vrlo dobro slaganje podataka po zakonitosti koju je utvrdio i koja će kasnije po njemu dobiti ime – Benfordov zakon, po kojem nalazi oko 30% brojčanih podataka koji počinju jedinicom i onda opadajuću učestalost za veće početne cifre, da bi samo oko 4% brojčanih podataka počinjalo cifrom 9. Od tada je ovo definitivno prihvaćeno kao *empirijski zakon*, ali je ostala misterija kakva je matematska podloga za ovu zakonitost, tj. zašto je to tako.
Tek 1996. godine, američki matematičar dr Theodore P. Hill sa *Georgia Institute of Technology – Atlanta* je objasnio i matematski dokazao da je Benfordov zakon rezultanta svih zakona raspodjele u matematskoj statistici (Gauss-ov, Poisson-ov, binomni i dr.) od kojih je svaki pojedinačno dominantan samo za određene vrste podataka (događaja, veličina).
Benfordov zakon ne važi za skupove podataka sa vještački postavljenim granicama za brojeve kao što su, na primjer, brojevi u telefonskom imeniku. Ne važi, također, za brojeve koji se izvlače za lutrije, lotto, itd., dakle, za čiste slučajne brojeve koji u stvari nisu podaci koji bi odražavali bilo kakvu stvarnu veličinu, nego predstavljaju samo "etikete označene brojevima". Proporcionalni udio prvih cifara takvih brojeva je po definiciji jednak. Benfordov zakon se, dakle, odnosi na brojeve koji zauzimaju "srednje područje" između kruto ograničenih brojeva i brojeva koji nemaju apsolutno nikakvu granicu. To su jedina ograničenja važenja i/ili primjene ovog zakona.
Benfordov zakon je fascinantan – fenomenološki zakon, ne samo zbog činjenice da se protivi našem intuitivnom osjećaju da bi se svi brojčani podaci s početnim ciframa od 1 do 9 trebali pojavljivati otprilike jednako često, neovisno od toga kojom cifrom počinju, nego što se pronalaze jako vrijedne primjene ovog zakona u raznim oblastima rada i života. Trenutno je najubjedljivija primjena Benfordovog zakona pri reviziji poslovnih knjiga radi otkrivanja eventualnog lažiranja podataka. Na primjer, spisak izdataka u poslovnim knjigama se mora slagati s Benfordovim zakonom tim bolje što je broj stavki veći, a ako to nije slučaj postoji vrlo jaka indicija da su podaci lažni i pristupa se detaljnoj provjeri drugim metodama. Čim se počelo s eksperimentalnom primjenom Benfordovog zakona u SAD, otkrivene su prijevare u najvećem broju slučajeva gdje je to indicirao Benfordov zakon.

2. OBJAŠNJENJE BENFORDOVOG ZAKONA

Benfordov zakon je, dakle, fenomenološki zakon. Nazivaju ga, također, i prvim zakonom o brojkama, prvim brojčanim fenomenom ili vodećim brojčanim fenomenom. Benfordov zakon kaže da se u brojčanim listama, statističkim tabelama itd., brojka 1, kao početna cifra neke brojne vrijednosti, pojavljuje s vjerovatnoćom od ~ 30%, što je mnogo više od očekivanih 11,1% (tj. jedna brojka od mogućih 9). Benfordov zakon se odnosi na podatke koji *nisu* bezdimenzionalni, pa numeričke vrijednosti podataka ovise o upotrijebljenim jedinicama. Ako postoji univerzalna vjerovatnoća raspodjele *P(x)* za takve brojeve, onda ona mora biti invarijanta u promjeni tretiranog brojnog sistema. Ako su, prema *Teoremi proizvoda* iz *Teorije vjerovatnosti*, događaji "*x*" i "*k*" među sobom zavisni, onda je:

$$P(kx) = f(k) \cdot P(x) \qquad \text{... (1)}$$

P(kx) - vjerovatnost da se istovremeno dese događaji "*k*" i "*x*"[13],[17]
f(k) - *uslovna vjerovatnost* događaja "*k*"(pod pretpostavkom da se desio događaj "*x*") [13],[17]
P(x) - vjerovatnost događaja "*x*" [13],[17]

