

Rys. 1. Przykład SAW-a o długości $k = 15$.

*Instytut Matematyki Stosowanej i Mechaniki, Uniwersytet Warszawski; Wydział Matematyki i Informatyki, Uniwersytet Mikołaja Kopernika, Toruń

w przyrodzie, wariant najlżejszy przestaje być najbardziej prawdopodobny – dominuje natomiast wariant z 42 dodatkowymi neutronami. Tak więc dzięki obliczeniu teoretycznego rozkładu izotopowego łatwo zorientować się, że osoba analizująca BSA spektrometrem mas powinna oczekiwać, że aparat zmierzy substancję cięższą o około 42 Da od masy najlżejszego wariantu tej cząsteczki.

Często istnieje wiele cząsteczek o zadanej masie, każda dodatkowa informacja może więc być przydatna w poprawnej identyfikacji. Taką informacją jest, na przykład, właśnie rozkład izotopowy. Rozważmy dwie cząsteczki: propan (C_3H_8) oraz dwutlenek węgla (CO_2). Mają one taką samą masę wariantu najlżejszego (44 Da), w obu przypadkach też najwyższe jest prawdopodobieństwo tego wariantu. Gdy obliczymy stosunek prawdopodobieństw $\frac{q_1}{q_0}$ dla każdego z nich, okaże się, że choć obie wartości są bardzo małe, to jednak w przypadku pierwszej cząsteczki jest ona około trzy razy większa niż w przypadku drugiej (dokładne obliczenia pozostawiamy Czytelnikowi).

Przy przewidywaniu tego, jak będzie wyglądał dla danej cząsteczki wynik pomiaru w eksperymencie wykorzystującym spektrometr mas, pomocne może okazać się więc obliczenie jej rozkładu izotopowego. Problem ten sprowadziliśmy zaś do czysto matematycznego zagadnienia potęgowania wielomianów (por. równanie (2)). To już jednak temat na inną opowieść...

Monte Carlo, spacer i polimery

Wojciech NIEMIRO*

Z czym kojarzy się Monte Carlo? Z kasynami, hazardem, ruletką. A więc Monte Carlo jest symbolem potęgi *przypadku*, który jednym pozwala zbijać fortuny, innych rujnuje. Są jednak ludzie, którzy przypadkowość, losowość potrafią okiełznać i wykorzystać z pożytkiem (przynajmniej dla siebie). Oczywiście, takimi ludźmi są właściciele kasyn gry, ale nie tylko. Również matematycy zaprzęgają losowość do pożytecznej roboty. Metody obliczeniowe oparte na generowaniu (symulowaniu komputerowym) zdarzeń losowych nazywają się... oczywiście, metodami Monte Carlo.

Liczenie przedmiotów nie wydaje się zadaniem wymagającym wyrafinowanej matematyki, wykraczającej ponad program przedszkolny. Nie widać też, jaką rolę w czynności zliczania miałby odgrywać przypadek. A jednak! Rozważymy pozornie proste zadanie zliczania dróg, w którym ujawnią się nieoczekiwane trudności i pokażemy, jak pomysłowe metody Monte Carlo wynaleziono, aby te trudności pokonać.

Wyobraźmy sobie, że idziemy na spacer. Załóżmy, że mieszkamy w mieście, w którym ulice tworzą regularną, kwadratową siatkę, tak jak na rysunku 1. Ruszamy z punktu początkowego, wybierając jedną z czterech ulic. Idziemy zatem na północ, na zachód, na południe lub na wschód. Przechodzimy do sąsiedniego skrzyżowania i znowu wybieramy jeden z 4 kierunków. Powtarzamy całą procedurę, powiedzmy, k razy. Przebyta droga jest łamaną składającą się z k odcinków o jednostkowej długości. Na rysunku 1 mamy przykład drogi o długości $k = 15$. Używając układu współrzędnych „zaczepionego w punkcie początkowym”, możemy tę drogę zapisać jako ciąg kolejno odwiedzanych punktów:

$$(0, 0) \rightarrow (0, 1) \rightarrow (1, 1) \rightarrow (1, 2) \rightarrow (0, 2) \rightarrow \dots \rightarrow (2, 0) \rightarrow (2, -1).$$

Inny sposób zakodowania tej samej drogi to podanie kierunków ruchu w kolejnych „krokach”: $N, E, N, W, W, \dots, E, S$. Ile jest wszystkich dróg o długości k ? Odpowiedź jest łatwa: 4^k . Musimy tu wyjaśnić, że licząc drogi wychodzące z tego samego punktu, utożsamiamy drogi z ciągami literek z alfabetu $\{N, W, S, E\}$.

k	Z_k
1	4
2	12
3	36
4	100

Odpowiedź można znaleźć w dalszej części tego artykułu.

