**BENFORDOV ZAKON I IZBORI U SRBIJI**

# Dragan Azdejković[[1]](#footnote-1), e-mail: dragan.azdejkovic@ekof.bg.ac.rs

**Apstrakt.** Neki analitičari primenjuju Benfordov zakon na izborne rezultate da bi proverili da li postoje nepravilnosti, pod pretpostavkom da bi prirodno prebrojavanje glasova trebalo da prati ovu raspodelu. Dok je test prve cifre češći, test druge cifre se ponekad koristi jer je manje podložan namernim protivmerama. Međutim, on je takođe osetljiviji na osnovnu strukturu podataka, kao što je metod agregacije glasova, pa se njegovi rezultati moraju tumačiti oprezno. U ovom radu koristimo Benford zakon za analizu poslednjih izbora u Srbiji. U ovom radu pokazaćemo da takvi testovi rutinski ne uspevaju u podacima sa izbora u Srbiji.

**Ključne reči:** Benfordov zakon, izbori, druga cifra

**BENFORD'S LOW AND ELECTIONS IN SERBIA**

**Abstract.** Some analysts apply Benford’s Law to election results to check for irregularities, under the assumption that naturally occurring vote counts should follow this distribution. While the first-digit test is more common, the second-digit test is sometimes used because it is less susceptible to intentional countermeasures. However, it is also more sensitive to the underlying structure of the data, such as the method of vote aggregation, so its results must be interpreted cautiously.In this paper we use Benford low to analyze latest elections in Serbia. I this paper I will show that such tests routinely fail in data from elections in the Serbia.

**Key words:** Benford’s Low, elections, second digit.

**JEL klasifikacija (classification)**: C44, C46.

**Uvod**

Optužbe za prevaru i izbornu krađu deluju kao uvek prisutna komponenta demokratskog procesa. Iako se stvari možda nisu mnogo promenile istorijski, danas, bez obzira da li se glasanje tiče Argentine ili Francuske, Teksasa ili Istanbula, Srbije ili Crne Gore, pobednici se raduju, a poraženi tvrde da se radi o prevari. U najkorumpiranijim i autokratskim režimima koji traže demokratski legitimitet, posao posmatrača izbora liči na nemoguću misiju. Teškoća u postizanju konačne procene izbora na osnovu direktnog posmatranja je u tome što režimi mogu da podignu ogromne administrativne barijere koje onemogućavaju bilo kakav objektivan i održiv nadzor. S obzirom da su posmatrači podložni optužbama da rade sa političkim programima koji ne podstiču slobodne i poštene izbore, postoji neophodnost razvoja statističkih alata i indikatora koji, kada se primenjuju na zvanične izveštaje, pojačavaju nalaze i zaključke posmatranja i usmeravaju dalje istrage o legitimnosti izbora. Potraga za objektivnim metodama korišćenjem podataka koje je obezbedila državna izborna komisija počela je analizom Sobjanjina i Suhovolskog[[2]](#footnote-2) ruskih parlamentarnih izbora 1993. i ustavnog referenduma. Sobjanin primetio empirijski odnos koji se odnosi na širok spektar fenomena. Konkretno, pretpostavimo imamo promenljivu , gde ona referiše na populaciju grada *i*, prodaju korporacije *i*, energiju subatomske čestice *i*, ili čak populaciju vrste insekata *i*, gde je i posmatrajmo dijagram . Primetićemo da se date tačke mogu skoro savršeno aproksimirati pravom linijom sa negativnim keoficijentom pravca. Primenjujući ovu ideju na apsolutne glasove koje je nekoliko partija dobilo na ruskim parlamentarnim izborima 1993. godine sa proporcionalnom zastupljenošću održanim uporedo sa ustavnim referendumom, Sobjanin je primetio odstupanja od linearnog odnosa i to tumačio kao pokazatelj veličine prevare. Problemi ovakvog pristupa su skoro očigledni. Naime, ako su birači sofisticirani i ponašaju se strateški, a stranke moraju da dostignu neki prag zastupljenosti, onda možemo predvideti nelinearni pad udela glasova partija za koje se ne očekuje da će osvojiti mesta. Štaviše, ako važi Duvergerov zakon na izborima za jedan mandat, onda će se pojaviti još izraženiji diskontinuitet u udelima među svim strankama koje su treće rangirane.

