– Бенфордов закон –
Или како погодити прву цифру непознатог броја!?

др Раде Живаљевић, Београд

Драги Саша,

Кажеш да си за овај и наредни семестар изабрао међу осталима и курсеве STAT–400 и STAT–420 на одсеку за статистику.[†]

Мислим да си добро урадио и да ћеш се лепо забављати. У данашње време не постоје „чисте науке" па и овај курс не може бити изузетак, поготово што ти долазе као лепа допуна курсева из математике и физике које си већ одслушао. Теорија вероватноће, а са њом и математичка статистика, имају много лица и ту има много интересатног за свакога. Мени се посебно допада геометријска вероватноћа или интегрална геометрија са проблемима као што су „израчунавање" броја π бацањем игле, налажење вероватноће да космички зрак који прође кроз кабину космичког брода погоди неког члана посаде итд. У овој области се користе технике из диференцијалне геометрије (диференцијалне форме), Лиове групе, Харове мере итд.

У последње време још више уживам у везама са комбинаториком, статистичком физиком, топологијом итд. Има ту дакле много чега али за ову прилику ево једне причице која надам се прилично добро одражава један од комбинаторних аспеката теорије вероватноће.

Изабери било који град у Србији и запамти прву цифру броја који представља његов годишњи новчани бруто приход за рецимо 1985. годину. Колика је (емпиријска) вероватноћа да ће та цифра бити број 3? Да ли је већа, мања или једнака вероватноћи да ће та цифра бити број 5? Када кажемо „емпиријска вероватноћа", подразумевамо да ћеш насумице одабрати на пример 100 градова и затим установити колико често је та цифра управо број 3 (или 5). Изабери било који други природан број. Висину твог суседа у клупи, површину неке државе, број бактерија у капљици кише итд. Шта се може рећи о првој цифри (или прве две цифре) ових бројева?

Све ово можда звучи помало смешно и неинтересантно. Зашто би уопште постављали ова питања? Сагласно је са „здравим разумом" да ће насумице изабрани бројеви попут ових увек бити у истој мери заступљени. Нема ничег посебног што би разликовало цифре 3 или 5, па је за очекивати да ће се све цифре $1, 2, \ldots, 9$ појављивати са истом вероватноћом. Укратко, ако је узорак довољно велик, логично је да ће у просеку један од 9 бројева у нашем узорку почињати цифром 3.

Тако се и мени чинило а онда сам неочекивано наишао на чланак Владимира Игоревича Арнолда „Антинаучна револуција и математика"[‡]. Арнолд је један од мојих математичких хероја. Све што он напише читам са великом пажњом и уживањем, без обзира да ли је блиско

[†] Систем у коме се могу бирати курсеви са различитих факултета и комбиновати према жељама студената, у сагласности са неким генералним принципима прописаним од стране универзитета, типичан је за многе европске и америчке универзитете. Сасвим је извесно да ће и код нас будуће генерације студената имати прилику да активније утичу на своје образовање, у оквиру „Болоња" или неких других реформи универзитета.

[‡] „The antiscientifical revolution and mathematics",
`http://www.pdmi.ras.ru/~arnsem/Arnold/arn-papers.html`

мом тренутном научном интересу и истраживачком раду. И свакако нисам једини! Владимир Игоревич је један од најутицајнијих живих математичара. Још као студент московског универзитета овенчао се славом као решавач једног од чувених Хилбертових математичких проблема. Родоначелник је низа научник теорија и важи за водећег специјалисту из области механике, динамичих система, теорије сингуларитета (укључујући и теорију „катастрофа") и др. Наука, а посебно математика, је за Арнолда и његове следбенике јединствен „живи организам". Он не признаје парохијске поделе на анализу, геометрију, алгебру, топологију, комбинаторику итд. већ убедљиво демонстрира да је математика најлепша и најмоћнија када комбинује технике више различитих математичих дисциплина.

