ROCZNIKI POLSKIEGO TOWARZYSTWA MATEMATYCZNEGO
Seria II: WIADOMOŚCI MATEMATYCZNE XX (1976)

wmptm 1897

Ministerstwo Nauki i Szkolnictwa Wyższego

PATRICK BILLINGSLEY (Chicago)

Liczby pierwsze a ruch Browna*

Każda liczba całkowita wyznacza pewien rodzaj trajektorii ruchu Browna; jest to związane z faktem, że rozkłada się ona na iloczyn liczb pierwszych, i własności ruchu Browna mogą być wykorzystane dla uzyskania twierdzeń dotyczących tej faktoryzacji. Pomimo utrzymujących się poglądów głoszących, że rezultaty wysłowione w języku probabilistycznym są raczej mniej prawdziwe niż gdyby były wypowiedziane w innym języku, przedstawię te wyniki jako twierdzenia teorii prawdopodobieństwa, a nawet dam im probabilistyczne dowody. W istocie rzeczy, prawdziwych dowodów będzie tutaj mało, gdyż w przeważającej części pracy zilustruję tylko ogólne rezultaty za pomocą przykładów i przypadków szczególnych. Postępowanie to jest poparte autorytetem Williama Fellera, który zwykle mówił do nas, swych studentów, że najcenniejszą rzeczą w matematyce, podobnie jak w sztuce, literaturze i innych dziedzinach, jest ogólność ujęta w konkrecie. Pomimo że początkowo myślałem, iż jest to wyraz antymilitarnego nastawienia (angielski zwrot „general embodied in the concrete" można też tłumaczyć „generał zatopiony w betonie" — *przyp. tłum.*), zrozumiałem to ostatecznie jako jego intelektualno-estetyczną zasadę i od tej pory zawsze usiłowałem traktować ją jako podstawę swojej działalności.

Praca ma trzy rozdziały. W rozdziale 1 zdefiniowano matematyczny model dla cząsteczki poruszającej się ruchem Browna i opisano kilka jego własności. Rozdział 2, który stanowi ogniwo łączące między ruchem Browna a liczbami pierwszymi, jest poświęcony spacerowi losowemu: rzucając kolejno monetę przesuwamy się wzdłuż osi o jednostkę w kierunku dodatnim lub ujemnym w zależności od tego, czy wypadła reszka, czy orzeł. Pokazano tutaj, że jeśli oglądać spacer losowy z daleka, to wygląda on jak ruch Browna i pokazano, jak model ruchu Browna prowadzi do twierdzeń granicznych dotyczących spaceru losowego. Rozdział 3 omawia spacer losowy, który generuje wybrana losowo liczba całkowita poprzez rozkład na czynniki pierwsze: badamy kolejne liczby pierwsze 2, 3, 5, ... i stopniowo przesuwamy

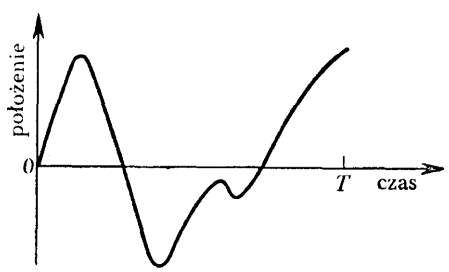
* Jest to tłumaczenie artykułu, który ukazał się w The American Mathematical Monthly 80 (1973), str. 1099–1115.

się wzdłuż osi o jednostkę w kierunku dodatnim lub ujemnym w zależności od tego, czy badana liczba pierwsza występuje w rozkładzie na czynniki czy nie. Okazuje się, że ze względu na znane twierdzenie głoszące, iż różne liczby pierwsze dzielą z osobna liczbę całkowitą wtedy i tylko wtedy, gdy dzieli ją ich iloczyn, faktoryzacyjny spacer losowy (p. str. 47) ma wiele własności zwykłego spaceru losowego związanego z rzutami monetą; w szczególności, może on być również aproksymowany przez ruch Browna, a pokazano również, jak fakt ten prowadzi do twierdzeń granicznych dotyczących rozkładu na czynniki pierwsze.

Poza elementami analizy, praca korzysta tylko z takich koncepcji probabilistycznych, jak średnia, wariancja, niezależność i rozkład normalny.

