Kitalálható-e az első számjegyekből a korrupció mértéke?

Pete Gábor és Timár Ádám

Kivonat

Korrupciókutatók megállapításai szerint a 2009 és 2015 közötti EU által finanszírozott magyar közbeszerzések értékeiben a Benford törvény megjelenése lenne várható, de az adatok nem követik a törvényt. Ebből a szerzők arra következtetnek, hogy a közbeszerzési árak manipulálása, felülárazás érhető tetten. Egy általuk kifejlesztett módszerrel kvantitatív becslést is adnak a túlárazás mértékére, és mint irják, „az 1,7 szerestől a közel 10 szeres túlárazás sem lehet ritka". Cikkünkben arra szeretnénk rámutatni, hogy az említett módszer matematikailag hibás, a Benford eloszlás ilyen típusú számszerű becslésre lényegében alkalmatlan. Ettől függetlenül, a kiindulási hipotézis is kétséges, vagyis, hogy a felhasznált adatsor túlárazás nélkül ezt az eloszlást követné.

1. Bevezetés

Hogyan szűrjük ki, ha egy adóbevallás fiktív számlaösszegeket tartalmaz, anélkül, hogy akár csak egy pillantást vetnénk a számlákra?

A Benford törvény választ ad a problémára [1, 2]. Azt mondja ugyanis, hogy az ilyen típusú adathalmazoknál az első számjegyeknek valamilyen jól meghatározott gyakoriságot (az úgynevezett Benford eloszlást) kell követniük. Például, azon megfigyelések gyakorisága, ahol az első számjegy egyes, az összes adatnak körülbelül 30%-a, ahol kettes, az 18%, és így tovább. „De hiszen ez lehetetlen! Ha egy véletlen bevételi elemet kell hasraütés-szerűen bemondanom, bármelyik számjegyet ugyanolyan eséllyel választom kezdő számjegynek! Miért lenne más ez a természet által generált véletlen számoknál??" — reagálnak sokan, mikor először hallanak a törvényről. Hogy milyen jelenség áll a Benford törvény hátterében, azt hamarosan kifejtjük. Képzeletbeli vitapartnerünk pedig, aki hasraütéssel generált véletlen számokat, pont azzal bukna le, ha ilyen adatsort probálna „generálni", hogy az ő adatsorában nagyjából ugyanolyan eséllyel szerepelne bármelyik véletlen számjegy, míg a valódi, adózott jövedelmekből származó adatsorban a Benford eloszlást követné. Ez a leleplező erő áll a Benford eloszlás sikeres alkalmazásai mögött, amikor például adócsalások felderítésére szolgált. Az elv tehát: ha tudjuk, hogy az adatsornak a „természetben" Benford eloszlást kellene követnie, és mégis nagyon eltér tőle, akkor valaki óhatatlanul manipulálta azt.

A közelmúltban nagy nyilvánosságot kapott egy friss magyar kutatás [3, 4]. Tóth István János és Hajdu Miklós, a Corruption Research Center Budapest munkatársai arra jutottak, hogy a 2009 és 2015 közötti, EU által finanszírozott magyar közbeszerzéseknél az árak manipuláltak, „akár 1,7-10-szeres túlárazások is voltak". Módszerükben központi szerepet kap a Benford eloszlás, és annak újszerű alkalmazása. Ezzel az alkalmazással azonban komoly problémák vannak. Mint a továbbiakban bemutatjuk, a szerzők számszerű konklúzióit nem bizonyítja, de még csak kevéssé rigorózus formában sem támasztja alá a módszerük. Az persze egy érdekes és figyelemre méltó észrevétel, hogy a magyar forrásból finanszírozott közbeszerzések teljesítik a Benford törvényt, míg az EU által finanszírozottak nem. Ámde, majdnem elhagyagolható mértékű túlárazással is el lehet érni a Benford-törvény tapasztalt megsértését, és az sem látszik, hogy milyen természetes korrupciós mechanizmusok alakíthatnák ki pontosan ezt a fajta torzulást. Így teljesen megalapozatlannak tűnik a túlárazás mértékére vonatkozó következtetés.

