0. Tři věty o prvočíslech

Martin Mareš

Úvodem

Při analýze algoritmů se často využívají různá tvrzení o prvočíslech. Většina z nich byla poprvé dokázána v 19. století velikány analytické teorie čísel (Pafnutij Lvovič Čebyšev, Charles-Jean de la Vallée Poussin a další).

Zde ukážeme, že k získání o něco slabších, ale pro naše účely zcela postačujících výsledků lze dojít i elementárními kombinatorickými úvahami. Průkopníkem tohoto přístupu byl Paul Erdős – jeho první článek, vydaný za studií, se týkal právě kombinatorického důkazu Bertrandova postulátu [1]. Půjdeme v jeho stopách a také se trochu inspirujeme Galvinovým článkem [2] a Shoupovou znamenitou knihou o algoritmické teorii čísel [3].

Dokážeme následující tři tvrzení. Předem ještě dodejme, že v celém textu bude p vždy značit prvočíslo a log dvojkový logaritmus.

Bertrandův postulát: Pro každé $n \geq 1$ existuje alespoň jedno prvočíslo p takové, že $n < p \leq 2n$.

Tuto hypotézu vyslovil v roce 1845 Joseph Bertrand a o 5 let později ji dokázal Čebyšev. Informatikům se hodí například při konstrukci hešovacích tabulek: pokud chceme, aby velikost tabulky byla prvočíselná, vždy ji stačí zvětšit nejvýše dvakrát.

Věta o hustotě prvočísel: Označíme-li $\pi(x)$ počet prvočísel od 1 do x, platí $\pi(x) = \Theta(x/\log x)$.

Odhad hustoty prvočísel je užitečný například při generování velkých prvočísel pro šifrovací klíče RSA. Neznáme algoritmus, který by přímo generoval náhodná prvočísla, ale umíme rozhodnout, zda je číslo prvočíslem. Můžeme tedy volit náhodná čísla mezi n a $2n$ tak dlouho, než se strefíme do prvočísla, a hustotní věta nám zaručuje, že nám v průměru bude stačit $\Theta(\log n)$ pokusů, než nějaké najdeme.

Analytická teorie čísel nabízí přesnější odhad $\pi(x) \sim x/\ln x$ ($f \sim g$ znamená, že $f(x)/g(x)$ konverguje k 1 pro $x \to \infty$).

Prvočíselná harmonická řada: Pro libovolné n platí $\sum_{1 \leq p \leq n} 1/p = \mathcal{O}(\log \log n)$.

Tento součet vystupuje například v analýze Eratosthenova síta. Vyškrtáváním násobků prvočísla p strávíme čas $\mathcal{O}(n/p)$. Celkem tedy $\mathcal{O}(\sum_{1 \leq p \leq n} n/p) = \mathcal{O}(n \cdot \sum_{1 \leq p \leq n} 1/p) = \mathcal{O}(n \log \log n)$.

Pozoruhodné kombinační číslo

Naše následující úvahy se budou točit okolo různých pozoruhodných vlastností kombinačního čísla

$$K := \binom{2n}{n} = \frac{(2n)!}{(n!)^2} = \frac{2n \cdot (2n-1) \cdot (2n-2) \cdot \ldots \cdot (n+1)}{n \cdot (n-1) \cdot (n-2) \cdot \ldots \cdot 1},$$

kde n je parametr, jehož hodnotu zvolíme později. Jak je toto číslo velké?

Lemma A:

$$\frac{4^n}{2n+1} \le \binom{2n}{n} \le 4^n.$$

Důkaz: Z binomické věty (nebo z toho, že kombinační čísla počítají podmnožiny dané velikosti) víme, že

$$\binom{2n}{0} + \binom{2n}{1} + \ldots + \binom{2n}{2n} = 2^{2n} = 4^n.$$

Horní odhad je snadný: všichni sčítanci jsou nezáporní, takže žádný z nich nemůže být větší než součet. Pro dolní odhad si uvědomíme, že číslo $\binom{2n}{n}$ je největším ze sčítanců, takže musí být alespoň tak velké, jako jejich průměr, tedy $4^n/(2n+1)$. ♡

Lemma A':

$$\binom{2n+1}{n} \le 4^n.$$

Důkaz: Součet všech kombinačních čísel, která mají nahoře $2n+1$, činí $2^{2n+1} = 2 \cdot 4^n$. Hodnota $\binom{2n+1}{n}$ se ovšem v tomto součtu vyskytuje dvakrát – podruhé jako $\binom{2n+1}{n+1}$ díky symetrii kombinačních čísel. Takže nepřekročí polovinu součtu. ♡

Nyní se ukáže, jak naše kombinační číslo K souvisí s prvočísly:

Lemma D: K je dělitelné všemi prvočísly p takovými, že $n+1 \le p \le 2n$.
Důkaz: Vzpomeňme na zápis

$$K = \frac{2n \cdot (2n-1) \cdot \ldots \cdot (n+1)}{n \cdot (n-1) \cdot \ldots \cdot 1}.$$

Libovolné p mezi $n+1$ a $2n$ se v prvočíselném rozkladu čitatele vyskytuje právě jednou a v rozkladu jmenovatele ani jednou, takže musí dělit K. ♡

Na základě toho, že všechna prvočísla mezi $n+1$ a $2n$ dělí K, můžeme dokonce shora odhadnout, kolik takových prvočísel existuje. Označíme-li $\pi(a,b)$ počet prvočísel $p \in \{a, a+1, \ldots, b\}$, platí:

Důsledek D: $\pi(n+1, 2n) \le 2n/\log n$.
Důkaz: Každé takové prvočíslo dělí K, takže i jejich součin musí dělit K. Všech $\pi(n+1, 2n)$ činitelů tohoto součinu je nicméně větších než n, takže platí nerovnost $n^{\pi(n+1,2n)} \le K$. Proto $\pi(n+1, 2n) \le \log_n K = \log K / \log n$. Jelikož $K \le 4^n$, je $\log K \le 2n$ a jsme hotovi. ♡

Prvočíselná harmonická řada

Připomeňme nejprve součet obyčejné harmonické řady:

$$H(n) = \sum_{1 \le i \le n} \frac{1}{i}.$$

Tuto sumu lze snadno odhadnout integrálem z funkce $1/x$, zde ji ale sečteme jiným způsobem. Budeme předpokládat, že n je mocnina dvojky a řadu rozdělíme na úseky, jejichž hranice budou nižší mocniny dvojky:

$$H(n) = H(1,2) + \sum_{i=2}^{\log n} H(2^{i-1}+1, 2^i), \quad \text{kde} \quad H(a,b) = \sum_{a \le i \le b} \frac{1}{i}.$$

Nyní si všimneme, že $H(2^{i-1}+1, 2^i)$ je součtem 2^{i-1} členů, jejichž hodnoty leží mezi $1/2^i$ a $1/2^{i-1}$. Celý součet tedy leží mezi $1/2$ a 1. Dosadíme-li navíc $H(1,2) = 1 + 1/2 = 3/2$, dostaneme:

$$3/2 + 1/2 \cdot \log n \le H(n) \le 3/2 + \log n.$$

Z toho ihned plyne, že $H(n) = \Theta(\log n)$. Dodejme ještě, že kdyby n nebylo mocninou dvojky, můžeme $H(n)$ omezit zdola i shora hodnotami pro nejbližší nižší a vyšší mocninu dvojky, které se od sebe liší nejvýše konstanta-krát.

Sčítejme nyní stejným způsobem prvočíselnou harmonickou řadu:

$$P(n) = P(1,2) + \sum_{i=2}^{\log n} P(2^{i-1}+1, 2^i), \quad \text{kde} \quad P(a,b) = \sum_{a \le p \le b} \frac{1}{p}.$$

Zaměřme se na jeden úsek $P(2^{i-1}+1, 2^i)$. Sčítanci leží mezi $1/2^i$ a $1/2^{i-1}$ a podle Důsledku D jich je $\mathcal{O}(2^i/i)$. Součet úseku tedy můžeme odhadnout shora výrazem $\mathcal{O}(1/i)$.

To nyní společně s $P(1,2) = 1/2$ dosadíme do hlavní sumy pro $P(n)$:

$$P(n) = \frac{1}{2} + \sum_{i=2}^{\log n} \mathcal{O}(1/i).$$

Tento výraz můžeme odhadnout obyčejnou harmonickou řadou a dostaneme $P(n) = \mathcal{O}(H(\log n)) = \mathcal{O}(\log \log n)$. ♡

Intermezzo o exponentech

Pro důkaz zbylých dvou prvočíselných vět se nám budou hodit následující úvahy o exponentech prvočísel.