1. Ruch Browna. Wyobraźmy sobie zawieszoną w roztworze cząsteczkę, która jest bombardowana przez molekuły podczas ruchu termicznego. Cząsteczka będzie wykonywać ów nieregularny i pozornie losowy ruch opisany po raz pierwszy przez biologa Roberta Browna w 1828 r. Ponieważ będziemy interesować się dokładnie jedną składową tego ruchu, wyobraźmy sobie, że został on zrzutowany na pionową oś współrzędnych: w każdej chwili czasu t określamy wysokość $x(t)$ wspomnianej cząsteczki ponad ustaloną poziomą osią współrzędnych.

Jeżeli cząsteczka startuje z 0, to jej ruch w ciągu T jednostek czasowych jest opisany przez wartości $x(t)$ dla $0 \leqslant t \leqslant T$, to znaczy, przez ciągłą funkcję x o wartościach rzeczywistych określoną na przedziale $[0, T]$ spełniającą warunek $x(0) = 0$. Prowadzi nas to do rozważania rodziny $C_0[0, T]$ takich właśnie funkcji x.



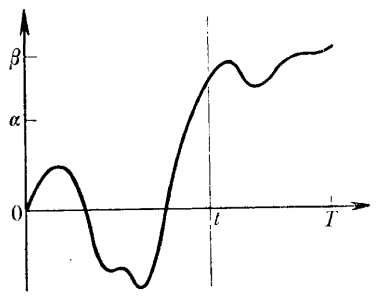
Rys. 1

Z powodów technicznych uczynimy z $C_0[0, T]$ przestrzeń metryczną, definiując odległość pomiędzy dwoma jej elementami jako maksymalną pionową odległość pomiędzy ich wykresami.

Topologia ta, tzn. topologia jednostajna, jest dla nas sama w sobie mało interesująca; wprowadzamy ją tutaj raczej jako świadectwo, że rozważania, o których będzie dalej mowa, można poprawnie sformalizować.

Losowy ruch cząsteczki jest opisany przez przyporządkowanie prawdopodobieństw $P_T(A)$ podzbiorom A przestrzeni $C_0[0, T]$: $P_T(A)$ ocenia możliwość tego, że trajektoria wyznaczona przez cząsteczkę leży w zbiorze A lub inaczej, że jest opisana przez funkcję x należącą do A. Prawdopodobieństwa te odzwierciedlają relatywne częstości z długich serii doświadczeń. Obserwując łączną sumę oczek na dwóch kostkach otrzymamy 2, 3, ..., 12 jako możliwe wyniki. Jeśli rzucamy wie-

loma parami jednorodnych kostek niezależnie, to relatywna częstość tych rzutów, które prowadzą do wyniku 7 jest w przybliżeniu równa $\frac{1}{6}$. Rozpatrując cząsteczkę w ruchu Browna podczas T jednostek czasu otrzymamy jako możliwe wyniki różne elementy z $C_0[0, T]$. Przy obserwacji wielu poruszających się niezależnie cząstek, stosunek liczby tych cząstek, których trajektorie leżą w A do liczby wszystkich cząstek jest równy w przybliżeniu $P_T(A)$. Jakkolwiek przy intuicyjnej interpretacji prawdopodobieństwa robi się tego rodzaju różnorodne obserwacje, w matematycznej teorii mówimy o pojedynczym rzucie kostkami, przypisując rzutowi prowadzącemu do wyniku 7 prawdopodobieństwo równe $\frac{1}{6}$; podobnie mówimy o pojedynczej cząsteczce, określając prawdopodobieństwo wpadnięcia jej trajektorii do zbioru A przez $P_T(A)$.



Rys. 2

Zbiór $[x\colon \alpha \leqslant x(t) \leqslant \beta]$, składający się z tych trajektorii, które przechodzą przez bramkę na rys. 2, reprezentuje zdarzenie polegające na tym, że w chwili t cząsteczka znajdzie się między α a β; prawdopodobieństwo tego zdarzenia wynosi

$$P_T[x\colon \alpha \leqslant x(t) \leqslant \beta] = \frac{1}{\sqrt{2\pi t}} \int_{\alpha}^{\beta} e^{-u^2/2t}\,du. \tag{1}$$

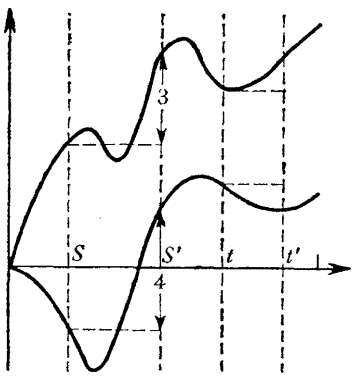
A zatem rozkład położenia cząsteczki w chwili t jest rozkładem gaussowskim ze średnią 0 i wariancją t. Średnia równa 0 odzwierciedla fakt, że cząsteczka może z równym prawdopodobieństwem przemieszczać się zarówno w górę jak i w dół. Jest ona także pozbawiona bezwładności. Wariancja t rośnie liniowo; pokazuje to, że cząsteczka ma tendencję do ucieczki od punktu startowego, a po takiej ucieczce nie podlega działaniu żadnej siły skłaniającej ją do powrotu do tegoż punktu. Równanie (1) można napisać ogólniej: przyrost na przedziale $[s, t]$ ma rozkład gaussowski ze średnią 0 i wariancją $t-s$.

