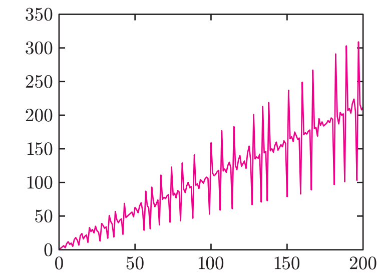
Ciąg EKG, czyli zaskakująca zabawa z teorią liczb

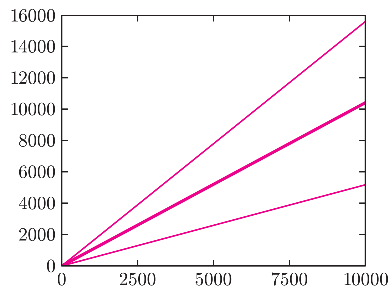
Marcin PILIPCZUK[*]

Teoria liczb w wielu miejscach jest zaskakująca i nieprzewidywalna, co – na poziomie elementarnym – objawia się przede wszystkim przez nieregularne i trudne do opisania rozmieszczenie liczb pierwszych w zbiorze liczb naturalnych. Wraz z Piotrem Hofmanem zmierzyliśmy się z tym fenomenem, badając tzw. ciąg EKG: prosty do zdefiniowania, a z bardzo ciekawymi właściwościami.

Ciąg EKG a_n określamy następująco: $a_1 = 1$, $a_2 = 2$ i dalej a_{n+1} jest najmniejszą liczbą całkowitą dodatnią niewystępującą wcześniej w ciągu, taką że $\mathrm{NWD}(a_n, a_{n+1}) > 1$. Jako pierwszy zajmował się tym ciągiem Ayres, następnie badali go Lagarias, Rains i Sloane.



Rys. 1. Pierwsze 200 wyrazów ciągu EKG.



Rys. 2. Pierwsze 10 000 wyrazów ciągu EKG.

Skąd nazwa ciąg EKG? Pierwsze 40 wyrazów ciągu nam może tego nie powie:

1	2	4	6	3	9	12	8	10	5
15	18	14	7	21	24	16	20	22	11
33	27	30	25	35	28	26	13	39	36
32	34	17	51	42	38	19	57	45	40

ale jeśli narysujemy wykres pierwszych 200 wyrazów (rys. 1), to zobaczymy coś ciekawego. Nasuwa się pytanie: czemu zawdzięczamy te „skoki" na wykresie, przypominające zachowanie elektrokardiogramu? Prześledzenie wyrazów ciągu EKG pokazuje, że te skoki to trójki kolejnych wyrazów postaci $2p, p, 3p$ dla liczb pierwszych p.

Lagarias, Rains i Sloane wykazali, między innymi, że ciąg a_n jest permutacją liczb całkowitych dodatnich, że liczby pierwsze pojawiają się w nim w kolejności rosnącej i że asymptotycznie zachowuje się on liniowo (dokładniej, $n/260 \leqslant a_n \leqslant 14n$ dla każdego n). Rysunek 2, pokazujący pierwszych 10 000 wyrazów ciągu EKG, sugeruje, że powinniśmy potrafić udowodnić dużo dokładniejsze ograniczenia: poza skokami, a_n jest równe prawie dokładnie n.

Oznaczmy $a'_n = 2p$, jeśli $a_n = p$ lub $a_n = 3p$ dla liczby pierwszej p, i $a'_n = a_n$ w przeciwnym przypadku. Ta definicja „łagodzi" skoki. Spodziewamy się więc, że będzie $a'_n \sim n$, tj. $\lim_{n\to\infty} a'_n/n = 1$. Bazując na eksperymentach numerycznych, Lagarias, Rains i Sloane postawili następujące hipotezy:

1. w ciągu EKG każda liczba pierwsza $p > 2$ występuje w trójce $2p, p, 3p$;
2. $\lim_{n\to\infty} a'_n/n = 1$, a dokładniej $a'_n \sim n(1 + 1/(3\log n))$.