2. Mi a Benford-eloszlás és miért olyan gyakori?

Milyen jelenseg áll a Benford törvény mögött? Mi az oka, amikor az a természetben megfigyelhető? Jelöljük X-szel azt a véletlen számot, ami egy véletlenszerűen választott közbeszerzés valódi (túlárazás nélküli) ára volna.

Ha X folytonos egyenletes eloszlású lenne például az $[1, 1000000]$ intervallumon, akkor az első számjegye nyilván egyenletes eloszlású lenne az $\{1, 2, \ldots, 9\}$ halmazon. De ha X az $[1, 2000000]$ intervallumon lenne egyenletes eloszlású, akkor az első számjegye több mint 50% valószínűséggel 1-es lenne. De melyik volna a természetesebb modell sok nagyságrenden átívelő véletlen számokra? A Benford törvény szerint egyik sem, hanem a következő. Vegyük észre, hogy X-nek a 10-es számrendszerbeli felírása pontosan akkor kezdődik k-val (ahol $k = 1, 2, \ldots, 9$), ha $\log_{10} X$ törtrésze $\log_{10}(k+1)$ és $\log_{10} k$ között van. **A Benford törvény szerint $\log_{10} X$ törtrészének kell egyenletes eloszlásúnak lenni a $[0, 1]$ intervallumon.** Így például, mivel $\log_{10} 1 = 0$ és $\log_{10} 2 \approx 0,301$,

$$\mathbb{P}\big[X \text{ első számjegye } 1\big] \approx \mathbb{P}\big[\log_{10} X \text{ törtrésze } 0 \text{ és } 0,301 \text{ közé esik}\big] \approx 0,301\,,$$

és így tovább: a Benford törvénynek engedelmeskedő adathalmazban azon számok aránya, amiknek k az első számjegye, $\log_{10} \frac{k+1}{k}$-hoz közeli. Természetesen nagyon sok ilyen X eloszlás van; talán a legegyszerűbb, ha $\log_{10} X$ maga egyenletes eloszlású egy $[N, M]$ alakú intervallumon, ahol N, M egészek, vagy $M - N$ jó nagy; avagy, ami ezzel ekvivalens, X sűrűségfüggvénye $f(x) = c/x$ alakú ezen az intervallumon. De a $\log_{10} X$ egyenletes eloszlásánál sokkal könnyebben teljesíthető feltétel, hogy az eloszlás egy nagy intervallumon pozitív, és sűrűségfüggvénye lassan változik, azaz minden egyes $[N, N+1]$ alakú részintervallumon közel egyenletes. Ilyenkor is közelítőleg a Benford-eloszlást fogjuk kapni.