Ruch Browna ma też inną ważną własność: przypuśćmy, że $s < s' < t < t'$ i rozważmy np. zdarzenie $A = [x\colon x(s')-x(s) \geqslant 3]$ polegające na tym, że cząsteczka przesuwa się w górę o co najmniej 3 jednostki w interwale czasowym $[s, s']$ oraz zdarzenie $B = [x\colon x(t')-x(t) < 0]$ opisujące sytuację, w której w przedziale $[t, t']$ położenie końcowe cząsteczki jest niższe niż początkowe. Górna trajektoria na rys. 3

leży w zbiorze A, lecz nie leży w B, natomiast dolna trajektoria należy równocześnie do A i do B.

Prawdopodobieństwa zdarzeń A, B oraz ich przekroju $A \cap B$ są powiązane wzorem:

$$P_T(A \cap B) = P_T(A)P_T(B). \tag{2}$$



Rys. 3

A i B spełniają zatem definicję niezależności; znaczy to, że przemieszczenie, jakiemu podlega cząsteczka w czasie $[s, s']$, nie wpływa w żaden sposób na to, jakiemu przemieszczeniu ulegnie ona w czasie $[t, t']$. Można to interpretować jako pewien rodzaj braku pamięci. Jakkolwiek zachowanie się cząstki w przyszłości zależy od jej obecnego położenia, nie zależy jednak od tego, jak cząsteczka położenie to osiągnęła. Równanie (2) ma również postać bardziej ogólną pokazującą, że przyrosty na dowolnej liczbie rozłącznych przedziałów czasowych są statystycznie niezależne.

Równania (1) i (2), łącznie z ich ogólniejszymi wersjami, definiują rodzinę prawdopodobieństw $P_T(A)$. (Pomijamy tutaj sprawę czysto techniczną: $P_T(A)$ może być zdefiniowane nie dla wszystkich podzbiorów A z $C_0[0, T]$, lecz tylko dla zbiorów borelowskich A, to znaczy dla każdego A z σ-ciała generowanego przez zbiory otwarte w topologii jednostajnej.) Jednym z osiągnięć Norberta Wienera było udowodnienie w 1923 r., że istnieje rodzina prawdopodobieństw spełniająca powyższe prawa, i w związku z tym P_T (odpowiadająca tej rodzinie miara na zbiorach borelowskich) jest nazywana miarą Wienera. Istnienie jej uważamy tutaj za rzecz oczywistą.

Ruch Browna opisany przez miarę Wienera spełnia prawo transformacji mające zdumiewające i głębokie konsekwencje. Przypuśćmy, że cząsteczka wykonuje ruch Browna w ciągu T jednostek czasu, oraz załóżmy, że w funkcji reprezentującej jej trajektorię zmniejszymy skalę czasu T razy, a skalę wartości $\sqrt{T}$ razy. Zgodnie z prawem, o którym wyżej mowa, nowa trajektoria będzie odpowiadała dokładnie trajektorii cząsteczki wykonującej ruch Browna podczas 1 jednostki czasu.

Aby zrozumieć dlaczego tak jest załóżmy, że x i y oznaczają starą i nową trajektorię, tak więc x leży w $C_0[0, T]$, y leży w $C_0[0, 1]$ oraz