Označme $o_p(n)$ exponent prvočísla p v rozkladu čísla n:

$$n = \prod_p p^{o_p(n)}.$$

Snadno nahlédneme, že platí:

$$o_p(xy) = o_p(x) + o_p(y),$$
$$o_p(x/y) = o_p(x) - o_p(y).$$

Rovněž můžeme snadno spočítat, jaký je exponent daného prvočísla v rozkladu $n!$:

$$o_p(n!) = \lfloor n/p \rfloor + \lfloor n/p^2 \rfloor + \lfloor n/p^3 \rfloor + \dots .$$

Od 1 do n totiž leží $\lfloor n/p \rfloor$ čísel dělitelných p, z nichž $\lfloor n/p^2 \rfloor$ je navíc dělitelných p^2 a tak dále. Z toto snadno získáme následující lemma o našem oblíbeném kombinačním čísle K:

Lemma X: Pro libovolné prvočíslo p platí $p^{o_p(K)} \leq 2n$.
Důkaz: Z předchozí úvahy plyne $o_p(K) = o_p((2n)!/(n!)^2) = o_p((2n)!) - 2o_p(n!)$, a tedy

$$o_p(K) = \left(\sum_{i \geq 1} \left\lfloor \frac{2n}{p^i} \right\rfloor \right) - 2 \left(\sum_{i \geq 1} \left\lfloor \frac{n}{p^i} \right\rfloor \right) = \sum_{i \geq 1} \left(\left\lfloor \frac{2n}{p^i} \right\rfloor - 2 \left\lfloor \frac{n}{p^i} \right\rfloor \right).$$

Pro každé reálné číslo x přitom platí $\lfloor 2x \rfloor - 2\lfloor x \rfloor \leq 1$. Navíc pro $i > \log_p 2n$ sčítáme samé nuly. Proto $o_p(K) \leq \log_p 2n$, a tedy $p^{o_p(K)} \leq p^{\log_p 2n} \leq 2n$. ♡

Ještě pomocí kombinačních čísel dokážeme jedno pomocné tvrzení:

Lemma Y:

$$\text{Pro libovolné } n \text{ platí } \prod_{p \leq n} p \leq 4^n.$$

Důkaz: Indukcí podle n. Pro malá n tvrzení evidentně platí.

Pokud je $n > 2$ sudé, je indukční krok triviální, protože n není prvočíslo:

$$\prod_{p \leq n} p = \prod_{p \leq n-1} p \leq 4^{n-1} \leq 4^n.$$

Liché n zapíšeme jako $2m+1$ a součin rozdělíme na dvě části:

$$\prod_{p \leq 2m+1} p = \left(\prod_{p \leq m+1} p \right) \cdot \left(\prod_{m+2 \leq p \leq 2m+1} p \right).$$

Levá část podle indukčního předpokladu nepřesáhne 4^{m+1}. Pravá podle Lemmatu D dělí kombinační číslo $\binom{2m+1}{m}$, které už jsme shora odhadli číslem 4^m (viz Lemma A'). Celý součin tedy nepřekročí $4^{m+1} \cdot 4^m = 4^{2m+1} = 4^n$. ♡

Bertrandův postulát

K důkazu Bertrandova postulátu uvážíme prvočíselný rozklad kombinačního čísla K a zapíšeme ho jako součin $K = A \cdot B \cdot C \cdot D$, kde:

- A zahrnuje prvočísla $p \leq \sqrt{2n}$,
- B zahrnuje $\sqrt{2n} < p \leq 2n/3$,
- C zahrnuje $2n/3 < p \leq n$ a
- D zahrnuje $n < p \leq 2n$.

Ukážeme postupně, že části A, B a C jsou malé, takže aby se splnil dolní odhad $K \geq 4^n/(2n+1)$ z Lemmatu A, musí být D dostatečně velké. To speciálně znamená, že $D > 1$, takže musí existovat alespoň jedno prvočíslo mezi $n+1$ a $2n$.

Část A: Do části A přispívá nejvýše $\sqrt{2n}$ prvočísel, každé z nich podle Lemmatu X přispěje nejvýše $2n$. Celkem tedy $A \leq (2n)^{\sqrt{2n}}$.

Část B: Prvočísla z části B musí mít podle Lemmatu X exponent $o_p(K) \leq 1$. Přispějí tedy dohromady nejvýše svým součinem, který odhadneme pomocí Lemmatu Y: $B \leq 4^{2n/3}$.

Část C: Pro p z části C je $o_p(n!) = 1$ a $o_p((2n)!) = 2$. Proto $o_p(K) = o_p((2n)!) - 2o_p(n!) = 2 - 2 = 0$, takže žádné z prvočísel nepřispěje ničím a $C = 1$.

Část D: Kdyby bylo $D = 1$, dostali bychom složením předchozích odhadů nerovnost

$$\frac{4^n}{2n+1} \leq K \leq (2n)^{\sqrt{2n}} \cdot 4^{2n/3}.$$

Tu můžeme nadále upravovat:

$$\begin{aligned} 2^{2n-\log(2n+1)} &\leq 2^{(\log 2n)\sqrt{2n}+4n/3}, \\ 2n - \log(2n+1) &\leq (\log 2n)\sqrt{2n} + 4n/3, \\ 2/3 \cdot n &\leq (\log 2n)\sqrt{2n} + \log(2n+1). \end{aligned}$$

Levá strana přitom roste asymptoticky rychleji než pravá, takže nerovnost může platit pouze pro konečně mnoho n. Najít největší takové n není lehké, ale snadno ověříme, že pro $n \geq 1024$ již bezpečně neplatí: $\log 2n \leq \sqrt{n/8}$ (tyto dvě funkce mají právě jeden průsečík a $n = 1024$ leží bezpečně za ním), takže $(\log 2n)\sqrt{2n} \leq \sqrt{n/8} \cdot \sqrt{2n} = n/2$. Analogicky získáme $\log(2n+1) \leq n/10$. Pravá strana nerovnosti tedy nepřeroste $n/2 + n/10 = 6/10 \cdot n$, což je méně než $2/3 \cdot n$ na levé straně.