Funkciju vjerovatnosti *P(x)* zadovoljava relacija

$$P(x \in B) = \int_B P(x)\,dx$$

a pod realnim uslovima ($x \in \Re$)

$$P(x \in \Re) = \int_{\Re} P(x)\,dx = P(-\infty < x < \infty) = \int_{-\infty}^{\infty} P(x)\,dx \equiv 1 \qquad \text{... (2)}$$

za posebne slučajeve je

$$P(a \le x \le b) = \int_a^b P(x)\,dx$$

$$P(a \le x \le a + da) = \int_a^{a+da} P(x)\,dx \approx P(a)\,da$$

$$P(x = a) = \int_a^a P(x)\,dx = 0$$

dakle, ako je $\int P(x)\,dx = 1$ ... onda je po definiciji

$\int P(kx)\,dx = \dfrac{1}{k}$ ... što implicira da je funkcija *uslovne vjerovatnosti* $f(k) = \dfrac{1}{k}$ ... (3)

diferenciranjem po "*k*" slijedi

$$d\int P(kx)\,dx = d\left(\frac{1}{k}\right)$$

$$P(kx)\,dx = -\frac{1}{k^2}\,dk \qquad ; \qquad P(kx) = \frac{1}{k}P(x)$$

$$\frac{1}{k}P(x)\,dx = -\frac{1}{k^2}\,dk$$

$$P(x) = -\frac{1}{k}\frac{dk}{dx} \qquad \text{... (4)}$$

Da bismo dobili rješenje za raspodjelu funkcije vjerovatnosti *P(x)*, moramo potražiti rješenje diferencijala *dk* iz relacije

$$d(kx) = \frac{\partial(kx)}{\partial x}dx + \frac{\partial(kx)}{\partial k}dk = k \cdot dx + x \cdot dk$$

$$dk = \frac{d(kx)}{x} - k \cdot \frac{dx}{x} \cong -k \cdot \frac{dx}{x} \quad \Rightarrow \quad \frac{dk}{dx} = -\frac{k}{x} \qquad \text{... pa je}$$

$$P(x) = -\frac{1}{k} \cdot \left(-\frac{k}{x}\right) = \frac{1}{x} \qquad \text{... (5)}$$

što znači da funkcija vjerovatnosti *P(x)* ima logaritamsku raspodjelu. Iako ovo nije prava vjerovatnost raspodjele (jer divergira zbog različitih vrijednosti baze logaritma), zakoni fizike i konvencije dozvoljavaju izvjesna odstupanja. Na primjer, ako su ulične adrese jednoliko

raspodijeljene u nizu od jedan do neke maksimalne vrijednosti (vještački prekinut niz brojeva), onda će se uklapati u zakonitost raspodjele blisku Benfordovom zakonu.
Ako veliki broj potencija od 10 leži unutar dozvoljenih odstupanja, onda je vjerovatnost da je prva (decimalna) brojka D data logaritamskom raspodjelom. U praksi se u problemima matematičke statistike umjesto $P(x)$ uvodi *relativna frekvencija* P_D pojavljivanja slučajnog događaja "x":

$$P_D = \frac{M}{N} = \frac{\int_D^{D+1} P(x)\,dx}{\int_1^{10} P(x)\,dx} \qquad \text{...(6)}$$

P_D - *statistička vjerovatnost* (*relativna frekvencija*) pojavljivanja slučajnog događaja "x" [17]
M - broj osmotrenih (koji su se desili) događaja "x" [17]
N - broj svih pokušaja u kojima se događaj "x" mogao desiti [17]
$P(x)$ - teoretska ili matematička vjerovatnost pojavljivanja slučajnog događaja "x" [17]

$$M = \int_D^{D+1} P(x)\,dx = \int_D^{D+1} \frac{1}{x}\,dx = \ln x \Big|_D^{D+1} = \ln(D+1) - \ln D = \ln\frac{D+1}{D} = \ln\left(1+\frac{1}{D}\right) \text{... (7)}$$

$$N = \int_1^{10} P(x)\,dx = \int_1^{10} \frac{1}{x}\,dx = \ln x\Big|_1^{10} = \ln 10 - \ln 1 = \ln 10 \qquad \text{... (8)}$$

$$P_D = \frac{\ln\left(1+\frac{1}{D}\right)}{\ln 10} = \frac{\log_{10}\left(1+\frac{1}{D}\right)}{\log_{10} 10} = \log_{10}\left(1+\frac{1}{D}\right) \qquad \text{... (9)}$$

pa je konačan izraz (procentualno izraženog) Benfordovog zakona:

$$P_D = 100\cdot\log_{10}\left(1+\frac{1}{D}\right)\ [\%] \qquad \text{... za } D = 1,2,\ldots,9 \qquad \text{... (10)}$$

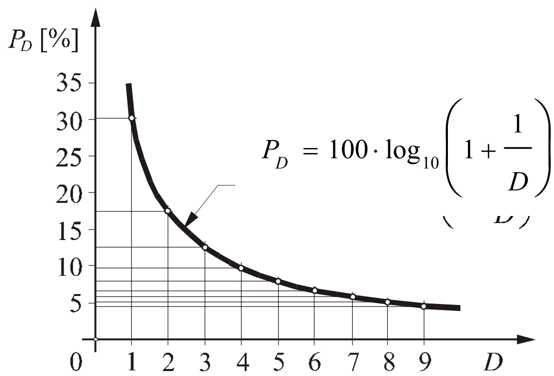
P_D [%] - vjerovatnost da brojčane vrijednosti podataka počinju cifrom D [13]