$$y(t) = \frac{1}{\sqrt{T}} x(tT), \quad 0 \leqslant t \leqslant 1. \tag{3}$$

Naturalnie (3) definiuje odwzorowanie

$$C_0[0, T] \to C_0[0, 1]. \tag{4}$$

Prawo transformacji powiada, że jeśli x jest losowo wybraną trajektorią w $C_0[0, T]$, której rozkład jest zgodny z P_T, wówczas y jest losową trajektorią w $C_0[0, 1]$ o rozkładzie zgodnym z P_1. (Formalnie, jeśli Φ_T oznacza odwzorowanie (4), to $P_1 = P_T\Phi_T^{-1}$.) Otóż na mocy (1), funkcja $x(tT)$ jest zmienną gaussowską ze średnią 0 i wariancją tT. Mnożenie gaussowskiej zmiennej losowej przez stałą c prowadzi do zmiennej losowej o rozkładzie także gaussowskim, przy czym średnia zostaje pomnożona przez c, wariancja zaś przez c^2. A więc rozkład zmiennej $y(t)$ zdefiniowanej wzorem (3) jest rozkładem gaussowskim o średniej 0 i wariancji t (ponieważ $T^{-\frac{1}{2}} \cdot 0 = 0$, $(T^{-\frac{1}{2}})^2 \cdot tT = t$), co jest pierwszym warunkiem dla ruchu Browna. Skurczenie czasu w skali 1: T prowadzi od trajektorii określonej na przedziale $[0, T]$ do trajektorii określonej na $[0, 1]$, a zmiana skali pionowej polegająca na pomnożeniu jej przez czynniki $1/\sqrt{T}$ powoduje, że otrzymujemy właściwe wariancje. Ponadto x ma (na rozłącznych przedziałach) przyrosty niezależne, a jest rzeczą intuicyjnie jasną, że monotoniczna zmiana skal na osi czasu i osi położenia nie może zamieniać przyrostów niezależnych na zależne. A zatem transformacja (3) musi także zachować drugą własność ruchu Browna, mianowicie tę, która dotyczy przyrostów niezależnych. Powyższe argumenty, które czynią prawo transformacji wiarogodnym, mogą być uzupełnione w szczegółach do kompletnego dowodu.

Za pomocą funkcji y zdefiniowanych przez transformację (3) można pokazać, że dla dowolnych liczb dodatnich ε i K trajektoria brownowska na $[0, 1]$ będzie miała z prawdopodobieństwem większym niż $1-\varepsilon$ cięciwę o nachyleniu większym niż K. Pomysł jest następujący: chcemy, aby na odcinku $[0, 1]$ trajektoria brownowska y miała stromą cięciwę. Otrzymamy to nie wprost, lecz przez zastosowanie transformacji (3) do trajektorii ruchu Browna x na $[0, T]$ z odpowiednio wybranym T. Wybierzmy T tak duże, aby x, z prawdopodobieństwem przewyższającym $1-\varepsilon$, miało gdzieś cięciwę o nachyleniu większym niż np. 1. Taka liczba T istnieje, ponieważ w długim przebiegu mogą mieć miejsce najbardziej niezwykłe zdarzenia (podobnie jak wówczas, gdy na maszynie do pisania piszą małpy), a wystąpienie cięciwy o nachyleniu większym niż 1 jest w rzeczywistości wymaganiem zupełnie skromnym. Równocześnie wybierzmy T tak, aby przekroczyło ono K^2. Jeśli x ma cięciwę o nachyleniu większym niż 1, oraz jeśli x i y są związane wzorem (3), to y ma cięciwę o nachyleniu większym niż $\sqrt{T}$, co z kolei przekracza K.

Ponieważ ε może być dowolnie małe, a K dowolnie duże, trajektoria ruchu Browna na $[0, 1]$ z prawdopodobieństwem 1 musi mieć cięciwy o dowolnie dużych nachyleniach. Muszą także wystąpić cięciwy z dowolnie małymi ujemnymi nachyleniami, a w rzeczywistości, cięciwy (być może bardzo krótkie) z maksymalnymi nachyleniami leżą gęsto wzdłuż trajektorii. Mówiąc ściślej i w bardziej staranny sposób, powyższe argumenty pokazują, że jeśli A oznacza zbiór trajektorii w $C_0[0, 1]$ o wahaniu nieograniczonym, to $P_1(A) = 1$. Trajektoria o wahaniu nieograniczonym

przedstawia ruch cząsteczki, która błądząc tam i z powrotem przebywa nieskończoną drogę i z tego też powodu fizycy przestają się nią zajmować ze względu na swoją obsesyjną potrzebę powiązań z rzeczywistością. Jednakże fakt ten jest interesujący z matematycznego punktu widzenia, podobnie zresztą jak fakt, że $P_1(A) = 1$, jeśli przez A oznaczymy zbiór funkcji w $C_0[0, 1]$, które nie są różniczkowalne w żadnym punkcie. Skonstruowanie ciągłej, nigdzie nieróżniczkowalnej funkcji jest trudne, ale losowy wybór elementu z $C_0[0, 1]$ zgodnie z miarą P_1 daje nam taką właśnie funkcję z prawdopodobieństwem 1.