Zmieńmy teraz nasze „reguły spacerowania”. Przypuśćmy, że nie chcemy powtórnie przechodzić przez punkt, który już wcześniej odwiedziliśmy. W ciągu opisującym drogę nie może dwukrotnie pojawić się ten sam punkt. Tę własność ma np. droga na rysunku 1. Po angielsku taka droga nazywa się *Self Avoiding Walk*, w skrócie SAW. Ile jest SAW-ów o długości k ? Wystarczy je wszystkie narysować i policzyć, prawda? Tą metodą można zbudować widoczną obok tabelkę. Liczbę SAW-ów o długości k oznaczmy przez Z_k . Ale co dalej? Im dalej, tym gorzej. Ile jest SAW-ów o długości $k = 50$? W istocie, liczenie SAW-ów jest na tyle trudnym zadaniem, że matematycy (i fizycy, o czym dalej) uciekają się do przybliżonego zliczania metodą Monte Carlo. Najprostszy algorytm tego rodzaju jest niemal oczywisty. Ustalmy k . Oznaczmy przez \mathcal{X} zbiór wszystkich dróg o długości k , zaś przez $\mathcal{Z} \subset \mathcal{X}$ zbiór SAW-ów o długości k . Oczywiście $|\mathcal{Z}| = Z_k$ i $|\mathcal{X}| = 4^k$ (kreseczki $| \cdot |$ oznaczają liczbę elementów zbioru; dla uproszczenia przy symbolach zbiorów zrezygnowaliśmy z indeksu k). Jeśli wybierzemy losowo jedną z dróg ze zbioru \mathcal{X} , to prawdopodobieństwo otrzymania „przypadkiem” SAW-a jest równe $|\mathcal{Z}|/|\mathcal{X}|$. Jeśli wybierzemy losowo i niezależnie n dróg, to spośród nich znajdzie się pewna liczba, powiedzmy n_1 , SAW-ów. Widać, że dla dużych n powinniśmy mieć

$$(1) \quad \frac{n_1}{n} \approx P(\mathcal{Z}) = \frac{|\mathcal{Z}|}{|\mathcal{X}|}.$$

Przybliżenie staje się coraz lepsze w miarę zwiększania n . Ścisłe sformułowanie tego stwierdzenia nazywa się Prawem Wielkich Liczb (PWL) i jest podstawą rachunku prawdopodobieństwa. Nie trzeba jednak wielkiej wiedzy z tej dziedziny, wystarczy trochę zdrowego rozsądku, aby w przybliżoną równość (1) uwierzyć i ją zrozumieć. Z naszego punktu widzenia ważne jest wykorzystanie PWL do oszacowania nieznaney liczby SAW-ów. Wiemy, że $|\mathcal{X}| = 4^k$. Zatem

- $\frac{n_1}{n} 4^k \approx |\mathcal{Z}|$ jest oszacowaniem liczby Z_k prostą metodą Monte Carlo.

W zasadzie możemy osiągnąć dowolną dokładność dla odpowiednio dużych n . Zauważmy, że bardzo łatwo wylosować drogę ze zbioru \mathcal{X} w taki sposób, aby dla każdego $x \in \mathcal{X}$ prawdopodobieństwo wybrania x było jednakowe, równe $P(x) = 1/4^k$. Na każdym kroku wybieramy jeden z 4 kierunków losowo. Możemy dwa razy rzucić monetą i umówić się: jeśli otrzymamy OO, to N , jeśli OR, to W , jeśli RO, to S , jeśli RR, to E . W ten sposób generujemy ciąg kodujący drogę „błądzenia przypadkowego”. Komputer pozwala szybko „rzucić monetą” np. $(2k) \cdot 10000$ razy i wygenerować $n = 10000$ takich dróg.