Wraz z Piotrem Hofmanem udowodniliśmy pierwszą hipotezę i częściowo drugą: wykazaliśmy, że istnieje taka stała uniwersalna C, że

$$\left|\frac{a'_n}{n} - 1\right| \leqslant \frac{C}{\log\log\log n},$$

co oczywiście implikuje $\lim_{n\to\infty} a'_n/n = 1$. Dowód powyższego oszacowania jest żmudny i wymaga wielu drobnych technicznych kroków. Za to dowód pierwszej hipotezy jest dość krótki i przedstawię go w dalszej części artykułu.

Jeśli q jest liczbą pierwszą dzielącą zarówno a_n, jak i a_{n-1}, to q nazywamy *liczbą sterującą* wyrazu a_n (to nie jest definicja jednoznaczna, dla jednego a_n możemy mieć wiele liczb sterujących). Zauważmy, że a_n jest najmniejszą liczbą niewystępującą wcześniej w ciągu, a podzielną przez q.

Kluczowe jest następujące spostrzeżenie. Ustalmy liczbę rzeczywistą $x \geqslant 0$. Wówczas, dla każdej liczby pierwszej q istnieje co najwyżej jeden taki wyraz ciągu a_n, że $a_{n-1} < x \leqslant a_n$ i q jest liczbą sterującą a_n. Istotnie, zauważmy, iż jeśli q jest liczbą sterującą a_n i $x \leqslant a_n$, to wszystkie liczby mniejsze od x podzielne przez q wystąpiły wcześniej w ciągu. Czyli jeśli dla pewnego $n' > n$ mamy $a_{n'-1} < x$, to q nie dzieli $a_{n'-1}$, a więc nie może być liczbą sterującą $a_{n'}$.

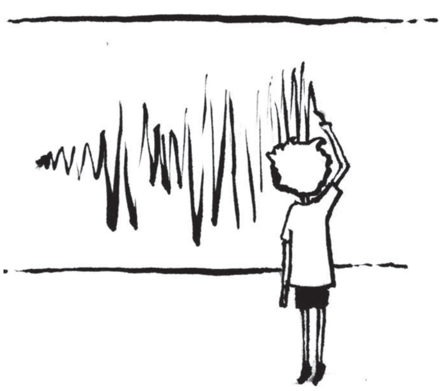
Załóżmy, że pewna liczba pierwsza p pojawia się w ciągu niepoprzedzona wyrazem $2p$, czyli $a_{n+1} = p$, $a_n = mp$ dla pewnych całkowitych $n > 1$ i $m > 2$. Lagarias, Rains i Sloane udowodnili, że wówczas a_n jest pierwszym wyrazem ciągu podzielnym przez p. W szczególności oznacza to, że $2p$ nie wystąpiło w ciągu przed pozycją n.

[*] doktorant, Instytut Informatyki, Uniwersytet Warszawski

Rozpatrzmy zbiory

$$A = \{1 < i \leqslant n : a_{i-1} < 2p \leqslant a_i\} \quad \text{i} \quad B = \{1 \leqslant i < n : a_i \geqslant 2p > a_{i+1}\}.$$

Zauważmy, że $|A| - |B| = 1$, gdyż $a_1 < 2p$ i $a_n > 2p$. Doprowadzimy do sprzeczności, dowodząc, że jeśli $m > 2$ i p jest wystarczająco duże, to zbiór B jest dużo większy niż A.



Jak wykazali Lagarias, Rains i Sloane, liczby pierwsze pojawiają się w ciągu EKG w kolejności rosnącej, czyli przed pozycją n nie pojawiła się liczba pierwsza większa niż p. Jeśli q jest liczbą sterującą wyrazu $a_{n'}$ i $n' \leqslant n$, to q jest kandydatem na wyraz $a_{n'}$, więc pojawia się w ciągu na pozycji n' lub wcześniejszej. Wnioskujemy stąd, że wszystkie liczby sterujące do pozycji n są mniejsze niż p. Z podanego przed chwilą „kluczowego spostrzeżenia” wiemy, że dla każdej liczby pierwszej $q < p$ istnieje co najwyżej jeden taki indeks i, że $a_{i-1} < 2p \leqslant a_i$ i q jest liczbą sterującą a_i. Liczb pierwszych nie większych niż p jest nie więcej niż $cp/\log p$ dla pewnej stałej uniwersalnej c, czyli