Mi okozza, hogy sok való életből vett adat teljesíti a Benford törvényt? Először is, megemlítjük a Benford eloszlás egy különleges tulajdonságát, a *skálafüggetlenség*et. Ha az adatsor számait ugyanazzal a c konstanssal megszorozzuk, az adatsornak továbbra is teljesítenie kell a Benford törvényt, hiszen ha $\log_{10} X$ törtrésze egyenletes a $[0, 1]$-en, akkor $\log_{10} c + \log_{10} X$ törtrésze is az. Így például ha az adathalmazunk a mai napon keletkezett számlákbol áll, nem számít, hogy mindegyik forintban, vagy mindegyik euróban van-e megadva. Hasonlóképpen, $\log_{10}$ helyett bármilyen más alapú logaritmussal is dolgozhatnánk, amiből pedig következik, hogy ha az adatsorunkat valamilyen d alapú számrendszerben írjuk le, akkor az első számjegy $\log_d(k+1)/k$ valószínűséggel lesz k. Namármost, könnyű bizonyítani, hogy a skálafüggetlenség nem csak konstans számmal való szorzásra teljesül, hanem akkor is, ha egy az adatsortól független véletlen változóval szorozzuk meg az adatsor elemeit. Például, ha az adatsorban szereplő mindegyik számot megszorozzuk 2-vel vagy elosztjuk 3-mal, érmefeldobás alapján, a keletkező új adatsor is teljesíti a törvényt, ha az eredeti teljesítette azt. Egy ennél erősebb tulajdonság is igaz: tegyük fel, hogy a kiindulási adasorunk „nagyjából” (legfeljebb 2%-os hibával, mondjuk), teljesítette a Benford törvényt. Szorozzuk most meg egy független véletlen változóval az adatokat. Ekkor a keletkező új adatsor is teljesíti nagyjából a Benford törvényt, ráadásul a hibaszázalék jó eséllyel csökken! Pont ez az *abszorpciós tulajdonság* (amiből következik a Benford törvény) magyarázhatja, miért fordul elő olyan gyakran a természetben: ha független véletlen változókat szorozgatunk össze, a szorzat egyre inkább Benford-szerű viselkedést mutat. Tehát ha például az adatsorunk hosszú ideig exponenciális növekedésnek volt kitéve, független véletlennek tekinthető hányadosokkal, mint például különböző termékek árai sokévi infláció hatására, vagy nagyvárosok lélekszáma az urbanizáció hatására, akkor a folyamat végeredménye reálisan kielégítheti a Benford törvényt. Ezért feltételezhetjuk azt is, hogy a közbeszerzések (túlárazás előtti) árai engedelmeskednének Benford törvényének: egy villanykörte ára, az egy utcában található villanykörték száma, egy kisváros utcáinak száma... ezek szorzataként áll elő egy sikeres közvilágítás-tender költsége.

Mélyebb matematikai képzettséggel rendelkező olvasóink számára egy kicsit konkrétabb (bár még mindig pontatlan) érvelést is adunk — a többiek nyugodtan ugorják át a bekezdést. Tegyük föl, mint az előbb, hogy az X végső ár sok véletlen tényező szorzata, $X = X_1 \cdots X_n$. Ekkor $\log X = \log X_1 + \cdots + \log X_n$, és az a kérdés, vajon ennek a törtrésze miért egyenletes eloszlású a $[0,1]$ intervallumon. Igen gyakori jelenség, hogy valószínűségi változók összege valamilyen skálázással valamilyen véletlen határértékhez tart:

$$\frac{Y_1 + \cdots + Y_n - \mu(n)}{\sigma(n)} \to Y\,.$$

A legismertebb példa a Centrális Határeloszlás Tétel: ha az Y_i-k független azonos eloszlásúak, véges μ várható értékkel és σ szórással, akkor $\mu(n) = n\mu$ és $\sigma(n) = \sqrt{n}\sigma$ választással az Y limesz standard normális. De rengeteg egyéb példa van, sokkal általánosabb körülmények között is, így egyáltalán nem furcsa föltételezés, hogy a mi $\log X_1 + \cdots + \log X_n$ összegünk is legyen jól közelíthető $\sigma(n)Y + \mu(n)$ alakban, ahol $\sigma(n)$ tart a végtelenbe, Y pedig valamilyen épeszű valószínűségi változó, egy egyenletesen folytonos $f(x)$ sűrűségfüggvénnyel. Ekkor, tetszőleges $0 < a < b < 1$-re,

$$\begin{aligned}
\mathbb{P}\big[\sigma(n)Y + \mu(n) \text{ törtrésze } (a,b)\text{-ben van}\big] &= \sum_{k=-\infty}^{\infty} \mathbb{P}\big[\sigma(n)Y + \mu(n) \in (a,b) + k\big] \\
&= \sum_{k=-\infty}^{\infty} \mathbb{P}\Big[Y \in \frac{(a,b) + k - \mu(n)}{\sigma(n)}\Big] \\
&\approx \sum_{k=-\infty}^{\infty} f\left(\frac{k + a - \mu(n)}{\sigma(n)}\right) \frac{|b-a|}{\sigma(n)}\,,
\end{aligned}$$