Niestety, prosta metoda Monte Carlo jest bardzo nieefektywna, bo dla dużych k prawdopodobieństwo $P(\mathcal{Z}) = |\mathcal{Z}|/|\mathcal{X}|$ szybko zbliża się do zera. Oczekiwanie na przypadkowe trafienie w zbiór \mathcal{Z} , czyli wylosowanie SAW-a, zaczyna przypominać poszukiwanie igły w stogu siana. Możliwym rozwiązaniem tego problemu jest metoda „wzrostu” zaproponowana przez Ariannę i Marshalla Rosenbluthów polega na losowaniu kolejnych kroków błądzenia przypadkowego tylko spośród „dopuszczalnych punktów”, to znaczy punktów wcześniej nieodwiedzonych. W każdym kroku, z wyjątkiem pierwszego, mamy co najwyżej 3 możliwości. Rysunek 2 (następna strona) pokazuje kolejne kroki prowadzące do zbudowania SAW-a z rysunku 1. Widać, że kolejne kroki wybieraliśmy spośród

4, 3, 3, 3, 2, 3, 2, 2, 3, 2, 3, 3, 2, 1, 3

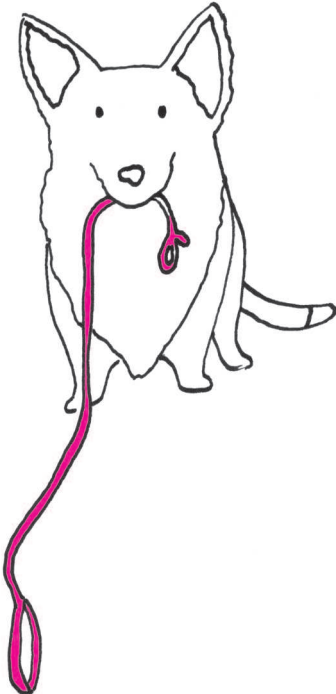
możliwych. Nasz SAW został zatem wylosowany z prawdopodobieństwem

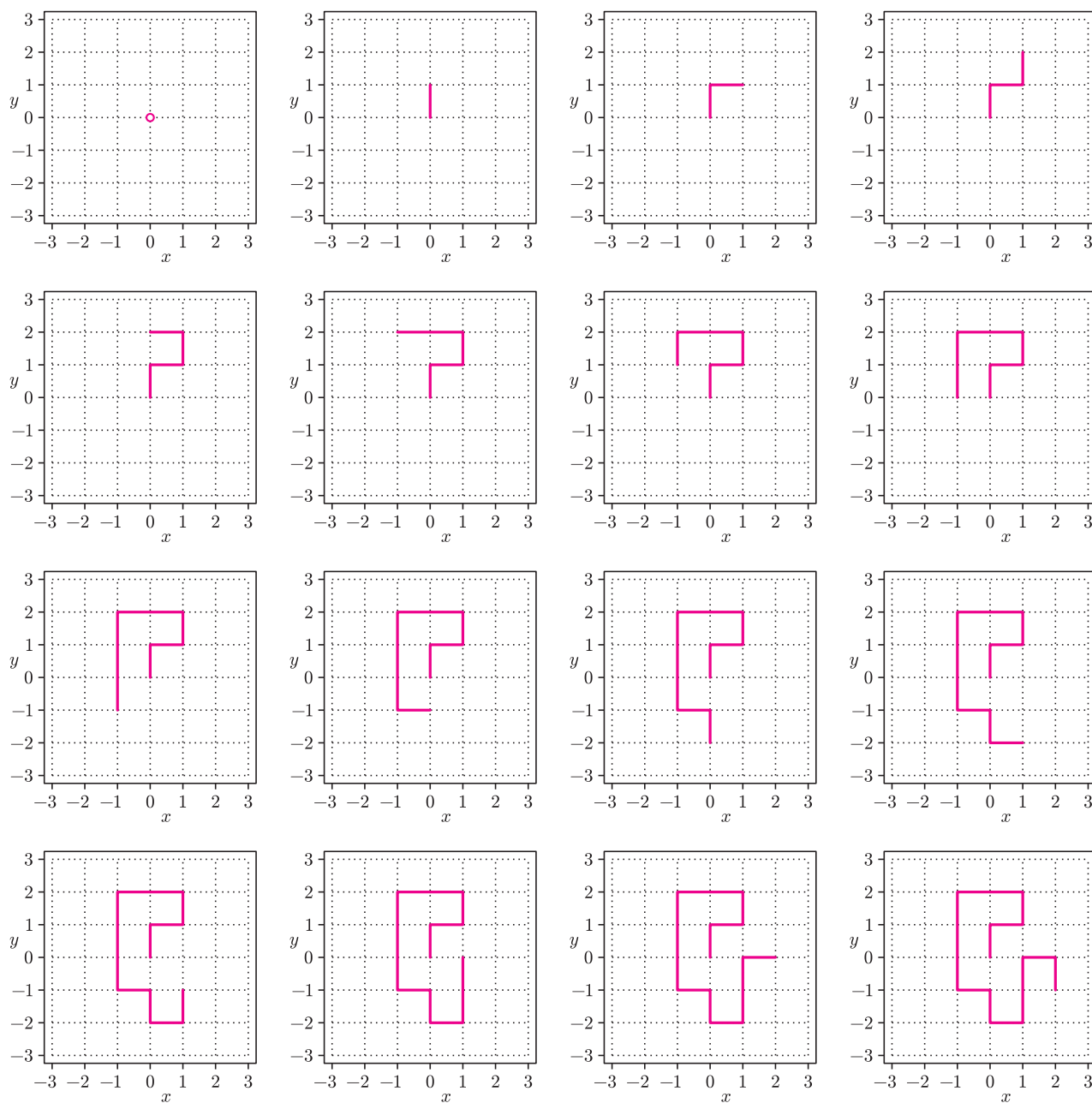
$$\frac{1}{4} \cdot \frac{1}{3} \cdot \frac{1}{3} \cdot \frac{1}{3} \cdot \frac{1}{2} \cdot \frac{1}{3} \cdot \frac{1}{2} \cdot \frac{1}{2} \cdot \frac{1}{3} \cdot \frac{1}{2} \cdot \frac{1}{3} \cdot \frac{1}{3} \cdot \frac{1}{2} \cdot \frac{1}{1} \cdot \frac{1}{3}.$$