što tabelarno, dijagramski i grafički predstavljeno ima slijedeći izgled:

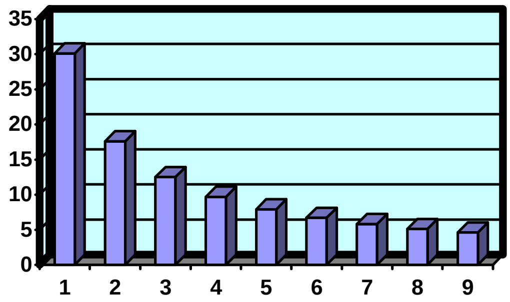
TABELA 1. – BENFORDOV ZAKON (FREKVENCIJA PRVIH CIFARA)

D	**P_D**	
	$\log_{10}\left(1+\frac{1}{D}\right)$	[%]
1	0,30103	30,1
2	0,176091	17,6
3	0,124939	12,5
4	0,09691	9,7
5	0,0791812	7,9
6	0,0669468	6,7
7	0,0579919	5,8
8	0,0511525	5,1
9	0,0457575	4,6

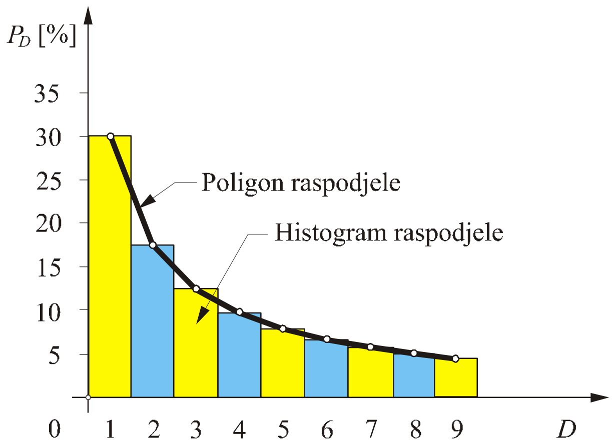
SLIKA 1. – TABELARNI PRIKAZ BENFORDOVOG ZAKONA



SLIKA 2. – DIJAGRAMSKI PRIKAZ BENFORDOVOG ZAKONA



Slika 3. – Grafički prikaz BENFORDOVOG ZAKONA



Slika 4. – Histogram i poligon raspodjele prvih BROJKI PO BENFORDOVOM ZAKONU

Međutim, Benfordov zakon se ne odnosi samo na podatke koji ne variraju u zavisnosti od brojnog sistema, nego i na brojeve iz mnoštva različitih izvora. Kako se broj promjenljivih povećava, gustina dobijene funkcije se približava funkciji logaritamske raspodjele. Za objašnjenje ove činjenice su potrebna obimnija istraživanja koja je obavio dr Theodore P. Hill i pokazao da je "distribucija distribûcîjâ" data u slučajnim uzorcima uzeta iz mnoštva različitih distribucija, de facto, Benfordov zakon.

Upečatljiv primjer potvrde Benfordovog zakona daje 54 miliona stvarnih konstanti u Plouffe-ovoj bazi podataka *"Inverse Symbolic Calculator"*, od kojih 30% počinje brojkom 1.

Uzevši podatke iz nekoliko odvojenih izvora (20.229 podataka), dr Frank Benford je, u svom originalnom radu iz 1938. godine, sastavio tabelu (tabela br.2) koja pokazuje raspodjelu prvih brojki u različitim sistemima brojeva.

TABELA 2. – BENFORDOV ZAKON (RASPODJELA PRVIH BROJKI U RAZLIČITIM SISTEMIMA BROJEVA)