Potraga za lako primenjivim indikatorom prevare nije se, naravno, završila sa Sobjanjinom. Nedavne primene Benfordovog zakona na izbore privlače pažnju kao lako primenljivi forenzički alat sa pratećom statističkom i matematičkom strogošću.

Benfordov zakon kaže da u mnogim skupovima podataka koji se pojavljuju u prirodi vodeće cifre nisu ravnomerno raspoređene; umesto toga, češće se pojavljuju manje cifre. Zakon se takođe može proširiti na druge cifre, iako je raspodela ravnija u poređenju raspodele sa prvim ciframa.

Dok prve cifre prate dobro poznatu logaritamsku raspodelu:

druge cifre (0-9) prate raspodelu koja je manje zakrivljena. Verovatnoća druge cifre data je sa:

Generalno, značajne cifre kod prebrojavanja glasova ne prate Benfordov zakon, ali je u nekoliko nedavnih istraživanja pronađeno da Benfordov zakon za drugu značajnu cifru približno dobro opisuje raspodelu drugih cifara kod glasanja. Iz sasvim uverljivih razloga napušten je fokus na prvu cifru izbornih podataka. Argument, u svom najjednostavnijem obliku, možda najbolje ilustruje zapažanje da ako se konkurentska trka dva kandidata događa u okruzima čija veličina varira između 100 i 1000, modalna prva cifra za glasove svakog kandidata neće biti 1 ili 2, već pre 4, 5, 3, ili 6. Još uvek ne postoji model – bilo kakva teorija – koja nas primorava da verujemo da nas izmanipulisani zbir glasova udaljava od predviđanja Benfordovim zakonom za drugu značajnu cifru i da slobodno i pošteno glasanje daje podatke koji su u skladu sa tom raspodelom.

U ovom radu tvrdimo da ovaj „zakon“ nije ništa manje sumnjiv kao sredstvo za otkrivanje izborne prevare nego što je Sobjanjinova adaptacija *log-rang* modela.

**Pregled literature**

Dijagnostičku upotrebu Benfordovog zakona druge cifre prilikom brojanja glasova na referendumu u Venecueli 2005 godine prvi su predložili Periči i Tores[[3]](#footnote-3), ali smo u izveštaju Kerterovog centra[[4]](#footnote-4) naišli na skeptičan odgovor zapažanjem da cifre u prebrojavanju glasova ne prate Benfordov zakon. Kako bi dijagnostifikovao verovatnu prevaru na izborima u Iranu 2009 Mebane[[5]](#footnote-5) je takođe koristio Benfordov zakon druge cifre (2BL). U radu Peričija i Toresa[[6]](#footnote-6) tvrdi se da Benfordov zakon primenjen na druge cifre u prebrojavanju glasova pruža dovoljan standard za dijagnostikovanje izborne prevare, ali se i ovde opet javljaju skeptični glasovi u radovima Šikana i Maka[[7]](#footnote-7), Lopeza[[8]](#footnote-8) i Mebanea[[9]](#footnote-9). Kantu i Sajah[[10]](#footnote-10) su otkrili da Benfordov zakon približno opisuje raspodelu prve cifre u rezultatima nekih izbora na nivou okruga na nekim izborima u Argentini.

U radu Mebanea[[11]](#footnote-11) naglašava se da uslovna sredina drugih značajnih cifara prebrojavanja glasova na biračkim mestima – statistika koju on označava ili – pomaže u dijagnosticiranju strategija koje birači koriste na izborima u Sjedinjenim Državama, Nemačkoj, Kanadi, Meksiku i drugim mestima. Prebrojavanje glasova je mešavina nekoliko različitih vrsta procesa: neki od njih određuju broj birača sa pravom glasa u svakoj jedinici; neki određuju koje kandidate bira svaki birač; neki način na koji se beleži izbor birača. Prema Rodrigezu[[12]](#footnote-12) i Grendaru, Džadžu i Šišteru[[13]](#footnote-13) takva mešavina može proizvesti brojeve koji prate raspodele slične Benfordu, ali ne i Benfordov zakon. Testove koje slede je predložio Mebane[[14]](#footnote-14).