Чланак који ми је привукао пажњу проистекао је из једног од многих Арнолдових предавања[§] у којима он на жив, полемичан и математички врло инспиративан начин анализира различите проблеме везане за однос математике, науке и друштва уопште. Ево исечка са самог почетка овог чланка:

> „. . . Посматрајмо прву цифру броја који представља површину неке земље. Ова цифра може бити било који од бројева $1, 2, \ldots, 8, 9$. Расподелимо ли све земље света у групе према првим цифрама њихових површина установићемо да је оваква подела веома неравномерна. Земље код којих је прва цифра 1 чине чак 30% свих земаља света, док је број оних чија је цифра 9 око шест пута мањи. Такође, показује се што је ова цифра већа то је мањи број таквих земаља у свету. Мерне јединице у којима меримо ове површине не утичу на саму расподелу. Можемо их мерити у квадратним километрима, миљама или центиметрима. Ова неравномерна расподела првих цифара примећена је у мноштву других појава и позната је као „Бенфордов емпириски закон". На пример прве цифре популација земаља света се понашају слично. . . "

У данашње време кад наиђемо на нешто занимљиво и неочекивано, први импулс нам је да проверимо уз помоћ „Google"-а или неког сличног интернет претраживача о чему се ту ради и какав је стварни карактер те појаве. И заиста, на упит „Benford's distribution", помоли се невероватна историја. Оставићу теби да сам извршиш овај експеримент. Обавести ме ако ти је нешто посебно привукло пажњу!

Не могу да одолим а да не завршим ово писмо једноставним математичким проблемом који по мом мишљењу идеално објашњава елементарну природу Бенфордовог закона.

Задатак. Одредити прву цифру броја 2^{1000}.

Решење. Претпоставимо да је A цео број који има $m+1$ цифара и да је његова прва цифра a при чему $a \in \{1, 2, 3, 4, 5, 6, 7, 8, 9\}$. Другим речима претпостављамо да је

$$a \cdot 10^m \leqslant 2^{1000} < (a+1) \cdot 10^m. \tag{1}$$

Логаритмовањем обе стране једначине (1) (логаритам се узима са основом 10) и коришћењем чињенице да логаритми претварају множење у сабирање а степеновање у множење, добијамо

$$\log a + m \leqslant 1000 \log 2 < \log (a+1) + m. \tag{2}$$

[§] Предавање је одржано у оквиру циклуса „Changing concepts of nature at the turn of the millennium", Pontifical Academy at Vatican, October 26, 1998.

Неједнакост (2) је заиста веома лепа! Она нам говори да уколико желимо да одредимо логаритам од a, а тиме и непознату цифру a, довољно је да пронађемо разломљени део $\{x\}$ од $x = 1000 \log 2$. Сетимо се да сваки реалан број x има свој цео део $[x]$, дефинисан као највећи цео број m такав да је $m \leqslant x$, и одговарајући разломљени део, дефинисан као $\{x\} = x - [x]$. На пример,

$$[3,141592] = 3, \quad \{3,141592\} = 0,141592, \quad [-5,27] = -6, \quad \{-5,27\} = 0,73.$$

Пошто је $1 \leqslant a < a + 1 \leqslant 10$, знамо да је $0 \leqslant \log a < \log(a+1) \leqslant 1$ па из неједнакости (2) следи

$$\log a \leqslant \{1000 \log 2\} < \log(a+1). \tag{3}$$

Сада већ знамо шта треба да радимо. Поново користећи „Google“, проналазим „online“ научни калкулатор и израчунавам log 2. Тачност од 5 децимала би требало да је довољна али у нападу ентузијазма користим сву моћ калкулатора и користим максималну дату прецизност. Ово је забавно (што ме помало чуди пошто су ме у прошлости теоретска и филозофска разматрања увек више интересовала!) тако да одлучујем да пронађем не само $\log 2$ него и логаритме свих целих бројева од 2 до 9 (већ знам наравно да је $\log 1 = 0$ и $\log 10 = 1$).

$$\begin{array}{ll} \log 1 = 0 & \log 2 = 0,30102999566398 \\ \log 3 = 0,47712125471966 & \log 4 = 0,60205999132796 \\ \log 5 = 0,69897000433602 & \log 6 = 0,77815125038364 \\ \log 7 = 0,84509804001426 & \log 8 = 0,90308998699194 \\ \log 9 = 0,95424250943932 & \log 10 = 1. \end{array} \tag{4}$$