Kol.	**Naziv**	**Prva brojka [%]**									**Podaci**
		1	**2**	**3**	**4**	**5**	**6**	**7**	**8**	**9**	
A	Dužine rijeka	31,0	16,4	10,7	11,3	7,2	8,6	5,5	4,2	5,1	335
B	Naseljenost	33,9	20,4	14,2	8,1	7,2	6,2	4,1	3,7	2,2	3.259
C	Konstante	41,3	14,4	4,8	8,6	10,6	5,8	1,0	2,9	10,6	104
D	Novine	30,0	18,0	12,0	10,0	8,0	6,0	6,0	5,0	5,0	100
E	Specifična toplina	24,0	18,4	16,2	14,6	10,6	4,1	3,2	4,8	4,1	1.389
F	Pritisak	29,6	18,3	12,8	9,8	8,3	6,4	5,7	4,4	4,7	703
H	Molekul. težina	26,7	25,2	15,4	10,8	6,7	5,1	4,1	2,8	3,2	1.800
J	Atomska težina	47,2	18,7	5,5	4,4	6,6	4,4	3,3	4,4	5.5	91
K	n^{-1} ; $\sqrt{n}$	25,7	20,3	9,7	6,8	6,6	6,8	7,2	8,0	8,9	5.000
N	Troškovi	32,4	18,8	10,1	10,1	9,8	5,5	4,7	5,5	3,1	741
O	X-zraci, volti	27,9	17,5	14,4	9,0	8,1	7,4	5,1	5,8	4,8	707
P	Am. baseball liga	32,7	17,6	12,6	9,8	7,4	6,4	4,9	5,6	3,0	1.458
Q	Crno tijelo	31,0	17,3	14,1	8,7	6,6	7,0	5,2	4,7	5,4	1.165
R	Adrese	28,9	19,2	12,6	8,8	8,5	6,4	5,6	5,0	5,0	342
S	$n^1, n^2, \ldots n!$	25,3	16,0	12,0	10,0	8,5	8,8	6,8	7,1	5,5	900
T	Mortalitet	27,0	18,6	15,7	9,4	6,7	6,5	7,2	4,8	4,1	418
	Prosjek:	**30,6**	**18,5**	**12,4**	**9,4**	**8,0**	**6,4**	**5,1**	**4,9**	**4,7**	**1.011**
	Vjerovatna greška:	**± 0,8**	**± 0,4**	**± 0,4**	**± 0,3**	**± 0,2**	**± 0,2**	**± 0,2**	**± 0,3**	**-**	

3. STATISTIČKA DERIVACIJA BENFORDOVOG ZAKONA

Dr Theodore P. Hill je, u svom radu iz marta 1996. godine, kojim je matematički dokazao Benfordov zakon, zapisao [4]:
Benfordov zakon (zakon "značajnih brojki") je takva empirijska opservacija da se u mnogim tablicama numeričkih podataka javljaju "vodeće" cifre koje nisu distribuirane kako se očekuje i umjesto očekivane, djelimično prate logaritamsku distribuciju. Prva poznata pisana referenca u ovom kontekstu je rad američkog astronoma i matematičara dr Simona Newcomba, objavljen u *The American Journal of Mathematics* 1881. godine, u kojem je, između ostalog, zapisano: *"Zakon vjerovatnosti slučajnih brojeva odgovara svim mantisama njihovih logaritama koje su podjednake".*
Sjetimo se da je mantisa dekadskog (Briggsovog) logaritma pozitivnog realnog broja "x" jedinstven broj "r" u $[{}^{1}/_{10},1)$ s relacijom $x = r \cdot 10^n$ za neki cijeli broj "n". Na primjer, mantise od 314 i 0,0314 su (za oba) .,314. Ovaj zakon podrazumijeva da se vodeća cifra "1" pojavljuje s vjerovatnosti $P = \log_{10}2 \approx 0{,}301$, vodeća cifra "2" s vjerovatnosti $P = \log_{10}({}^{3}/_{2}) \approx 0{,}176$ i tako monotono opadajući do vjerovatnosti $P \approx 0{,}046$ pojavljivanja vodeće cifre "9", odnosno:

$$P(\text{prva značajna cifra} = d) = \log_{10}\left(1 + d^{-1}\right) \qquad ; \qquad d = 1,2, \dots ,9 \qquad \dots (11)$$

Tačna zakonitost pojavljivanja prve dvije vodeće cifre (koju je, također, utvrdio dr Simon Newcomb) je:

$$P(\text{druga značajna cifra} = d) = \sum_{k=1}^{9} \log_{10}\left[1 + (10k + d)^{-1}\right] \quad ; \quad d = 1,2, \dots ,9 \qquad \dots (12)$$

Generalni oblik zakona:

$$P(\text{ mantisa} \leq t/10 \text{ }) = \log_{10} t \qquad ; \qquad t \in [1,10) \qquad \dots (13)$$

čak označava zajedničku distribuciju značajnih cifara.