Ne postoji formalna teorija koja bi podržala interpretacije obrazaca u drugim značajnim ciframa pri prebrojavanu glasova na biračkim mestima. Mebane[[15]](#footnote-15) otkriva da se u nekoliko zemalja — i sa višekružnim i sa mešovitim izbornim sistemima — druge značajne cifre u prebrojavanju glasova na biračkim mestima ponašaju na pravilan način koji odgovara strategijama koje birači koriste. Mobilizacija strategijom je samo jedna vrsta mobilizacije koja ostavlja redovne tragove u raspodelama cifara pri prebrojavanja glasova.

Pad na testu Benfordovog zakona ne znači nužno prevaru. Benfordov zakon je pozivan u diskusijama o izbornim prevarama, kao na primer tokom predsedničkih izbora u SAD 2020. godine, ali stručnjaci su istakli da to nije pouzdan alat za otkrivanje izborne prevare.

Zakon najbolje funkcioniše na skupovima podataka koji sadrže vrednosti u rasponu od veoma malih do veoma velikih brojeva, što nije uvek slučaj sa izbornim rezultatima.

**Podaci i metodologija**

Imajući u vidu osetljivost zaključaka na veličinu uzorka, saglasnost empirijske raspodele relativnih frekvencija na skupu podataka sa Benfordovim zakonom za drugu značajnu cifru proveravaćemo koristeći nekoliko statističkih testova i mera.

Pirsonov Hi kvadrat test je neparametarski statistički test koji se najčešće koristi prilikom poređenja raspodele frekvencija stvarnih podataka sa Benfordovim zakonom. Test ima veliku snagu za velike uzorke tako da će čak i sasvim mala odstupanja od Benfordovog zakona biti statistički značajna. Test je baziran na sledećoj statistici:

gde je *N* broj prebrojavanja sa 10 ili više glasova (tako da postoji druga cifra), je broj prebrojavanja koji ima drugu cifru i

Jedan od najpopularnijih neparametarskih testova za proveru saglasnosti dve raspodele je test Kolmogorov-Smirnova test (KS). Za datu kumulativnu funkcije raspodele pretpostavljene raspodele i empirijsku funkciju raspodele posmatranih podataka, ta test statistika je data sa:

Imajući u vidu da se u našem slučaju radi o Benfordovoj raspodeli za drugu značajnu cifru, koja je diskretna, test statistika u našem slučaju glasi:

gde je i .

Treća statistika inspirisana radom Grendara, Džadža i Šištera[[16]](#footnote-16), je srednja vrednost drugih cifara, označena . Ako raspodela drugih cifara kod prebrojanja ima frekvencije kao što je dato u Benfordovoj raspodeli, onda je srednja vrednost za drugu cifru

U matematičkoj statistici, Kulbek–Lejbler (KL) divergencija je vrsta statističke metrike, odnosno mere koliko se raspodela verovatnoće modela razlikuje od prave raspodele[[17]](#footnote-17). Ako je ta mera jednaka nuli, onda možemo reći da su dve date raspodele identične. U kontekstu našeg problema KL divergenciju možemo izračunati na sledeći način

U statistici, Bhatačarijevo rastojanje je broj kojim se meri sličnost između dve raspodele verovatnoće[[18]](#footnote-18). Slično kao i kod Kulbek-Lejblerove mere, bliskost 0 refereriše na sličnost između dve posmatrane raspodele. Bhatačarijevo rastojanje između Benfordove raspodele za drugu značajnu cifru i raspodele dobijene iz podataka računamo na sledeći način:

Konačno, koristićemo i test medijane apsolutnih odstupanja (MAO) koji je predložio Nigrini:

gde se empirijskim putem došlo graničnih vrednosti za MAD na osnovu kojih opisujemo saglasnost sa raspodelom.