ahol a közelítés azért állja meg a helyét, mert a két végtelen összeg tagjai az egyenletes folytonosság miatt egyenletesen közel vannak egymáshoz. Az utolsó szumma viszont a

$$|b-a| \int_{-\infty}^{\infty} f\left(x + \frac{a - \mu(n)}{\sigma(n)}\right) dx = |b-a| \int_{-\infty}^{\infty} f(x)\,dx = |b-a|$$

integrál egy közelítőösszege. Így hát a kérdéses valószínűség közel van $|b-a|$-hoz, ha n elég nagy, azaz közelítőleg a Benford-törvényt kaptuk.

3. A túlárazásra adott becslés hibái

De térjünk vissza az idézett szerzők állítására: „az 1,7-szerestől a közel 10-szeres túlárazás sem lehet ritka”. Ez a kijelentés önmagában is különös: az 1,5-szeres túlárazás ritka lenne, de az 1,7-szeres már nem? De nézzük, hogyan áll elő a fenti becslés. Ha a közbeszerzések Benford eloszlást követnének, ahhoz képest a megfigyelt adatsorban „az 1-es, és 2-es számjegyek, mint első számjegyek gyakrabban, és a 3-as, 4-es, 5-ös, és 6-os számjegyek a vártnál ritkábban fordulnak elő az EU által finanszírozott közbeszerzéseknél”. (Vegyük észre, hogy a megfigyelt eloszlás pont az ellenkező irányba tér el a Benfordtól, mint a szokásos adatmanipulálási esetekben szokott: még kevésbé egyenletes.) A szerzők ebből arra következtetnek, feltéve, hogy az alulárazás nem életszerű, hogy a „hiányzó” 3-mal, 4-gyel, 5-tel vagy 6-tal kezdődő számok helyett 1-gyel vagy 2-vel kezdődő számok szerepelnek, tehát például 60 milliós projekt 100 milliósra lett túlárazva (1,7-szeres szorzó), és 30 milliós 299 milliósra (közel 10-szeres szorzó). Figyelmen kívül hagyják azonban azt a lehetőséget, hogy egy „hiányzó” 6-tal kezdődő projektből úgy is eljuthatunk egy „extra” 1-gyel kezdődőbe, ha például egy 60 millióst 80 millióra, egy 80 millióst pedig

100 millióra áraztak. Ebben a (szintén hasraütés-szerű) esetben mindkét túlárazás kevesebb, mint 1,34-szeres! Ha az ilyen lehetőségeket teljes mértékben kihasználjuk, akkor a két valószínűségi eloszlás (az első számjegyek elméleti — Benford — és megfigyelt eloszlása) közti úgynevezett optimális transzport (vagy Wasserstein) távolságot kell vennünk, ami lényegében azt ragadja meg, hogyan lehetséges a lehető legkevesebb torzítással átvinni egyik eloszlást a másikba. Mint látni fogjuk, **szinte elhanyagolható túlárazással** is meg lehet oldani a Benford eloszlás megfigyelt torzulását. A szerzők érvelésében egy másik probléma, hogy ha egy 3-mal kezdődő szám 2-vel kezdődőre árazódik túl, ők automatikusan élnek azzal a feltevéssel, hogy 30 milliósból kaptunk 299 milliósat. Miért nem 39 milliósból 200 milliósat? (Így közel tízszeres szorzó helyett csupán ötszöröst kapnánk!) Egy alsó becsléshez ez a feltevés lenne egyedül elfogadható. Máskülönben önkényesen akár azt is feltehetnénk, hogy egy 30 milliós munka 2 vagy 20 milliárdra lett túlárazva...