Ako označimo sa D_1, D_2, ... (za bazu logaritma – 10) funkcije značajnih cifara [npr. D_1(.,0314) = 3, D_2(.,0314) = 1, D_3(.,0314) = 4], generalna zakonitost (13) dobija slijedeći oblik:
Za sve pozitivne cijele brojeve "k", svako $d_1 \in \{1,2,\dots,9\}$ i svako $d_j \in \{0,1,\dots,9\}$; $j = 2, \dots ,k$

$$P(D_1 = d_1, \dots, D_k = d_k) = \log_{10}\left[\ 1 + \left(\ \sum_{i=1}^{k} d_i \cdot 10^{k-i}\right)^{-1}\right] \qquad \dots (14)$$

Pojedinačno, $P(D_1 = 3, D_2 = 1, D_3 = 4) = \log_{10}[1 + (314)^{-1}] \approx 0{,}0014$

Možda iznenađujuće djeluje zaključak generalne zakonitosti da su *značajne cifre zavisne*, a ne nezavisne kako se moglo očekivati. Iz (12) slijedi da je (bezuslovno) vjerovatnost pojave cifre "2" na drugoj poziciji nekog broja $P \approx 0{,}109$, ali iz (14) slijedi da je (uslovno) vjerovatnost pojave cifre "2" na drugoj poziciji nekog broja proizašlo iz vjerovatnosti pojave cifre "1" na istoj poziciji $P \approx 0{,}114$. Ova međusobna zavisnost značajnih cifara se rapidno smanjuje kao i razmak njihovog povećavanja, iz čega slijedi generalna zakonitost (14) da se distribucija n-tog značajnog broja približava jedinstvenoj distribuciji od {0,1, ... ,9} eksponencijalno brzo kako $n \to \infty$. Analogno generalnom obliku zakona (13), za druge baze logaritma $b > 1$, vjerovatnost pojave značajnih cifri je:

$$P[\text{mantisa (baza } b)] \leq t / b = \log_b t \qquad \text{za svako } t \in [1, b) \qquad \ldots (15)$$

U kontekstu izloženog su u tabeli br.3 date frekvencije pojavljivanja značajnih cifara u nekom broju na pozicijama od prve do četvrte. Uočljivo je da je frekvencija pojavljivanja značajnih cifara na prvoj poziciji nekog broja osjetno različita, na drugoj značajno umanjena ova razlika, da bi na trećoj, a posebno četvrtoj poziciji bila gotovo potpuno izjednačena.

TABELA 3. – BENFORDOV ZAKON (OČEKIVANA FREKVENCIJA CIFARA)

Cifra	Pozicija cifre u broju							
	Prva		Druga		Treća		Četvrta	
	Mantisa	[%]	Mantisa	[%]	Mantisa	[%]	Mantisa	[%]
0	-	-	.,11968	12,0	.,10178	10,2	.,10018	10,0
1	.,30103	30,1	.,11389	11,4	.,10138	10,1	.,10014	10,0
2	.,17609	17,6	.,10882	10,9	.,10097	10,1	.,10010	10,0
3	.,12494	12,5	.,10433	10,4	.,10057	10,0	.,10006	10,0
4	.,9691	9,7	.,10031	10,0	.,10018	10,0	.,10002	10,0
5	.,7918	7,9	.,09668	9,7	.,09979	10,0	.,09998	10,0
6	.,6695	6,7	.,09337	9,3	.,09940	9,9	.,09994	10,0
7	.,5799	5,8	.,09035	9,0	.,09902	9,9	.,09990	10,0
8	.,5115	5,1	.,08757	8,8	.,09864	9,9	.,09986	10,0
9	.,4576	4,6	.,08500	8,5	.,09827	9,8	.,09982	10,0

4. RAZUMIJEVANJE BENFORDOVOG ZAKONA

Dugo je vremena trebalo da zakon "anomalnih brojeva", ili Benfordov zakon, od obične matematičke zanimljivosti postane standardni alat istraživača. Dr Simon Newcomb je 1881. godine napisao prvi rad na temu frekvencije prvih cifara neke brojne vrijednosti. On je otkrio (a isto je, neovisno od njega, otkrio i dr Frank Benford 1938. godine) nešto što snažno protivrječi našoj intuiciji. Velika je vjerovatnost da će brojevi koji označavaju slučajne fizikalne veličine, kao što su dužine rijeka, broj stanovništva zemalja ili atomske težine hemijskih elemenata, imati prvu cifru "1" , "2" ili "3". Vjerovatnost da se kao prva cifra neke brojčane vrijednosti pojavi "1" je otprilike 30,1%, a za "2" je mogućnost pojavljivanja 17,6%. Vjerovatnost pojavljivanja svake naredne cifre je sve manja, tako da su izgledi da se cifra "9" pojavi kao prva cifra samo 4,6%.

