Koliko su veliki, veliki djelitelji?

Ivana Antoliš[1], *Bojan Basrak*[2]

Svaki prirodan broj veći od 1 na jedinstven način možemo rastaviti na proste faktore, tako je npr. $479\,451 = 79 \cdot 17^2 \cdot 7 \cdot 3$. Iako je rastav jednoznačan i potpuno određen, zanimljivo matematičko pitanje je odrediti ponašanje vodećih prostih faktora za "tipičan jako velik" prirodan broj. Da bismo matematički precizno postavili ovo pitanje koristit ćemo jezik vjerojatnosti.

Pretpostavimo da je broj $N = N_n$ slučajno izabran između brojeva $1, 2, \ldots, n$, odnosno tako da je $\mathbb{P}(N = k) = 1/n$ za sve $k = 1, 2, \ldots, n$. Rastav broja N je također jednoznačno određen, pa možemo pisati

$$N = P_1^{l_1} P_2^{l_2} \cdots P_k^{l_k}, \tag{1}$$

za neke prirodne brojeve $k, l_1, \ldots, l_k$, a pritom možemo pretpostaviti da prosti faktori zadovoljavaju $P_1 > P_2 > \ldots > P_k$. Uočite da su sad faktori $P_1, P_2, \ldots$ slučajni, a naše pitanje se, zapravo, tiče njihovog odnosa s N, i to za $n \to \infty$. U tom smislu ćemo pokušati odgovoriti na pitanje iz naslova. Nas će dakle zanimati vodeći, odnosno najveći, prosti djelitelji $P_1, P_2, \ldots$ slučajno odabranog broja. Ispostavlja se da je lakše razumjeti odnos prostih faktora $P_1, P_2, \ldots$ i slučajnog broja

$$T_n = P_1 P_2 \cdots P_k,$$

odnosno produkta prostih faktora broja N. Naime jasno je da vrijedi $T_n \leq N_n \leq n$, ali vrijedi i (za dokaz pogledajte [3]) $\log T_n / \log n \approx 1$ s velikom vjerojatnošću. Preciznije, za svaki $\varepsilon > 0$, $\mathbb{P}(1 \geq \log T_n / \log n > 1 - \varepsilon) \to 1$ za $n \to \infty$, stoga je naravno i $\log N_n \approx \log n$ u ovom istom smislu, gdje log označava *prirodan* logaritam. Ako odredimo ponašanje $\log P_i / \log T_n$ za velike n ono će nužno biti vrlo blizu ponašanja $\log P_i / \log N_n$, stoga uvodimo sljedeće slučajne veličine

$$X_1^n = \frac{\log P_1}{\log T_n}, \quad X_2^n = \frac{\log P_2}{\log T_n}, \quad \ldots, \quad X_k^n = \frac{\log P_k}{\log T_n}.$$

Ovi brojevi su jasno nenegativni, no važno je primijetiti da im je suma jednaka 1, naime

$$\sum_{i=1}^{k} \frac{\log P_i}{\log T_n} = \frac{\sum_{i=1}^{k} \log P_i}{\log T_n} = \frac{\log\left(\prod_{i=1}^{k} P_i\right)}{\log T_n} = 1.$$

Također, budući da smo proste faktore označili tako da su sortirani od većeg prema manjem vrijede nejednakosti