Pro $n \geq 1024$ musí tedy být $D > 1$, takže Bertrandův postulát platí. Pro zbývajících konečně mnoho n stačí uvážit posloupnost prvočísel

$$2, 3, 5, 7, 13, 23, 43, 83, 163, 317, 631, 1259.$$

V ní je každé prvočíslo menší než dvojnásobek předchozího, takže každý interval $\langle n, 2n\rangle$ pro $n < 1024$ obsahuje alespoň jedno z nich.

Tím je Bertrandův postulát dokázán. ♡

Hustota prvočísel

Nakonec dokážeme větu o hustotě prvočísel, tedy $\pi(x) = \Theta(x/\log x)$. Důkaz rozdělíme na horní a dolní odhad.

Dolní odhad: Logaritmus kombinačního čísla K můžeme pomocí prvočíselného rozkladu zapsat takto:

$$\log K = \log\left(\prod_{p \leq 2n} p^{o_p(K)}\right) = \sum_{p \leq 2n} \log p^{o_p(K)} = \sum_{p \leq 2n} o_p(K) \cdot \log p.$$

Jelikož podle Lemmatu X přispěje prvočíslo p nejvýše $2n$, musí být jeho exponent $o_p(K)$ nejvýše $\log_p 2n = \log 2n / \log p$. Dosadíme:

$$\log K \le \sum_{p \le 2n} \frac{\log 2n}{\log p} \cdot \log p \le \sum_{p \le 2n} \log 2n \le \pi(2n) \cdot \log 2n.$$

Z úvodních odhadů velikosti čísla K víme, že $K \ge 4^n/(2n+1)$, takže pro dost velká n nastane $\log K \ge n$. Z toho ihned dostaneme $\pi(2n) \ge n/\log(2n)$.

Pro sudé x je tedy $\pi(x) = \Omega(x/\log x)$ a jelikož funkce π neklesá, musí tento odhad platit i pro lichá x.

Horní odhad: Zvolme k tak, aby $2^{k-1} < x \le 2^k$. Odhadneme shora počet všech prvočísel mezi 1 a 2^k. Použijeme k tomu osvědčený trik s rozdělením řady na bloky podle mocnin dvojky a nerovnost $\pi(x+1, 2x) \le 2x/\log x$ z Důsledku D:

$$\pi(x) \le \pi(2^k) = \pi(1,2) + \sum_{i=2}^{k} \pi(2^{i-1}+1, 2^i) \le 1 + \sum_{i=2}^{k} \frac{2 \cdot 2^i}{i}.$$

Poslední sčítanec sumy vpravo (pro $i = k$) dává kýžený odhad řádu $x/\log x$, ale musíme se ubezpečit, že zbývajících $\log n$ členů klesá dost rychle na to, aby nám to nepokazily. Naštěstí ano:

$$\frac{2^{k-1}}{k-1} \Big/ \frac{2^k}{k} = \frac{k}{2(k-1)} \le \frac{3}{4} \qquad (\text{pro } k \ge 3),$$

takže celou sumu můžeme shora odhadnout geometrickou řadou:

$$\pi(x) \le \pi(2) + \pi(2^{k-1}+1, 2^k) \cdot \sum_{i=0}^{\infty} \left(\frac{3}{4}\right)^i \le 1 + \frac{2 \cdot 2^k}{k} \cdot \frac{4}{3} = \mathcal{O}(2^k/k) = \mathcal{O}(x/\log x).$$

Tím je věta o hustotě prvočísel dokázána. ♡

Literatura

[1] P. Erdős. Beweis eines Satzes von Tschebyschef. *Acta Sci. Math. (Szeged)*, 5(1930–1932):194–198, 1932.

[2] D. Galvin. Erdős's Proof of Bertrand's Postulate, 2006. Available online at `http://nd.edu/~dgalvin1/pdf/bertrand.pdf`. Cit. 2012-08-21.

[3] V. Shoup. *A* computational introduction to number theory and algebra. Cambridge University Press, 2009. Available at `http://www.shoup.net/ntb/`.