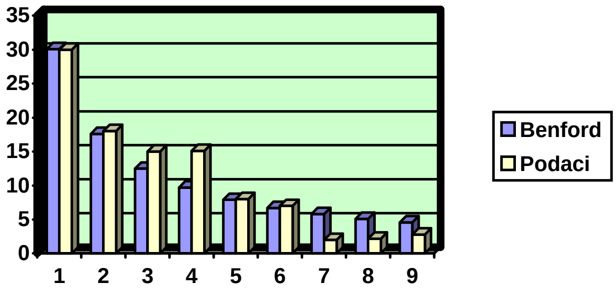
Ovakav ishod je toliko čudan da se većina ljudi koji se s njim susretnu u početku pitaju u čemu je trik. Jasno, prva cifra ne može biti "0" – pa bismo očekivali da svaka od preostalih cifara od "1" do "9" ima podjednake šanse, oko 11,1%. Međutim, matematičari, fizičari i računovođe stalno iznova pokazuju da to nije slučaj. Bilo da raspravljamo o količini odvijača ili tabli šperploče u skladištu, o netto vrijednosti dolara u zajedničkim fondovima, ili u postsezonskoj statistici američke baseball lige, proporcije početnih cifara su jako bliske Benfordovoj standardnoj vjerovatnosti brojeva. One imaju logaritamsku distribuciju.

Evo jednog primjera. Uzmimo jedan mali grad negdje u Bosni. Počnimo ga posmatrati od trenutka kada on ima 1.000 stanovnika. U tom kontekstu, uz godišnji prirast od 10%, trebat će 7,3 godina da mu se stanovništvo udvostruči i sa 1.000 naraste na 1.999 stanovnika. Svo to vrijeme, grad će imati "1" kao prvu cifru broja svoga stanovništva. Sada pretpostavimo da se broj stanovnika grada još jednom udvostruči, sa 2.000 na 3.999. Za ovo će, uz isti godišnji prirast stanovništva, ponovno trebati 7,3 godine – ali će ove godine biti podijeljene između cifara "2" i "3". Period rasta broja stanovništva od 4.000 do 7.999 će trajati, također, 7,3

godine i obuhvataće cifre "4" , "5" , "6" i "7" kao početne cifre broja stanovništva. Ako pretpostavimo da je stopa prirasta stanovništva konstantna, posmatranom gradu treba mnogo manje vremena za porast stanovništva sa 3.000 na 3.999, nego sa 1.000 na 1.999. I konačno, kada ukupan broj stanovništva dosegne 10.000, posmatrani krug porasta stanovništva ponovno započinje.
Tokom više decenija, prosječna proporcija vremena provedenog na datoj početnoj cifri teži ka Benfordovoj vrijednosti. Dakle, ako u bilo kom datom momentu uzmemo uzorak od 1.000 naseljenih mjesta, koji imaju različite stope prirasta stanovništva, doći ćemo do zaključka da će, bez obzira na njihove stvarne veličine – od 100 do 9,999.999 – oko 30% našeg uzorka imati broj stanovnika s prvom cifrom "1".
U godinama koje su uslijedile nakon Benfordovog rada iz 1938. godine, fizičari i matematičari su postepeno produbljavali naše razumijevanje ovih "anomalnih brojeva". Naučili smo, na primjer, da čak i kada grupa brojeva počinje s podjednakom distribucijom, interakcije s okolinom koja ih okružuje će vremenom brzo preurediti vrijednosti prema logaritamskom uređenju. Ako uzmemo 99 brojeva s vrijednostima od 1 do 99, tako da ima 11 brojeva koji počinju sa "1", 11 brojeva koji počinju sa "2" itd., i svaki od njih pomnožimo s drugim slučajnim brojem između 0 i 1 – npr. 0,45 ili 0,78 , rezultati množenja će odmah krenuti s početnom cifrom "1" i doći će do manjka rezultata s početnom cifrom "9".
Dakle, kakav je značaj ovoga? Dugo vremena su istraživači smatrali da sve ovo nema značaja i da Benfordov zakon ne može imati korisnu primjenu. Amerikanac Goudsmit, čiji je rad iz 1944. godine bio prvi koji je slijedio Benfordov, je čak 1978. godine napisao da je jedina praktična primjena Benfordovog zakona "klađenje sa lakovjernim kolegama na digitalne frekvencije". Međutim, u proteklih nekoliko godina je nastao ogroman val interesa za Benfordovim zakonom i otkrivale mogućnosti njegove primjene na mnogim poljima ljudske djelatnosti. Naročito kontrolori računa imaju koristi od Benfordovog zakona, jer im znanje o učestalosti pojavljivanja prvih cifara neke brojne vrijednosti omogućava da otkriju, inače neuhvatljive, primjere grešaka, nezakonitog trošenja i obmana u velikim datotekama bankovnih podataka.
Ovdje ključnu ulogu ima psihologija. Kada poreski obveznici ili pronevjeritelji izmišljaju brojeve, oni slijede onu istu pogrešnu matematičku intuiciju koja Benfordov zakon čini tako iznenađujućim. Kao rezultat takve intuitivnosti, u falsificiranim podacima gotovo uvijek ima previše slučajeva s početnim ciframa 7,8 ili 9 i premalo slučajeva s početnim ciframa 1,2 ili 3. Falsifikatori, također, često ponavljaju određene, jako specifične kombinacije – na primjer, dvije početne cifre "89" u iznosima vrijednosti krivotvorenih čekova se pojavljuju na 2 ili 3 od 20 krivotvorenih čekova, mada je statistička vjerovatnost pojave dvije početne cifre "89" iznosa vrijednosti čeka manja od 1%.
Benfordov zakon se, također, može koristiti za provjeru efikasnosti u radu ili organizaciji posla. Ako neka firma mnogo koristi usluge kurira (na primjer komunalna preduzeća, preduzeća za distribuciju grijanja, vode i sl.), a ne izvrši fakturisanje svojih usluga u prikladnom mjesečnom obračunskom obrascu, lako može završiti na obradi hiljada identičnih faktura od 8,73 KM (ili nekoj drugoj fiksnoj vrijednosti). Čak može doći do pisanja hiljada identičnih čekova za plaćanje ovih faktura. U nekoliko novijih studija izrađenim u velikim američkim korporacijama, skoro polovina ukupnog broja obrađenih faktura je spadala u ovu kategoriju, s vrlo visokim troškovima manuelnog rada na obradi faktura, kao rezultatom neorganiziranosti i neefikasnosti.
Benfordov zakon je moguće koristiti za ocjenu kvaliteta izrađenih investicionih programa, projekata, elaborata, obavljenih laboratorijskih ispitivanja, testiranja novih proizvoda, itd., itd., ...