$$X_1^n > X_2^n > \ldots X_k^n.$$

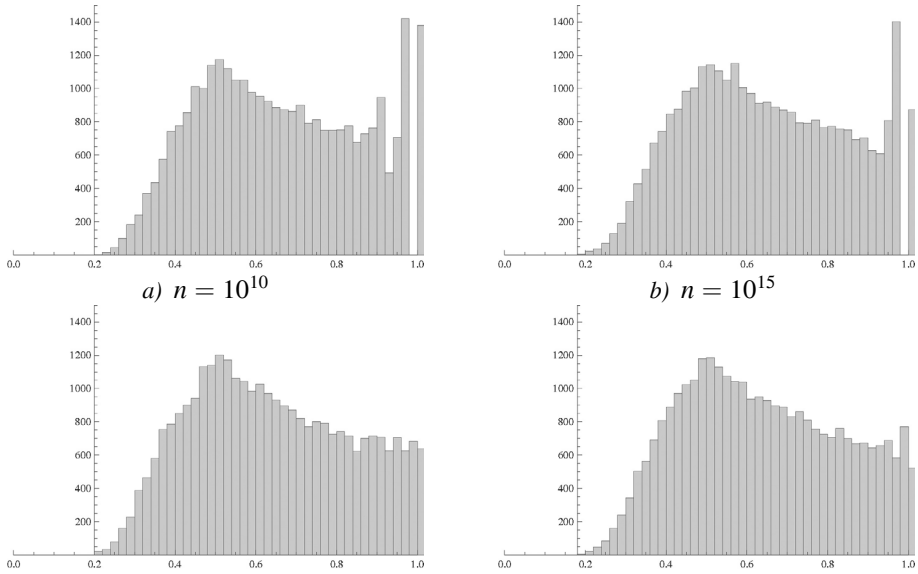
Premda je, jasno, niz $X_1^n, X_2^n, \ldots$ konačan, odnosno ima samo k elemenata, možemo ga formalno proširiti u beskonačan niz dodavanjem nula na njegov kraj, takav niz $X_1^n, X_2^n, \ldots, X_k^n, 0, 0, \ldots$ možemo smatrati slučajnim elementom skupa

$$\Delta := \{\mathbf{p} \in \mathbf{R}^\infty : p_1, p_2, \ldots \geq 0, p_1 + p_2 + \ldots = 1\}.$$

Nizove koji su elementi skupa Δ zovemo *razdiobama*, jer oni zaista svi određuju jednu vjerojatnosnu razdiobu npr. na skupu prirodnih brojeva. Tako da svaki naš izbor broja N određuje jednu takvu slučajno odabranu razdiobu.

[1] Autorica je asistentica na Matematičkom odsjeku PMF-a u Zagrebu; e-pošta: iantolis@math.hr

[2] Autor je izvanredni profesor na Matematičkom odsjeku PMF-a u Zagrebu; e-pošta: bbasrak@gmail.com



a) $n = 10^{10}$ *b)* $n = 10^{15}$

c) $n = 10^{20}$ *d)* $n = 10^{25}$

Slika 1. Vrijednosti $\dfrac{\log P_1}{\log T_n}$ *za različite* n.

Za ilustraciju možemo simulirati slučajnu varijablu $X_1^n = \dfrac{\log P_1}{\log T_n}$ za različite prirodne brojeve n i promotriti razdiobu, odnosno tzv. histograme njezinih realizacija. Na slici 1 prikazani su histogrami za slučajne uzorke od po $m = 30\,000$ elemenata dobivenih u slučaju $n = 10^{10}$, $n = 10^{15}$, $n = 10^{20}$ i $n = 10^{25}$ pomoću programskog paketa Wolfram Mathematica [4]. Uočimo da se na svim grafovima može uočiti rast otprilike do vrijednosti $1/2$ i pad nakon nje. Na grafovima se za $n < 10^{20}$ pojavljuju nepravilni skokovi pri kraju segmenta $[0,1]$. Razmislite kada se takva situacija događa. Uočimo da taj utjecaj postupno slabi kako n raste te da su grafovi za $n = 20$ i $n = 25$ već vrlo slični i relativno pravilni.