Dobra vest pri istraživanju saglasnosti podataka prilikom prebrojavanja glasova je dostupnost podataka, što kod drugih primena Benfordovog zakona obično nije slučaj. Za sprovođenje ovog istraživanja preuzeti su podaci sa predsedničkih izbora od 2002. godine do 2022. godine Republičkog zavoda za statistiku Republike Srbije[[19]](#footnote-19). Za testiranje i analizu podataka korišćen je programski jezik Pajton, a odgovarajuća radna beležnica je dostupna na Githab-u[[20]](#footnote-20).

Saglasnost raspodele ispitivana je za najmanje 3 kandidata koji su u prvom krugu dobili najviše glasova. Elementi uzorka su izborna mesta, a obimi uzoraka su veći od 8500.

**Rezultati i diskusija**

Rezultati testiranja dati su u sledećoj tabeli. Brojevi u drugoj koloni predstavljaju broj glasova za kandidate koji su u prvom krugu dobili najviše glasova. U koloni data je *p*-vrednost *t* statistike, dok je kod testova Hi kvadrat i KS dat samo ishod testiranja, gde + označava saglasnost. Ishodi testa MAD su bliska saglasnost (BS), prihvatljiva saglasnost (PS), slaba saglasnost (SS) i nesaglasnost (N), pa su na taj način dati ishodi u koloni MAD. Statističke mere DKL i DB nisu normalizovane, ali svakako njihova bliskost 0 ukazuje na saglasnost raspodela. Konačno, ako neki test sugeriše nesaglasnost, u koloni Z je opisano kod kojih cifara ili cifre dolazi do značajnog odstupanja, što je provereno *z*-statistikom.

|  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- |
| Tabela 1. Predsednički izbori u Republici Srbiji u periodu od 2022-2022 | | | | | | | | |
| Godina | Glasovi |  | KM | MAD |  |  |  | *Z* |
| 2002 | 1.123.420 | + | + | BS | 0.0003 | 0.0003 | 0.0001 |  |
| 995.200 | + | + | BS | 0.0010 | 0.0003 | 0.0001 |  |
| 854.308 | + | + | BS | 0.0008 | 0.0002 | 0.0001 |  |
| 2003 | 1.166.896 | + | + | BS | 0.0006 | 0.0002 | 0.0001 |  |
| 893.906 | + | + | BS | 0.0005 | 0.0001 | 0.0002 |  |
| 229.229 | - | - | BS | 0.0323 | 0.0081 | 0.0002 | 0, 8, 9 |
| 2004 | 954.339 | + | + | BS | 0.0015 | 0.0037 | 0.0000 |  |
| 853.584 | + | + | BS | 0.0006 | 0.0002 | 0.0000 |  |
| 568.691 | - | - | BS | 0.0015 | 0.0004 | 0.0002 | 8 |
| 414.971 | + | + | BS | 0.0009 | 0.0002 | 0 | 0 |
| 2008 | 1.646.172 | + | + | BS | 0.0005 | 0.0001 | 0 |  |
| 1.457.030 | + | + | BS | 0.0005 | 0.0001 | 0 |  |
| 305.828 | + | + | BS | 0.0009 | 0.0002 | 0 |  |
| 2012 | 989.454 | + | + | BS | 0.0005 | 0.0001 | 0 |  |
| 979.216 | + | - | BS | 0.0010 | 0.0002 | 0 | 8, 9 |
| 556.013 | - | - | BS | 0.0022 | 0.0006 | 0 | 0,1,6,7 |
| 2017 | 2.012.788 | + | - | BS | 0.0009 | 0.0002 | 0 | 7 |
| 597.728 | - | - | BS | 0.0016 | 0.0004 | 0 | 2, 7 |
| 344.498 | - | - | BS | 0.0021 | 0.0005 | 0 | 6, 9 |
| 2022 | 2.224.914 | + | - | BS | 0.0010 | 0.0002 | 0 | 1, 5 |
| 698.538 | + | + | BS | 0.0009 | 0.0002 | 0 |  |
| 226.137 | - | - | BS | 0.003 | 0.0008 | 0 | 0, 7, 8 |

Ono što je je očigledno iz prethodne tabele je da MAD test uvek ima ishod bliska saglasnost. Takođe, *p* vrednost statistike vrlo bliska nuli što sugeriše uvek isti zaključak: hipoteza da je srednja vrednost druge značajne cifre 4.187 se odbacuje.