W dalszej części pracy zajmiemy się głównie zbiorami, które ściślej wiążą się z rzeczywistością. Jakkolwiek rozdziały 2 i 3 obejmą badanie transformacji (3) dla T większych od 1, w dalszej części rozdziału 1 będziemy zakładać, że $T = 1$. Użyjemy wzoru (1) w przypadku $T = t = 1$:

$$P_1[x\colon \alpha \leqslant x(1) \leqslant \beta] = \frac{1}{\sqrt{\pi}} \int_{\alpha}^{\beta} e^{-u^2/2}\,du. \tag{5}$$

Przypuśćmy, że $\alpha \geqslant 0$ i rozważmy zdarzenie $[x\colon \max x(t) \geqslant \alpha]$ polegające na osiągnięciu przez cząsteczkę poziomu α w chwili t, $0 \leqslant t \leqslant 1$. Po pierwsze

$$P_1[x\colon \max x(t) \geqslant \alpha] = P_1[x\colon \max x(t) \geqslant \alpha \text{ i } x(1) \geqslant \alpha] + \\ + P_1[x\colon \max x(t) \geqslant \alpha \text{ i } x(1) < \alpha].$$

Można dowieść, że oba prawdopodobieństwa po prawej stronie są równe, a to z grubsza mówiąc z tego powodu, że skoro raz cząsteczka osiągnie poziom α jest rzeczą równie możliwą w przypadku braku bezwładności błądzić w górę i zakończyć wędrówkę powyżej α w chwili $t = 1$, jak błądzić w dół i zakończyć poniżej α. Stąd

$$P_1[x\colon \max x(t) \geqslant \alpha] = 2P_1[x\colon \max x(t) \geqslant \alpha \text{ i } x(1) \geqslant \alpha].$$

Ponieważ warunek $\max x(t) \geqslant \alpha$ jest zbyteczny przy założeniu $x(1) \geqslant \alpha$, prawa strona jest równa $2P_1[x\colon x(1) \geqslant \alpha]$, i (5) dla $\alpha \geqslant 0$ i $\beta = \infty$ da nam

$$P_1[x\colon \max x(t) \geqslant \alpha] = \frac{2}{\sqrt{2\pi}} \int_{\alpha}^{\infty} e^{-u^2/2}\,du. \tag{6}$$

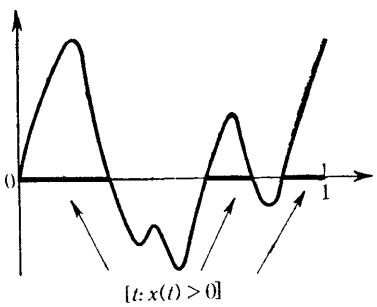
Mamy więc dystrybuantę zmiennej losowej równej największemu dodatniemu wychyleniu.

Jakkolwiek uściślenie powyższych rozumowań wymaga nieco wysiłku, wyprowadzenie wzoru (6) ma podstawy intuicyjne. Następny rezultat podamy bez dowodu i — jak wiele twierdzeń głoszonych *ex cathedra* — jest ono sprzeczne z intuicją. Rozważmy zbiór $[t\colon x(t) > 0]$ punktów czasowych t, $0 \leqslant t \leqslant 1$, w których cząsteczka przebywa powyżej poziomu 0. Zbiór ten jest sumą przedziałów (nieskończenie wielu,

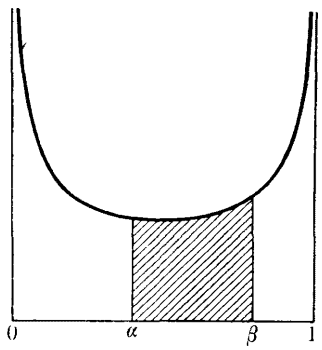
w przeciwieństwie do rys. 4). Oznaczmy pionowymi kreskami miarę Lebesgue'a tego zbioru, czyli sumę długości odcinków wchodzących w jego skład: $|[t\colon x(t) > 0]|$. Dystrybuanta tej zmiennej losowej, czyli całkowitego czasu spędzonego przez cząsteczkę powyżej poziomu 0, dana jest wzorem:

$$P_1[x\colon \alpha \leqslant |[t\colon x(t) > 0]| \leqslant \beta] = \frac{1}{\pi}\int_{\alpha}^{\beta} \frac{du}{\sqrt{u(1-u)}} \tag{7}$$

dla $0 \leqslant \alpha \leqslant \beta \leqslant 1$. Jest to prawo arc sin udowodnione przez Paula Lévy'ego, nazwane tak dlatego, że wykonanie całkowania prowadzi do funkcji arcus sinus.