$$|A| \leqslant cp/\log p.$$

Spójrzmy teraz na zbiór B. Niech q będzie liczbą sterującą wyrazu $a_n = mp$; oczywiście q jest dzielnikiem m. Wobec tego liczby kq dla $1 \leqslant k \leqslant (mp/q - 1)$ musiały pojawić się w ciągu wcześniej. Oznaczmy

$$B' = \{1 \leqslant i < n : 2q \mid a_i \text{ oraz } a_i > 2p\}.$$

Indeksy wszystkich parzystych wyrazów postaci kq (podzielnych przez 4, jeśli $q = 2$), większych od $2p$, należą do B', wobec tego

$$|B'| \geqslant \left\lfloor \frac{mp - 2p}{2q} \right\rfloor \geqslant \left\lfloor \frac{p}{6} \right\rfloor.$$

Z drugiej strony, jeśli $i \in B'$, to $2p$ jest potencjalnym kandydatem na wyraz a_{i+1}, czyli $a_{i+1} < 2p$. Mamy więc $B' \subseteq B$, zatem

$$|B| \geqslant \lfloor p/6 \rfloor.$$

Wobec tego dla odpowiednio dużych p mamy $|B| > |A|$.

Wyliczywszy dokładnie stałe w powyższym rozumowaniu, można dowieść, że $|B| > |A|$ już dla $p > 25\,000$. Teza dla mniejszych wartości p została sprawdzona numerycznie przez Lagariasa, Rainsa i Sloane'a, więc hipoteza została udowodniona dla wszystkich $p > 2$. Dowód zakończony.

Jak wyznaczać wyrazy ciągu EKG?

Jakub RADOSZEWSKI

Po przeczytaniu artykułu Marcina Pilipczuka trudno nie odnieść wrażenia, że nasz zasób wiedzy o zachowaniu ciągu EKG opiera się przede wszystkim na wynikach eksperymentów komputerowych, natomiast dowody otrzymanych w ten sposób hipotez pojawiają się z pewnym opóźnieniem. Co więcej, niektóre własności nie doczekały się jeszcze ścisłych dowodów, choć dane doświadczalne pokazują je na tyle wyraźnie, że ciężko byłoby nie wierzyć w ich prawdziwość. Widzimy tu, że informatyka zaczyna pełnić funkcję eksperymentalnej poddziedziny matematyki. Co ciekawe jednak, nie chodzi tu tylko o łatwiejsze stawianie hipotez – wyniki obliczeń mogą stanowić fragmenty *dowodu* odkrytych za ich pomocą własności! W przypadku ciągu EKG przejawiło się to w automatycznym sprawdzeniu prawdziwości hipotezy dla pewnej liczby początkowych wyrazów ciągu (dla pozostałych zadziałało rozumowanie czysto matematyczne), ale zdarzały się już w historii problemy, dla których cały dowód opierał się na komputerowej analizie przypadków – na przykład dowód faktu, że każdą mapę (graf planarny) można pokolorować czterema kolorami, tak aby żadne dwa sąsiednie państwa (wierzchołki) nie miały tego samego koloru. Niektórzy twierdzą, że takie rozumowanie, oparte w znacznej mierze na programach komputerowych i często obszernych wynikach ich działania, nie jest właściwie żadnym *dowodem*. Jest to argument poniekąd słuszny – ciężko ręcznie zweryfikować poprawność tego typu wywodu. Ten medal ma jednak dwie strony. Otóż przypomnijmy sobie chociażby liczne „dowody” Wielkiego Twierdzenia Fermata, publikowane bez mała przez trzy stulecia, które nie używały żadnych podejrzanych metod informatycznych, a mimo wszystko okazywały się nieprawdziwe (niepełne). Czy autorzy tych „dowodów”