A következő kérdéses pont, hogy vajon pontosan milyen manipuláció okozhatja, hogy adatsorunk elbukik a Benford törvényen, ráadásul nem a szokásos „egyenletes" módon? Ha nincs elképzelésünk, hogy az esetleges túlárazások hogyan magyarázhatják a megfigyelt torzulást, akkor érdemes lenne egyéb lehetséges okokat is keresni. Ha a túlárazás annyit jelentene, hogy minden közbeszerzés ára szorzódik 2-vel, a kapott új adatsor ugyanúgy követné a Benford törvényt, ha a kiindulási követte azt (mint ezt az „abszzorpciós tulajdonság" ismertetésénél láttuk). Ez általánosabban akkor is igaz, ha az összes árat egy tőlük független véletlen változóval szorozzuk. Tehát ilyen esetben nem várhatnánk, hogy a Benford törvény leplezze le a csalást. Akkor lehet esetleg esélye a módszernek, ha a túlárazást meghatározó szorzó függ a torzítatlan ártól (például ha fölfelé kerekítésekkel működik a túlárazás). Ez esetben viszont az válik problematikussá, hogy a szerzők a forintban és idegen devizában denominált közbeszerzéseket összevonva vizsgálják, az utóbbiakat az adott év árfolyamán váltva át forintra. A kerekítéses példánál maradva, egy euróban kerekített érték forintra átváltás után már nem lesz kerekítve, tehát az első számjegyekre gyakorolt hatás kiszámíthatatlanná válik.

Az a feltételezés, hogy a torzítatlan árak csak fölfelé változhatnak, elsőre jogosnak tűnik, hiszen ki akarna kevesebb pénzt kapni az EU-tól, mint amennyit gondolna, hogy a munkája ér. Ám ez sem feltétlenül ilyen egyszerű: ha a kiírás egy bizonyos munkára 1,99 millió euróban maximalizálja a költséget, de egy cég a szokásos haszonkulcsával 2,2 millió euróra hozná ki az árajánlatát, lehet, hogy meggondolja, hogy a nagy megrendelés kedvéért most kivételesen kisebb haszonkulccsal dolgozzon. Más szóval, könnyen lehet, hogy korrupció nélkül is lehet a Benford eloszlástól való eltérést indokolni: a közbeszerzési értékhatárok, az EU-s keretösszegek, amik ráadásul iparáganként különbözőek lehetnek, valószinűleg befolyásolják az adatok eloszlását. Kérdés persze, hogy indokolt-e, hogy ezek a hatások az EU-s közbeszerzések esetében torzítják a Benford eloszlást, míg a hazai forrású közbeszerzések esetében nem. Hogy ezt megítéljük, ahhoz egyrészt pontos jogi és közgazdasági információk kellenének, másrészt az összegyűjtött adatsor, amiből a szerzők dolgoztak, nem nyilvános; így ez ügyben csak spekulálni tudunk.

4. Hogyan lehetne helyesebben?

Most visszatérünk az **eloszlások optimális transzport távolságára**, amit használnunk kell, ha helyes alsó becslést akarunk adni a túlárazás mértékére. (És még itt se feledkezzünk meg arról, hogy két komoly feltevéssel élünk: egyrészt, hogy a manipulálatlan eloszlás Benfordot követ, másrészt, hogy manipuláció csak fölfelé történhet.) Általában, ha van két eloszlásunk egy metrikus téren, akkor az ő *optimális transzport* (más néven *földmunkás* vagy *Wasserstein*) távolságukon a következőt értjük [5]: gondoljunk úgy a kérdésre, hogy van egy földkupacunk az egyik eloszlás szerint szétosztva a térben, és ezt minél kisebb munkával szeretnénk átpakolni úgy, hogy a végeredmény a másik eloszlást kövesse. Azaz, az átlagos távolságot akarjuk minimalizálni, amennyivel a földszemcséket mozgatnunk kell.