Slučajnu razdiobu možemo konstruirati i na pregršt drugih načina, jedan od prirodnijih nudi algoritam *lomljenja štapića*: izaberimo nezavisno i uniformno slučajne brojeve $U_1, U_2, \ldots$ na segmentu $[0,1]$. Postavimo

$$G_1 := U_1,$$

$$G_i := (1-U_1)(1-U_2)\ldots(1-U_{i-1})U_i, \text{ za } i = 2,3,\ldots.$$

Intuitivno, G_1 dobijemo nasumično birajući broj U_1 i razdjeljujući segment $[0,1]$, odnosno štapić na dijelove duljine U_1 i $1-U_1$. Nakon toga, nastavljamo ovaj postupak na drugom od ovih dvaju dijelova, i tako dobijemo slučajnu varijablu $G_2 = (1-U_1)\,U_2$ dijeljenjem segmenta $[U_1,1]$, odnosno ostatka dosad slomljenog štapića. Dakle, nakon što od jediničnog segmenta odlomimo slučajan dio, na ostatku napravimo isto, i postupak nastavimo.

Ispostavlja se da postoji uska veza između razdiobe niza $(X_i^n)_i$ (proširenog nulama) i razdiobe niza $(G_i)_i$. Već je na osnovi simulacije na slici 1 jasno da ne možemo očekivati da je npr. X_1^n blizu razdiobe od G_1 koja je uniformna po konstrukciji. Prava veza je bitno suptilnija.

Ključna ideja u nastavku je slučajno uređivanje elemenata skupa Δ procesom koji se zove *pristrano uzorkovanje* od engl. *biased sampling*. Pretpostavimo da je $(p_n) \in \Delta$, preslikavamo ga na slučajan način u drugi element skupa Δ koji označimo s (Q_n), tako sukcesivno uvodimo:

$$\begin{aligned}
Q_1 &= p_{k_1} \text{ s vjerojatnošću } p_{k_1}, k_1 \in \mathbf{N},\\
Q_2 &= p_{k_2} \text{ s vjerojatnošću } \frac{p_{k_2}}{1-Q_1} \text{ za } k_2 \in \mathbf{N} \setminus \{k_1\},\\
Q_3 &= p_{k_3} \text{ s vjerojatnošću } \frac{p_{k_3}}{1-Q_1-Q_2} \text{ za } k_3 \in \mathbf{N} \setminus \{k_1, k_2\},\\
&\vdots
\end{aligned}$$

Uočite da je niz (Q_n) permutacija niza (p_n) te da je on slučajan čak i ako to (p_n) nije bio.

Ako je npr. $(p_n) = \left(\frac{1}{2}, \frac{1}{3}, \frac{1}{6}, 0, 0, \ldots\right)$, možemo zamisliti da je u odnosu $\frac{1}{2} : \frac{1}{3} : \frac{1}{6}$ odnosno $3 : 2 : 1$ podijeljena tabla čokolade na dijelove A, B i C, te da slučajno određujemo koji ćemo dio prvo pojesti proporcionalno veličini dijelova, a zatim na preostalim dijelovima napravimo isto. Tako je npr.

$$\mathbb{P}(\text{prvi je na redu dio B}) = \frac{1}{3}.$$

Zatim biramo između preostalih dijelova na isti način. Tako je npr.

$$\mathbb{P}(\text{drugi je na redu dio A}|\ \text{prvi je na redu bio dio B}) = \frac{\frac{1}{2}}{1-\frac{1}{3}} = \frac{3}{4}.$$

Niz $Q_1, Q_2, Q_3, \ldots$ predstavlja ovako slučajno odabrane omjere dijelova čokolade. U ovom primjeru je jasno $Q_k = 0$ za $k \geq 4$ s vjerojatnošću 1, zbog oblika niza (p_n).