Rys. 4



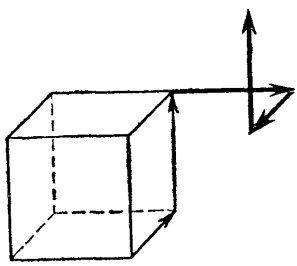
Rys. 5

Rysunek 5 pokazuje kształt gęstości rozkładu, pole zakreskowanego obszaru jest równe prawej stronie wzoru (7). Krzywa ma kształt litery *U*, tak więc jeśli długość przedziału $\beta - \alpha$ jest stała, to prawdopodobieństwo zdarzenia $[x\colon \alpha \leqslant |[t\colon x(t) > 0]| \leqslant \beta]$ rośnie wraz ze zbliżaniem się przedziału do 0 lub do 1, a jest najmniejsze, kiedy środkiem przedziału jest $\frac{1}{2}$. Fakt ten jest dziwny, ponieważ czas spędzony powyżej poziomu 0 ma średnią $\frac{1}{2}$ ze względu na symetrię, a wartości położone blisko wartości oczekiwanej zmiennej losowej występują zazwyczaj z większym prawdopodobieństwem niż wartości daleko oddalone od średniej, podczas gdy tutaj ma miejsce właśnie sytuacja przeciwna.

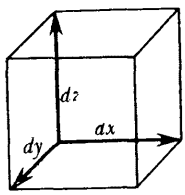
Ogólne informacje o ruchu Browna znajdują się w [4] i [7].

2. Spacer losowy. Wyobraźmy sobie cząsteczkę poruszającą się losowo po wierzchołkach siatki przestrzennej. Cząsteczka może się przesuwać w dowolnym spośród sześciu kierunków (północ, południe, wschód, zachód, w górę, w dół) aż do sąsiedniego wierzchołka. Kierunek określamy rzucając jednorodną (zrównoważoną) kostką do gry, cząsteczka przechodzi do następnego wierzchołka, ponownie rzucamy kostkę aby określić kierunek następnego posunięcia i dalej procedura powtarza się. Rysunek 6 pokazuje pięć kroków takiego właśnie losowego spaceru łącznie z jedną komórką siatki. Rysunek jest w duchu owej czcigodnej książki z analizy wektorowej, która dowód twierdzenia Gaussa rozpoczynała od nakazania

czytelnikowi rozważenia „nieskończenie małego elementu objętości o wymiarach *dx*, *dy* i *dz*". Nakazowi temu towarzyszył starannie wyrysowany diagram podobny do rys. 7, o którym mówiono, że pokazuje taki właśnie nieskończenie mały element objętości „w dużym powiększeniu". A więc rys. 6 jest też bardzo powiększony, ale jeśli kostki w siatce są w rzeczywistości bardzo małe, a cząsteczka porusza się z dużą szybkością od wierzchołka do wierzchołka, to jest rzeczą naturalną oczekiwać, że taki ruch przybliża ruch Browna.

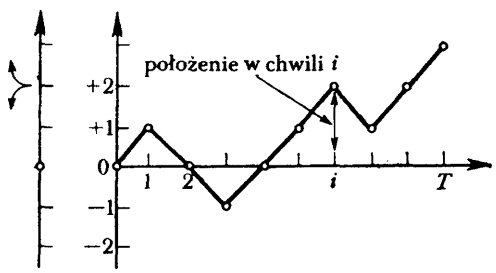


Rys. 6



Rys. 7

Wykorzystamy jednomiarową wersję tego pomysłu. Rozważmy pionową oś, na którą naniesiono punkty całkowite $0, \pm 1, \pm 2, \ldots$ Startujemy z 0, rzucamy monetę i przesuwamy się o jedną jednostkę w górę jeśli wypadnie reszka, a o jedną jednostkę w dół jeśli wypadnie orzeł. W nowym położeniu ($+1$ lub -1) rzucamy znowu monetę i przesuwamy się o jednostkę w górę lub w dół, w zależności od tego, czy wypadnie reszka, czy orzeł, a następnie prowadzimy opisany proces przez T kroków, T jest tutaj liczbą całkowitą. Jeśli każdy krok opisanego spaceru losowego wykonamy w jednostce czasu, a wierzchołki połączymy liniowo, nasze zmieniające się położenie będzie opisywane przez funkcję podobną do funkcji na rys. 8, tzn. do trajek-



Rys. 8

torii, która jest łamaną o wysokości w punkcie i równej położeniu w momencie i — a zatem położeniu po i krokach. Każda z 2^T możliwych trajektorii ma prawdopodobieństwo 2^{-T}. (Rozmaite fakty dotyczące spaceru losowego znaleźć można w [3].)