Powiedzmy ogólniej, że przy budowaniu SAW-a x o długości l mamy kolejno

$$m_1 = 4, m_2, \dots, m_l$$

możliwości. Nie jest przy tym wykluczone, że w pewnym kroku $l < k$ nie mamy żadnej dalszej możliwości, $m_{l+1} = 0$. Mówimy wtedy, że powstał nieprzedłużalny SAW o długości l . Niech \mathcal{Y} oznacza zbiór wszystkich SAW-ów o długości $l = k$ oraz nieprzedłużalnych SAW-ów o długości $l < k$. Oczywiście, $\mathcal{Z} \subset \mathcal{Y}$.





Rys. 2. Generowanie SAW-a metodą „wzrostu”.

Prawdopodobieństwo wylosowania SAW-a $x \in \mathcal{Y}$ jest równe

$$P(x) = \frac{1}{m_1} \cdot \frac{1}{m_2} \cdots \frac{1}{m_l}.$$

Niech, dla $x \in \mathcal{Y}$, będzie

$$W(x) = \begin{cases} m_1 \cdot m_2 \cdots m_k & \text{jeżeli } x \in \mathcal{Z} \text{ (a więc } l = k); \\ 0 & \text{jeżeli } x \notin \mathcal{Z} \text{ (a więc } l < k \text{ i } m_{l+1} = 0). \end{cases}$$

Można interpretować $W(x)$ jako „wagę” SAW-a x . Powtórzmy teraz losowanie metodą wzrostu n razy. Powstaje n losowych SAW-ów X_1, X_2, \dots, X_n ze zbioru \mathcal{Y} . Obliczamy „średnią wagę” wylosowanych SAW-ów i zauważamy, że

$$(2) \quad \bar{W} = \frac{1}{n} \sum_{i=1}^n W(X_i) \approx \sum_{x \in \mathcal{Y}} P(x) W(x) = \sum_{x \in \mathcal{Z}} 1 = |\mathcal{Z}|.$$

W rezultacie,

- $\bar{W} \approx |\mathcal{Z}|$ jest to oszacowaniem liczby Z_k „metodą wzrostu”.

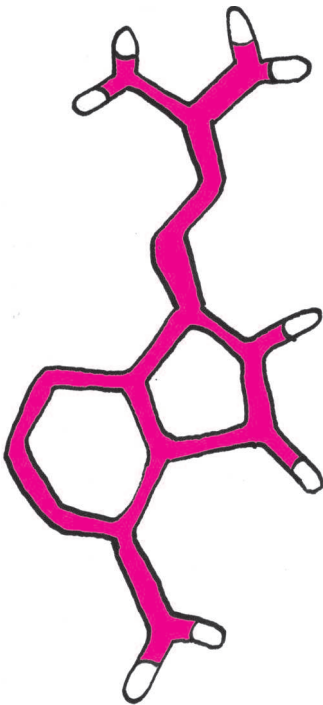
Dodajmy komentarz na temat wzoru (2). Po lewej stronie mamy średnią wagę obliczoną dla wylosowanych n elementów, czyli „średnią z próbki”. Po prawej mamy też średnią, ale „w całej populacji \mathcal{Y} ”. Wagi W zostały specjalnie dobrane tak, aby ta średnia populacyjna była równa $|\mathcal{Z}|$. Jeśli n jest dużo, dużo mniejsze od $|\mathcal{Z}|$, to obliczenie \bar{W} jest możliwe, a bezpośrednio policzenie elementów \mathcal{Z} jest niemożliwe. Kluczowy jest fakt, że średnia próbkowa przybliża średnią populacyjną. To jest odrobinę ogólniejsza wersja PWL. Poniższy rachunek jest próbą przekonania Czytelnika, że aproksymacja we wzorze (2) jest równie intuicyjna jak ta we wzorze (1). Niech n_s oznacza liczbę tych spośród wylosowanych elementów X_i , dla których $W(X_i) = s$ i $\mathcal{W}_s = \{x \in \mathcal{Y} : W(x) = s\}$. Podobnie jak we wzorze (1), na mocy PWL, $n_s/n \approx P(\mathcal{W}_s)$. Mamy zatem