Одавде са лакоћом израчунавамо жељени разломљени део (са великом прецизношћу!)

$$\{1000 \cdot \log 2\} = \{1000 \cdot 0,30102999566398\} \doteq \{301,0299\} = 0,299.$$

Брзи поглед на табелу (4) открива да је

$$\log 1 \leqslant 0,299 < \log 2,$$

што у светлу неједнакости (3) гарантује да је $a = 1$! Дакле прва цифра броја 2^{1000} је 1, у пуној сагласности са предвиђањем Бенфордовог закона.

Анализа коришћена у решавању нашег задатка даје нам заправо много више информација о понашању првих цифара бројева облика 2^k. Да се разломљени део $\{x\} = \log 2^k$ нашао у неком другом интервалу, рецимо у интервалу $[\log 4, \log 5]$, закључили би да је прва цифра од 2^k број 4 итд. Одавде закључујемо да је права вероватноћа појављивања цифре a пропорционална дужини одговарајућег интервала $[\log a, \log(a+1))$! Пошто је дужина интервала $[\log 1, \log 2)$ приближно 6 пута већа од дужине интервала $[\log 9, \log 10)$ то нам коначно објашњава зашто се цифра 1 појављује 6 пута чешће од цифре 9.

Саша надам се да ти се допала ова причица о Бенфордовом закону. У наставку ћеш наћи неколико надам се интересатних задатака које можеш покушати да решиш применом горе изложених идеја или неким другим оригиналним приступом.

Срећно у новој школској години! Твој ујак Р.

Задаци за вежбу

1. Колико цифара имају бројеви 2^{1000} и 3^{10000}?

2. Одредити другу цифру броја 2^{1000}.

3. Шта је веће, $\log_2 3$ или $\log_3 4$?

4. Доказати да је $\log 2$ ирационалан број. Општије, за које природне бројеве a и b је $\log_a b$ рационалан број?

5. Шта каже „Google" када му затражимо информацију о разломљеном делу (fractional part). Изабрати неку од путања које више обећавају (нпр. Wolfram MathWorld) и „детективски" осмотрити са ким се „разломљени део" виђао у последње време, ко су познаници, у којим операцијама је учествовао итд.

6. Ако је α ирационалан број и k било који природан број, онда постоји разломак $\dfrac{m}{n}$ такав да је $n \leqslant k$ и

$$\left|\alpha - \frac{m}{n}\right| \leqslant \frac{1}{nk}.$$

7.* Претпоставимо да су α и β позитивни, ирационални бројеви везани релацијом

$$\frac{1}{\alpha} + \frac{1}{\beta} = 1. \tag{5}$$

Показати да се међу бројевима $[\alpha], [2\alpha], [3\alpha], \ldots, [\beta], [2\beta], [3\beta], \ldots$ налазе сви природни бројеви и то без понављања. Другим речима ако је $A_\theta = \{[n\theta] \mid n \in \mathbb{N}\}$, онда уз услов (5) важи

$$A_\alpha \cap A_\beta = \phi \qquad \text{и} \qquad A_\alpha \cup A_\beta = \mathbb{N}.$$

8.* Нека је α позитиван ирационалан број. Показати да је низ $\{\alpha\}, \{2\alpha\}, \ldots, \{n\alpha\}, \ldots$ равномерно распоређен у интевалу $[0,1]$ у смислу да је вероватноћа да се неки члан низа нађе у интервалу $(a,b) \subset [0,1]$ једнака дужини $b-a$ тог интервала.

Упутство. За велико n наћи добру апроксимацију $\frac{m}{n}$ броја α (Задатак 6) и показати да важи $\{j\alpha\} = \{\frac{jm}{n}\}$ за „скоро све" бројеве $j = 1, 2, \ldots, n$.

Друго решење базирано на „вишој" математици, може се наћи у познатој збирци задатака „Задаци и теореме из анализе" (G. Pólya, G. Szegö).

Статијата прв пат е објавена во списанието ТАНГЕНТА на ДМ на Србија во 2006/07 година