Na trajektorię można też patrzeć jako na opis zmian w fortunie gracza. Położenie na osi pionowej przedstawia fortunę gracza (relatywnie do kapitału początko-

wego, a więc zaczyna on umownie od 0) i zmienia się ono w każdym kroku w górę lub w dół o jedną jednostkę — powiedzmy o jeden funt — stosownie do tego, czy wygrał on, czy przegrał następną grę.

Trajektoria spaceru losowego ma niektóre własności trajektorii ruchu Browna na $[0, T]$. Po pierwsze, dla liczb całkowitych spełniających $i < i' < j < j'$ zmiany położenia zachodzące na odcinkach czasowych $[i, i']$ i $[j, j']$ są niezależne, ponieważ zależą one od rozłącznych zbiorów rzutów monetą, o których zakładamy, że są niezależne (moneta nie ma pamięci). A zatem trajektoria ma zasadniczo przyrosty niezależne (dla przedziałów, których końcowe punkty nie są liczbami całkowitymi przyrosty mogą być nieznacznie zależne). Odległość przebywana w jednostce czasu ma średnią równą

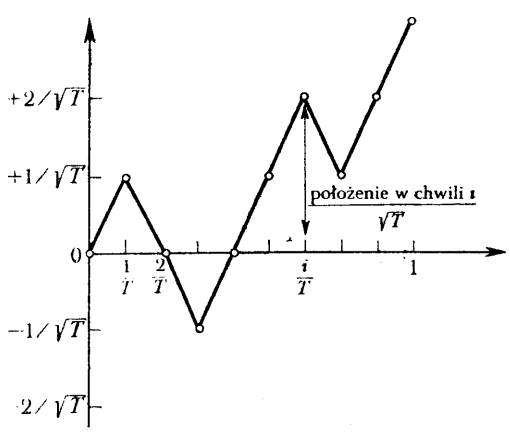
$$(+1)\cdot\tfrac{1}{2}+(-1)\cdot\tfrac{1}{2}=0 \tag{8}$$

i wariancję

$$(+1)^2\cdot\tfrac{1}{2}+(-1)^2\cdot\tfrac{1}{2}=1. \tag{9}$$

Tak więc położenie w chwili i ma średnią 0 oraz, ze względu na niezależność, wariancję równą i, podobnie jak ruch Browna (na mocy (1)). (Dla niecałkowitych t położenie w chwili t ma średnią 0, lecz wariancja jest tylko aproksymowana przez t). Pomimo że trajektorie ruchu Browna nie mają charakteru łamanych, zamiana obu skal (skali czasu i skali położenia) spowoduje, że w granicy, gdy $T \to \infty$, prostoliniowe odcinki na rys. 8 znikną.

Przypuśćmy, że zmieniamy skalę czasu za pomocą czynnika T, a skalę pionową za pomocą czynnika $\sqrt{T}$, i stosując przekształcenie (3) otrzymujemy z rysunku 8 rysunek 9. Na rys. 8 odcinki mają długość $\sqrt{2}$, podczas gdy na rys. 9 są one bardzo



Rys. 9

krótkie dla dużych T, mianowicie długość ich jest rzędu $1/\sqrt{T}$. Jeśli rys. 8 przedstawiał trajektorię ruchu Browna na $[0, T]$ wówczas, zgodnie z tym, co powiedziano w rozdziale 1, rys. 9 przedstawia trajektorię ruchu Browna na $[0, 1]$. Przekształcenie (3)

pozostawia bez zmian te charakterystyki (średnie, wariancje, niezależność przyrostów), które wyjściowa trajektoria ma takie jak ruch Browna, natomiast ma tendencję do zacierania tych cech (liniowość przedziałami), które trajektorię tę od ruchu Browna odróżniają. Z tych więc powodów możemy oczekiwać, że krzywa na rys. 9 będzie bardzo podobna do trajektorii ruchu Browna dla dużych T. I rzeczywiście, prawdą jest, że

$$\text{(10)} \qquad \mathrm{P}[\text{trajektoria} \in A] \to P_1(A), \quad \text{gdy } T \to \infty$$

dla podzbiorów A przestrzeni $C_0[0, 1]$, gdzie $P_1(A)$ oznacza miarę Wienera zbioru A. Trajektorii podobnych do tej na rys. 9 jest ogółem 2^T, a $\mathrm{P}[\text{trajektoria} \in A]$ jest równe 2^{-T} pomnożone przez liczbę trajektorii należących do zbioru A.