Jelen esetben a metrikus terünk igen egyszerű: a lehetséges árak halmaza, vagy leegyszerűsítve, az első számjegyeik által reprezentálva. Továbbá most nem a tényleges minimumot keressük, hanem élünk azzal a föltevéssel, hogy csak fölfelé mozgathatunk tömeget, így egyfajta „monoton transzport” távolságról beszélhetünk. Ha az értékek logaritmusával dolgozunk, akkor a Benford törvénynek az első számjegyeken adódó Benford eloszlás felel meg, és ennek a monoton transzport távolsága a valóságban mért eloszlástól pont annak fog megfelelni a valóságos (a logaritmus-vétel előtti) árakat illetően, hogy átlagosan legalább hányszoros túlárazást kell eszközölnünk.

Az értékek logaritmusán a valóságban megfigyelt eloszlást mi nem ismerjük, csak a szerzők által közölt eloszlást az értékek logaritmusainak első számjegyein. Mutatunk egy nagyon könnyen átlátható monoton transzportot a Benford eloszlásból ebbe az eloszlásba:

	EU	33,1	20,2	11,1	8,2	6,0	5,2	5,7	4,7	5,7
Benford		1	2	3	4	5	6	7	8	9
30,1	1		2,6							
17,6	2			0						
12,5	3				1,3					
9,7	4					2,8				
7,9	5						4,7			
6,7	6							5,7	0,5	
5,8	7								4,2	1,6
5,1	8	1,0								4,1
4,6	9	4,6								

1. táblázat. A Benford eloszlás egy monoton transzportja az EU-s közbeszerzéseknél megfigyelt eloszlásba. Például: a 8-as első számjegyen a Benford eloszlásban 5,1% súly van; ebből 4,1%-ot átrakunk a 9-esbe és 1%-ot az 1-esbe (persze az eggyel nagyobb nagyságrend 1-esébe); végül a 7-esből idehozunk 4,2%-ot, a 6-osból pedig 0,5%-ot, hogy meglegyen az EU-s eloszlás 4,7%-a.

Namármost, ha a szerzők megközelítését alkalmazzuk, és torzítatlannak veszünk minden ár-eloszlást, ami az első számjegyeken a Benfordot adja, akkor legalább mekkora túlárazás kell, hogy egy ilyenből el lehessen jutni egy olyan áreloszlásba, ami az EU-s eloszlást adja az első számjegyeken? Lehetséges, hogy például a Benfordot teljesítő eloszlás 6,7%-os súlya a 6-os számjegyen mind 699... alakú. Ha ezeket megszorozzuk 1,0015-tel, akkor 7-essel kezdődő árakat fogunk kapni. Azaz, a szomszédos számjegyek között tetszőlegesen kicsi túlárazással is tudunk súlyokat mozgatni (ha a Benfordot teljesítő eloszlásunk pont ilyen furcsán fölfelé billenő). Így biztos alsó korlátot a túlárazásra csak a $6 \to 8$, $7 \to 9$, és $8 \to 1$ esetekben kapunk, mégpedig 8 / 6,99 és 9 / 7,99 és 10 / 8,99 arányokkal, rendre 0,5%, 1,6%, és 1% súlyokkal. Mindez **átlagosan kevesebb mint 4 ezrelékes** túlárazást jelent.

Ha nem akarunk ennyire precízek lenni, minden lehetséges furcsa lehetőséget lefedni, akkor kissé önkényesen, de a valósághoz valószínűleg közelebb állóan, gondolhatunk úgy az első számjegyeken adott eloszlásokra, mint a valóságos eloszlások diszkrét közelítéseire, és így a k-s első számjegyből az ℓ-esbe mozgatáskor egyszerűen egy ℓ/k szorzóval számolhatunk. Ekkor a táblázatbeli transzport 8% körüli, azaz 1,08-szoros átlagos túlárazást ad. Ez már nem mérhetetlenül kicsi, de egyrészt még mindig nem a pontos optimum, másrészt az 1,7-10-szeres szorzóhoz képest még mindig elenyésző. Nem igaz tehát a szerzők azon állítása, hogy az „ilyen nagymértékű túlárazásoknak is számottevő gyakorisággal elő kellett fordulniuk az EU által finanszírozott közbeszerzéseknél Magyarországon 2009-15 között. Másképp nem lehetséges.”