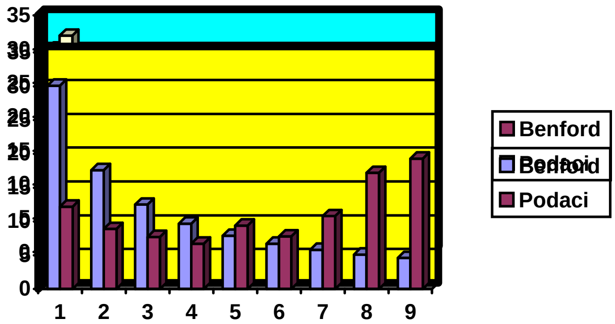
5. PRIMJERI PRIMJENE BENFORDOVOG ZAKONA U PRAKSI

5.1. Podaci o broju stanovništva za 109 općina u BiH po popisu iz 1991. godine



Komentar: S obzirom na mali broj podataka, raspodjela prvih cifara iz raspoloživih podataka je zadovoljavajuća, sa, uslovno, prihvatljivim odstupanjima u odnosu na raspodjelu po Benfordovom zakonu, ali dobijena odstupanja, u svakom slučaju, ostavljaju sumnju na korektnost obavljenog popisa stanovništva.

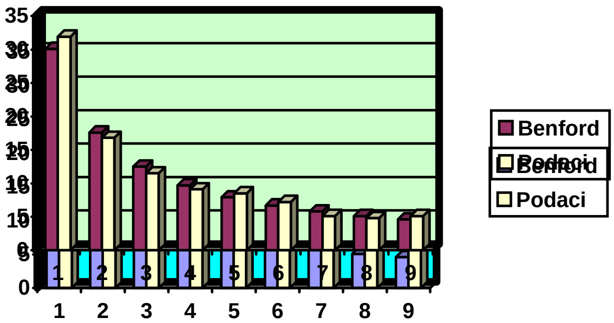
5.2. Podaci o broju stanovništva 191 okruga u SAD (764 podatka)



Komentar: Bez obzira na mali broj podataka, raspodjela prvih cifara iz raspoloživih podataka nije zadovoljavajuća, s vrlo velikim odstupanjima u odnosu na raspodjelu po Benfordovom zakonu, te postoji vrlo jaka indicija da su podaci lažirani. To znači da je neophodno izvršiti detaljnu provjeru zakonitosti poslovanja preduzeća "A" drugim metodama.