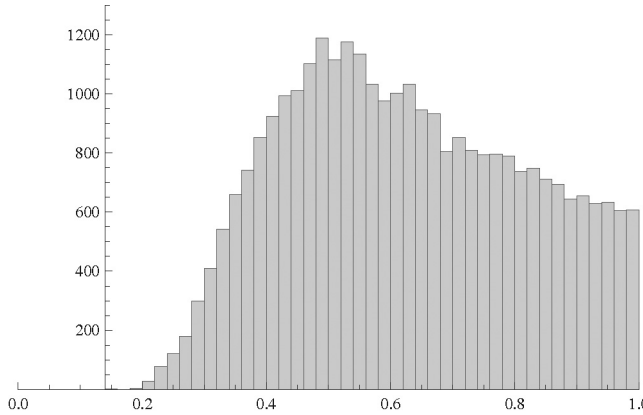
Pristrano uzorkovanje možemo (nezavisno) primijeniti na bilo koji niz u Δ, pa tako i na slučajni niz $(X_i^n)_i$, tako dobijemo novi niz

$$(X_i^n)_i \mapsto (Y_i^n).$$

Uočimo, da bismo dobili originalni niz $(X_i^n)_i$ od tako generiranog niza $(Y_i^n)_i$, dovoljno je ovaj drugi niz urediti po veličini počevši od najvećeg. Formalno se može definirati funkcija uređivanja po veličini $\rho : \Delta \to \Delta$ koja će sve nizove prostora Δ sortirati silazno. Uz takvu oznaku dakle imamo $(X_i^n)_i = \rho(Y_i^n)$.

Uočimo da je za definiciju funkcije ρ bitno da je na prostoru Δ suma $x_1 + x_2 + \ldots + x_k + \ldots = 1$. U suprotnom takva funkcija ne bi uvijek imala smisla. Razmislite ima li je smisla definirati npr. na nizu $0, 1, 0, 1, 0, 1, \ldots$.

Pogledajmo kako funkcija ρ djeluje na drugi niz o kojem smo govorili $(G_1, G_2, \ldots)$. U tom slučaju možemo zamisliti da i dalje lomimo štapić na nasumične dijelove, ali u ovom slučaju ih slažemo ne u poretku lomljenja nego od najvećeg silazno prema manjima. Kako bi dobili bolji osjećaj o tako dobivenom nizu, promotrimo njegovu prvu koordinatu odnosno najveći odlomljeni dio. Na slici 2 je prikazan histogram za slučajne uzorke od po $m = 30\,000$ elemenata dobivenih pomoću programskog paketa Wolfram Mathematica [4].



Slika 2. Vrijednosti najvećeg odlomljenog dijela štapića.

Ukoliko ovaj histogram usporedimo s onima na slici 1 za veće n možemo naslutiti ranije spomenutu suptilnu vezu između razdiobe niza $(X_i^n)_i$ (proširenog nulama) i razdiobe niza $(G_i)_i$.

Želimo li takve slutnje potvrditi važno nam je svojstvo neprekidnosti transformacija uređivanja po veličini ρ na skupu Δ. Da bismo objasnili što to znači, promotrimo nizove

$$x_1^1, x_2^1, x_3^1, x_4^1, \ldots$$
$$x_1^2, x_2^2, x_3^2, x_4^2, \ldots$$
$$x_1^3, x_2^3, x_3^3, x_4^3, \ldots$$
$$\vdots$$

na prostoru Δ takve da $x_1^n \to x_1$ kada $n \to \infty$, $x_2^n \to x_2$ kada $n \to \infty$ itd., gdje je niz $x_1, x_2, x_3, \ldots$ također iz prostora Δ. Činjenica da je funkcija ρ neprekidna znači da će isto vrijediti i nakon što na takve nizove djelujemo funkcijom ρ. Drugim riječima ako odgovarajuće koordinate niza iz Δ konvergiraju prema koordinatama niza koji je također u prostoru Δ onda će isto vrijediti i nakon što svaki od tih nizova sortiramo silazno. Zanimljivo je uočiti da je ovdje ponovno važan uvjet da je suma koordinata niza u Δ jednaka 1. Kad ne bi imali taj uvjet razmislite što bi vrijedilo u slučaju nizova

$$1, 0, 0, 0, 0, \ldots$$
$$0, 1, 0, 0, 0, \ldots$$
$$0, 0, 1, 0, 0, \ldots$$
$$\vdots$$

Kamo oni konvergiraju, a kamo nizovi koje dobivamo kada ih sortiramo?