Mint láthattuk, megnehezítette a pontos kvantitatív becslést, hogy a szerzők csak az első számjegyeken való Benford eloszlást használták, egy nagyon kis részletet tartva meg az adatsorból, önszántukból megszabadulva sok egyéb információtól. A teljes adatsor ismeretében esetleg használható alsó becslést lehetne adni a monoton transzport távolságra — ám nem hisszük, hogy ez lényegesen a fent említett 8% fölé menne.

Végezetül fölhívnánk a figyelmet a Benford törvény most tapasztalt érdekes tulajdonságára: habár a törvény robosztus abból a szempontból, hogy majdnem mindig számíthatunk a megjelenésére, ha egy adatsor számai sok tényező szorzataként állnak elő, de bizonyos szempontból rendkívül sérülékeny is: az adatok mindegyikének egészen kis utólagos módosításával már tönkre lehet tenni.

A tanulságot három pontban foglalnánk össze:

1. A Benford törvény nem tűnik alkalmasnak kvantitatív becslésekre, ezen belül a túlárazás mértékének számszerű becslésére.

2. Egy olyan módszer esetén, ami kifinomult matematikai megfontolásokon alapul, hasznos lett volna matematikusokkal is konzultálni az eredmények bejelentése előtt.

3. Az adatsort, melyből a kutatás készült, talán érdemes lenne nyilvánosan hozzáférhetővé tenni. A magyar állam arra irányuló politikai szándéka egyértelműnek tűnik, hogy ezen adatokat lehetőleg senki ne tudja tanulmányozni: eltérő és nehezen kezelhető formátumokban, szétszórva lelhetőek fel az adatok. Aki adatbázist szeretne ebből építeni, annak magának kell kigyűjtenie őket. A szerzők és munkacsoportjuk hatalmas érdeme, hogy ezt az adatbázist létrehozták. Talán megfontolandó, nem kellene-e ezt nyilvánossá tenni, hogy az érdeklődők ne csak az ő szűrőjükön keresztül juthassanak átfogó információk birtokába.

Hivatkozások

[1] English Wikipedia: Benford's law. https://en.wikipedia.org/wiki/Benford's_law

[2] T. Tao: Benford's law, Zipf's law, and the Pareto distribution. https://terrytao.wordpress.com/2009/07/03/benfords-law-zipfs-law-and-the-pareto-distribution/

[3] Tóth I. J., Hajdu M.: Miért feltételezhető, hogy az EU által támogatott közbeszerzéseknél az 1,7 szerestől a közel 10 szeres túlárazás sem lehet ritka? http://www.crcb.eu/wp-content/uploads/2016/09/eu_overpricing_benfords_law_160907.pdf

[4] Tóth I. J.: Olyan országot építünk, ami nem tart sehová. http://index.hu/gazdasag/2016/04/15/toth_istvan_janos_interju_a_korrupcios_kockazatrol

[5] English Wikipedia: Wasserstein metric. https://en.wikipedia.org/wiki/Wasserstein_metric

Pete Gábor
MTA Rényi Alfréd Matematikai Kutatóintézet, Budapest, és
Budapesti Műszaki és Gazdaságtudományi Egyetem, Matematika Intézet
http://www.math.bme.hu/~gabor
gabor 'at' math dot' bme dot' hu

Timár Ádám
MTA Rényi Alfréd Matematikai Kutatóintézet, Budapest
madaramit 'at' gmail dot' com

A kutatót az EU Kutatási és Technológiafejlesztési Hetedik Keretprogram Marie Curie Intra-európai Ösztöndíja támogatta.