5.3. Podaci o finansijskim rezultatima poslovanja preduzeća "A" (133 podatka)
5.4. Podaci o finansijskim rezultatima poslovanja preduzeća "B" (327 podataka)
5.5. Godišnji izvještaj Centralne banke BiH za 2001. godinu
5.6. Primjer američke korporacije "Enron" i drugih kompanija

U novembru 2001, američka korporacija *Enron* je nadležnim državnim organima dostavila korigirani finansijski izvještaj o svom poslovanju, što je pokrenulo lavinu, po nju neugodnih, događanja uključujući i zvaničnu prijavu stečaja, te dizanje optužnice, a potom i osudu njenog predsjednika Arthura Andersena za ometanje pravosuđa. Krajem 2001. i početkom 2002. godine računovodstva i menadžmenti velikog broja američkih kompanija su učinili isto što i korporacija *Enron*, što im je donijelo neuobičajeno negativan publicitet. Sva ova lažiranja su otkrivena, u prvom redu, zahvaljujući primjeni Benfordovog zakona.



Komentar: I ovo je primjer s velikim brojem podataka, raspodjela prvih cifara iz raspoloživih podataka je jako dobra, s vrlo malim odstupanjima u odnosu na raspodjelu po Benfordovom zakonu, što snažno indicira na korektnost predočenih podataka. Izvjesna odstupanja kod značajnih cifri 7, 8 i 9 su nastala zbog manjeg broja podataka sa vještački postavljenim granicama, te određenog broja podataka sa ovim početnim ciframa koji se ponavljaju više puta (Vladin ulog u MMF, domaći depoziti, ...).

Američki matematičar dr Mark J. Nigrini je krajem augusta 2002. godine objavio rad pod naslovom *Procjena promjene broja slučajeva manipuliranja podacima o dobiti nakon afere*

Enron/Andersen u kojem je analizirao izvještaje o prihodima korporacije *Enron* i velikog broja drugih kompanija (koristio podatke iz *Digest of Earnings* – objavljene u *The Wall Street Journal* u aprilu i maju 2001., te aprilu i maju 2002. godine). Metodologija koju je koristio je bila bazirana na analizi brojčanih podataka o prihodima tretiranih kompanija koristeći Benfordov zakon koji daje očekivane proporcije pojavljivanja prve dvije cifre u tabelarnim podacima, te na analizi raspodjele numeričkih podataka o dobiti po dionici s obzirom na izražene diskontinuitete (skokove) i vrhove u njima.

Osnovno pitanje u ovom istraživačkom radu je bilo: Da li je usmjeravanje pažnje javnosti na računovodstva kompanija koindiciralo sa učestalošću manipulacija podacima o prihodima. Rezultat ovog istraživanja je bio slijedeći: U standardnim obrascima za podatke koji se daju u izvještajima o prihodima je utvrđeno manipuliranje podacima koji nisu tako očigledni, niti se mogu lako detektirati, te da se čak povećalo u periodu od 2001. do 2002. godine. Za podatke o prihodima koji su očigledni i javno poznati, utvrđeno je malo umanjenje manipuliranja, što je u vezi sa suptilnim potezima uprava kompanija da prikriju svoje aktivnosti.

6. REFERENCE

[1] Newcomb, S. (1881.): Note on the frequency of use of the different digits in natural numbers, *The American Journal of Mathematics* 4, 39–40

[2] Benford, F. (1938.): The law of anomalous numbers *Proceedings of the American Philosophical Society* 78, 551–572

[3] Hill, T. (1995.): Base-invariance implies Benford's law *Proceedings of the American Mathematical Society* 123, 887–895

[4] Hill, T. (1996.): A statistical derivation of the significiant-digit law *Statistical Science* 10, 354–363

[5] Nigrini, M. (1996.): A taxpayer compliance application of Benford's law *Journal of the American Taxtacion Association* 18, 72–91

[6] Mark Nigrini (novinski članak): I've Got Your Number *Journal of Accountancy*, may, 1999

[7] Lee Berton (novinski članak): He's Got Their Number: Scholar Uses Math to Foil Financial Fraud, *The Wall Street Journal*, july 10, 1995 (page 81)

[8] Robert McNatt (novinski članak): Up Front (A Penny Earned Is A Penny Fudged?) *Business week*, december 14, 1998

[9] Malcolm W. Browne (novinski članak): Following Benford's law, or Looking Out for No.1, *The New York Times* (Science), august 4, 1998

[10] Kevin Maney (novinski članak): Baffled by math?, *USA Today*, june 11, 2000 http://www.usatoday.com

[11] Steve Stanek (novinski članak): Applying Benford's Law to internal audit http://www.knowledgespace.com

[12] *Take Command*: Volume 12, №1 (USA)

[13] Benford's Law, Probability Function: http://mathworld.wolfram.com

[14] http://www.nigrini.com/images/Abstract2.htm

[15] http://www.nigrini.com/images/VersionForNigriniCom.htm

[16] http://www.cbbh.gov.ba

[17] **it** – Priručnik, knjiga 1, matematika-mehanika, *Rad*, Beograd,1970