Uz ovakve oznake vrijedi sljedeći teorem (za dokaz pogledajte [2])

Teorem 1.

$$(Y_1^n, Y_2^n, Y_3^n, \ldots) \implies (G_1, G_2, G_3, \ldots),$$

za $n \to \infty$, gdje $\implies$ označava konvergenciju po distribuciji. Drugim riječima, za svaki $k \in \mathbf{N}$ i proizvoljne $x_1, \ldots, x_k \in \mathbf{R}$ kada $n \to \infty$

$$\mathbb{P}(Y_1^n \le x_1, Y_2^n \le x_2, \ldots, Y_k^n \le x_k) \to \mathbb{P}(G_1 \le x_1, G_2 \le x_2, \ldots, G_k \le x_k).$$

Budući da je funkcija uređivanja neprekidna, vrijedi

$$(X_1^n, X_2^n, X_3^n, \ldots) = \rho\,(Y_1^n, Y_2^n, Y_3^n, \ldots) \implies \rho\,(G_1, G_2, \ldots)\,.$$

Odnosno, dobivamo ono što smo za prvu koordinatu i ranije naslutili, da niz $(\log P_1/\log T_n, \log P_2/\log T_n, \ldots)$ konvergira prema nizu $\rho\,(G_1, G_2, \ldots)$ koji je u literaturi poznata kao Poisson-Dirichletova razdioba.

Posebno ako označimo $G_M = \max\{G_1, G_2, G_3, \ldots,\}$ kada $n \to \infty$

$$\mathbb{P}\left(\frac{\log P_1}{\log T_n} \le x\right) \implies \mathbb{P}\,(G_M \le x)\,.$$

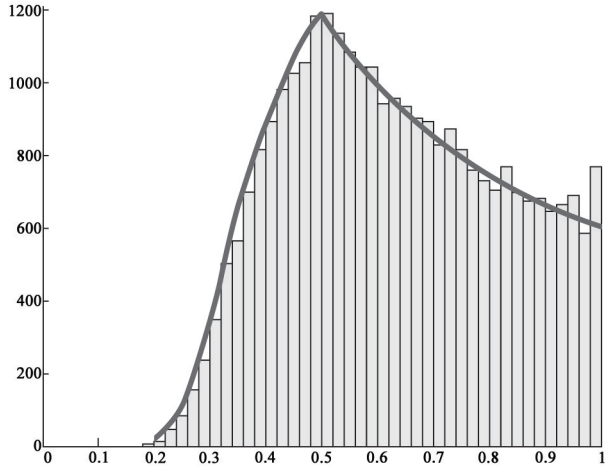
Može se posebno pokazati da je funkcija gustoće (za definiciju pogledajte [1]) slučajne varijable G_M dana donekle nezgrapnom formulom

$$f\,(x) = x^{-1} \sum_{0 \le k < x^{-1}-1} \frac{(-1)^k}{k!} \int\limits_{\substack{s_1,\ldots,s_k > 1 \\ \sum_{i=1}^{k} s_i < \frac{1-x}{x}}} \frac{ds_1 \ldots ds_k}{s_1 \cdot \ldots \cdot s_k}, \tag{2}$$

gdje se u slučaju $k = 0$ za vrijednost integrala uzima 1. Što znači da za $x \in [0,1]$

$$\mathbb{P}\left(\frac{\log P_1}{\log T_n} \le x\right) \implies \mathbb{P}\,(G_M \le x) = \int_0^x f\,(t)\,dt.$$

Za očekivati je da histogram vrijednosti $\log P_1/\log T_n$ slijede približno oblike ove gustoće za velike n, što se zaista i može vidjeti sa slike 3.