Imajući u vidu osetljivost ovih testova na veličinu uzorka ti rezultati su očekivani.

Hipoteza da je prebrojavanje glasova sa biračkih mesta raspoređeno po Benfordovom zakonu za drugu značajnu cifru se ne odbacuje za većinu kandidata kada se hipoteza testira korišćenjem podataka sa predsedničkih izbora u Republici Srbiji 2002–2022. Od 44 primera testiranja prikazanih u tabeli 1., hipoteza je odbačena 14 slučajeva. Srednja vrednost druge značajne cifre značajno se razlikuje od teorijske. Takođe, ta srednja vrednost nije stabilna tokom vremena.

Kulbek–Lejbler (KL) divergencija i Bhatačarijevo rastojanje ukazuju da druga značajna cifra ima raspodelu vrlo slična Benfordovom zakonu za drugu značajnu cifru.

Ovi obrasci vode do činjenice da birači u Republici Srbiji obično deluju strateški i delimično skladu sa instrumentalnom racionalnošću.

**Zaključak**

Podaci sa izbora u Srbiji ne samo da mogu pomoći u dijagnosticiranju strateškog glasanja, već su osetljivi i na druge aspekte normalne politike kao što su vrste mobilizacije koje prevazilaze okvire strateškog glasanja. Tvrdnja da kršenje Benfordovog zakona pruža dovoljan standard za dijagnosticiranje izborne prevare je gotovo sigurno lažna, ali Benfordov zakon nije irelevantan za forenzičko ispitivanje izbora. Prebrojavanje glasova na biračkim mestima u Srbiji često se poklapa sa raspodelom koju podrazumeva 2BL.

Nije lako odgovoriti na pitanje da li se obrasci u ciframa proizvedenim prevarama oštro razlikuju od obrazaca koje proizvodi normalna politika. Dosadašnja istraživanja pružaju slučajeve u kojima je verovatna prevara proizvela veoma karakteristične obrasce (kao što su iranski izbori 2009. i drugi izbori), ali i slučajeve u kojima prirodni politički šokovi proizvode obrasce koji bi se inače tumačili kao politička prinuda. Podaci sa izbora u Srbiji ne sadrže obrasce koji bi proizveli raspodelu relativnih frekvencija druge značajne cifre koja se značajno razlikuje od Benfordove raspodele.

**Literatura:**

Cantu F. and Saiegh F. (2011), „Fraudulent Democracy? An Analysis of Argentina’s Infamous Decade Using Supervised Machine Learning“, *Political Analysis* 19, no. 4, 409–433.

Dodge, Y. (Ed.). (2003). *The Oxford dictionary of statistical terms*. Oxford University Press, USA.

Kullback, S., & Leibler, R. A. (1951). On information and sufficiency. *The annals of mathematical statistics*, *22*(1), 79-86.

Lopez J. A. G. (2009), “Fraude Electoral?“, *Doble Hélice Ediciones*, Chihuahua, Mexico.

Grendar M., Judge G. and Schechter L. (2006), “An Empirical Non-parametric Likelihood Family of Data-Based Benford-Like Distributions“, *Physica A: Statistical Mechanics and its Applications.*

Mebane W. (2006), “Election Forensics: Vote Counts and Benford’s Law”, *Paper prepared for the 2006 Summer Meeting of the Political Methodology Society*, UC-Davis.

Mebane W. (2009), Note on the Presidential Election in Iran,

[http://www.umich.edu/∼](http://www.umich.edu/%E2%88%BCwmebane/note29jun2009.pdf)wmebane/note29jun2009.pdf.