$$\begin{aligned} \frac{1}{n} \sum_{i=1}^n W(X_i) &= \sum_s s \frac{n_s}{n} \approx \\ &\approx \sum_s s P(\mathcal{W}_s) = \sum_s s \sum_{x \in \mathcal{W}_s} P(x) = \sum_x P(x) W(x). \end{aligned}$$

Przy użyciu bardzo wyrafinowanych metod można obliczyć Z_k dla $k \leq 72$, ale o ile mi wiadomo, na tym koniec.

Poniższa tabelka podaje dokładne wartości Z_k , oszacowania tych liczb prostą metodą Monte Carlo (MC) i „metodą wzrostu”. Dla obu metod, dla każdego k , liczba symulowanych doświadczeń losowych była równa $n = 10000$.

k	Z_k	proste MC	metoda wzrostu
1	4	4	4
2	12	11,99	12,00
3	36	36,27	36,00
4	100	98,74	100,18
5	284	282,73	283,21
10	44100	41628,47	44041,61
20	897697164	659706977	893264097
30	16741957935348	0	$1,66208 \cdot 10^{13}$
40	300798249248474268	0	$3,067392 \cdot 10^{17}$
50	5292794668724837206644	0	$5,261991 \cdot 10^{21}$



Skąd wzięły się niepokojące zera w dolnej części tabelki? To się natychmiast stanie jasne, jeśli Czytelnik zechce obliczyć $Z_k/4^k$ dla $k = 30, 40, 50$. Z tabelki widać, że metoda wzrostu radzi sobie nie najgorzej nawet dla sporych wielkości k . Trzeba jednak dodać, że ta metoda, wynaleziona w latach pięćdziesiątych ubiegłego wieku, może być znacznie ulepszona. Istnieją bardziej efektywne algorytmy Monte Carlo, przeznaczone do zliczania SAW-ów. Niektóre z nich wykorzystują zupełnie inne pomysły niż metoda wzrostu. Niestety, opowieść o tym przekracza ramy tego artykułu.

SAW-y pojawiają się w fizyce jako najprostszy model budowy polimerów. Są to duże cząsteczki mające postać łańcucha złożonego z monomerów. Znajdują się w najróżniejszych materiałach: od włókien tkanin, tworzyw sztucznych, gumy i celulozy, aż po białka w żywych organizmach. Struktura przestrzenna łańcucha wpływa na własności fizyczne polimeru. W ogromnym uproszczeniu tę strukturę reprezentuje trójwymiarowy SAW (o długości k od ok. 10^3 do 10^5).

W przestrzeni trójwymiarowej SAW definiuje się bardzo podobnie jak dwuwymiarowy: jako ciąg punktów o trzech współrzędnych całkowitych, w którym nie ma powtórzeń. Okazuje się, że pewne istotne cechy polimerów są związane np. z odległością (euklidesową) między końcami łańcucha. Badanie tej wielkości było zasadniczą motywacją przytoczonych prac. Zadanie zliczenia SAW-ów pojawiło się „przy okazji” i zafrapowało matematyków. Na zakończenie przytoczę przypuszczenie, którego dotąd nie udało się ani udowodnić, ani obalić. Powróćmy do dwuwymiarowych SAW-ów o długości k i liczb Z_k . Przypuszcza się, że istnieje granica

$$\lim_{k \rightarrow \infty} \frac{Z_k}{k^{11/32} \mu^k} = A,$$

gdzie $0 < A < \infty$ i $2,6256 < \mu \leq 2,6792$. Coś, co zaczyna się jak przedszkolna zabawa, prowadzi do takich tajemnic.