Slika 3. Usporedba grafa funkcije f i histograma vrijednosti $\log P_1/\log T_n$.

Formula u (2) je vrlo kompliciranа i prirodno se postavlja pitanje može li se pojednostaviti. To se zapravo svodi na računanje integrala u k varijabli. Pogledajmo kako to izgleda za neke x.

Ako je $x \geq \frac{1}{2}$ sumacija se vrši po prirodnim brojevima k koji su veći ili jednaki 0 i manji od $\frac{1}{x} - 1 \leq 2 - 1 = 1$. Budući da to zadovoljava samo $k = 0$ za takve x naša funkcija poprima oblik $f(x) = \frac{1}{x}$.

Uvjerite se da ako je $\frac{1}{3} \leq x < \frac{1}{2}$ pomoću jednostrukog integrala dobivamo funkciju oblika $f(x) = x^{-1}\left(1 - \log\left(\frac{1-x}{x}\right)\right)$.

Problemi se počinju javljati za male x, recimo možete provjeriti da bi za $x < \frac{1}{5}$ morali izračunati integrale s jednom, dvije, tri i četiri varijable što postaje problematično. Za određivanje tih vrijednosti mogu se koristiti metode koje integrale ne računaju eksplicitno već nam daju njihovu približnu vrijednost. Tako je za crtanje grafa na slici 3 korištena metoda integracije Monte Carlo koja ima vjerojatnosnu podlogu.

Ostaje vidjeti u kojem smislu nam ovo odgovara na pitanje koliko su zapravo veliki tako odabrani djelitelji slučajnog broja. Možemo se recimo za slučajno odabrani prirodan broj N_n pitati je li njegov najveći prosti faktor veći od $\sqrt{N_n}$. Naravno odgovor nije uvijek potvrdan, ali sada znamo procijeniti s kojom vjerojatnošću, za dovoljno velik n jest, naime

$$\begin{aligned}\mathbb{P}\left(P_1 > \sqrt{N_n}\right) &= \mathbb{P}\left(P_1 > N_n^{\frac{1}{2}}\right) = \mathbb{P}\left(\log P_1 > \frac{1}{2}\log N_n\right) \\ &= \mathbb{P}\left(\frac{\log P_1}{\log N_n} > \frac{1}{2}\right) \approx \mathbb{P}\left(\frac{\log P_1}{\log T_n} > \frac{1}{2}\right) \approx \mathbb{P}\left(G_M > \frac{1}{2}\right) \\ &= \int_{\frac{1}{2}}^{1} f(t)\,dt = \int_{\frac{1}{2}}^{1} \frac{dt}{t} = \log 1 - \log\frac{1}{2} \approx 0.693.\end{aligned}$$

Sličnim postupkom možete pokušati procijeniti vjerojatnost $\mathbb{P}\left(P_1 > \sqrt[3]{N_n}\right)$. U računu se javljaju nešto kompliciraniji integrali, koji se mogu odrediti i pomoću raznih programa prilagođenih matematičkim problemima.

Granična Poisson-Dirichletova razdioba iz teorema 1 se javlja i kod proučavanja najduljih ciklusa slučajno odabrane permutacije brojeva $1, 2, \ldots n$, u primijenjenoj statistici te u teorijskoj biologiji i genetici.

Literatura

[1] N. SARAPA, *Teorija vjerojatnosti*, Školska knjiga, 2002.

[2] P. BILLINGSLEY, *Convergence of Probability Measures*, A Wiley-Interscience Publication, 1999.

[3] A. DUJELLA, *Uvod u teoriju brojeva*, PMF – Matematički odjel, Sveučilište u Zagrebu, `http://web.math.pmf.unizg.hr/~duje/utb/utblink.pdf`

[4] *Wolfram Research*, Inc. *Mathematica*, verzija 9.0, Champaign, IL, 2012.