Mebane W. (2010), “Election Fraud or Strategic Voting?“, *Paper prepared for the 2010 Annual Meeting of the Midwest Political Science Association*, Chicago, IL, April 22–25, 2010.

Mebane W. (2010), “*Fraud in the 2009 Presidential Election in Iran?“*, *Chance* 23, no. 1, 6–15.

Mebane W. (2011), “Comment on ‘Benford’s Law and the Detection of Election Fraud“, *Political Analysis* 19, no. 3, 269–272.

Mebane W. (2010), “Second-Digit Tests for Voters’ Election Strategies and Election Fraud*”*, Paper prepared for the 2012 Annual Meeting of the Midwest Political Science Association, Chicago, IL, April 11–14, 2012.

Mebane W. (2013), “ Election Forensics“, *Book MS*

Rodriguez R. (2004), “ First Significant Digit Patterns from Mixtures of Uniform Digits“, *American Statistician.*

Pericchi L. and Torres D. (2004), “La Ley de Newcomb-Benford y sus aplicaciones al Referendum Revocatorio en Venezuela*“*, *Reporte Técnico no- definitivo 2a. versión: Octubre 01, 2004*.

Pericchi L. and Torres D. (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, 502–516

Shikano S. and Mack V. (2009), “When Does the Second Digit Benford’s Law Test Signal an Election Fraud? Facts or Misleading Test Results“, *Jahrbücher fur Nationalökonomie und Statistik* 231, no. 5–6, 719–732.

Sobyanin, A., and Suchovolsky V., (1993), ‘’Elections and the referendum December 11, 1993, in Russia’’. *Unpublished report to the Administration of the President of the RF*, Moscow.

<https://www.cartercenter.org/documents/2020.pdf> [Pristupljeno: 27/03/25]

<https://www.stat.gov.rs/sr-latn/oblasti/izbori/predsednicki-izbori/>[Pristupljeno: 21/04/25]

<https://github.com/dazdejkovic/Serbian-elections-2BL>[Pristupljeno: 10/05/25]

1. Univerzitet u Beogradu, Ekonomski fakultet, Kamenička 6, 11000 Beograd, Srbija [↑](#footnote-ref-1)
2. Sobyanin (1993) [↑](#footnote-ref-2)
3. L. Pericchi and D. Torres (2004) [↑](#footnote-ref-3)
4. https://www.cartercenter.org/documents/2020.pdf [↑](#footnote-ref-4)
5. Mebane W. (2006), Mebane W. (2009), Mebane W. (2010) [↑](#footnote-ref-5)
6. L. Pericchi and D. Torres (2011) [↑](#footnote-ref-6)
7. Shikano S. and Mack V. (2009) [↑](#footnote-ref-7)
8. Lopez J. A. G. (2009) [↑](#footnote-ref-8)
9. Mebane W. (2011) [↑](#footnote-ref-9)
10. Cantu F. and Saiegh F. (2011) [↑](#footnote-ref-10)
11. Mebane W. (2013) [↑](#footnote-ref-11)
12. Rodriguez R. (2004) [↑](#footnote-ref-12)
13. Grendar M., Judge G. and Schechter L. (2006) [↑](#footnote-ref-13)
14. Mebane W. (2006) [↑](#footnote-ref-14)
15. Mebane W. (2013) [↑](#footnote-ref-15)
16. Grendar M., Judge G. and Schechter L. (2006) [↑](#footnote-ref-16)
17. Kullback, S., & Leibler, R. A. (1951). On information and sufficiency. *The annals of mathematical statistics*, *22*(1), 79-86. [↑](#footnote-ref-17)
18. Dodge, Y. (Ed.). (2003). *The Oxford dictionary of statistical terms*. Oxford University Press, USA. [↑](#footnote-ref-18)
19. https://www.stat.gov.rs/sr-latn/oblasti/izbori/predsednicki-izbori/ [↑](#footnote-ref-19)
20. https://github.com/dazdejkovic/Serbian-elections-2BL [↑](#footnote-ref-20)