W celu zilustrowania powyższego twierdzenia przypuśćmy, że A we wzorze (10) jest zbiorem $[x\colon \alpha \leqslant x(1) \leqslant \beta]$ tych trajektorii z $C_0[0, 1]$, które w punkcie $t = 1$ leżą między α a β. Ponieważ wysokość w punkcie $t = 1$ na rys. 9 jest równa iloczynowi $1/\sqrt{T}$ i położenia w chwili T w wyjściowym spacerze losowym, z (10) i (5) wynika razem wzór

$$\text{(11)} \qquad \mathrm{P}\left[\alpha \leqslant \frac{\text{położenie w chwili } T}{\sqrt{T}} \leqslant \beta\right] \to \frac{1}{\sqrt{2\pi}} \int_\alpha^\beta e^{-u^2/2} du.$$

Powyższy wzór to klasyczne centralne twierdzenie graniczne De Moivre'a–Laplace'a dla prób Bernoulliego. Opisuje ono położenie po wykonaniu dużej liczby kroków w spacerze losowym lub majątek gracza-hazardzisty pod koniec tury, w której stawiał on T razy. Jeśli $-\alpha = \beta = 0{,}9$, granica w (11) jest równa w przybliżeniu 0,6. Dla $T = 100$ gracz ma wówczas szansę równą w przybliżeniu 0,6 zakończenia wieczoru z sumą różniącą się od jego początkowego kapitału o nie więcej niż $0{,}9 \times \sqrt{100} = 9$ funtów.

Przypuśćmy teraz, że A jest zbiorem w (6), tzn. zbiorem trajektorii z $C_0[0, 1]$ mających w jakimś punkcie wysokość co najmniej α (zakładamy tu, że $\alpha \geqslant 0$). Trajektoria na rys. 9 należy do A, jeśli w pewnym momencie gry majątek gracza przewyższa o $\alpha\sqrt{T}$ funtów jego pierwotny kapitał, a ze względu na (10) prawdopodobieństwo tego zdarzenia zmierza do prawej strony wzoru (6). Dla $\alpha = 1{,}7$ wartość tej granicy jest w przybliżeniu równa 0,1. Dla $T = 100$, daje to prawdopodobieństwo około 0,1, że gracz zarobi co najmniej $1{,}7 \times \sqrt{100} = 17$ funtów w momencie, w którym opuści grę.

Przypuśćmy na koniec, że A jest zbiorem $[x\colon \alpha \leqslant |t\colon x(t) > 0| \leqslant \beta]$ w (7). W czasie wieczoru gracz wygrywa przez pewną frakcję czasu; jeśli krzywa na rys. 9 opisuje historię jego majątku, to należy ona do zbioru A wtedy i tylko wtedy, gdy frakcja ta leży między α a β. Prawdopodobieństwo tego zdarzenia jest na mocy (10) i (7) niemal równe polu obszaru zakreskowanego na rys. 5. Po obliczeniu pól okazuje się, że prawdopodobieństwo wygrywania przez gracza w przeciągu od 45% do 55% czasu gry wynosi zaledwie około 0,06, podczas gdy prawdopodobieństwo, że wygrywa on przez więcej niż 90% czasu wynosi około 0,2. Zatem w ciągu jednego

wieczoru na pięć gracz będzie wówczas wygrywał przez więcej niż 90% czasu gry. Ze względu na symetrię, w ciągu jednego wieczoru na pięć gracz będzie wygrywał przez mniej niż 10% czasu gry. Przekonanie go w pierwszym (drugim) przypadku, że wynik jego doświadczenia jest zupełnie przypadkowy i nie jest spowodowany tym, że jest on dzieckiem szczęścia (pechowcem), będzie rzeczą trudną (niemożliwą).

Zastosowaliśmy więc (10) do trzech interesujących zbiorów. Jeśli A jest zbiorem funkcji z $C_0[0, 1]$ o nieskończonym wahaniu, wówczas $P_1(A) = 1$ na mocy tego, co powiedziano w rozdziale 1, podczas gdy P[trajektoria $\in A$] = 0, ponieważ krzywa na rys. 9 ma w oczywisty sposób wahanie skończone. Zatem (10) nie zachodzi dla pewnych podzbiorów A z $C_0[0, 1]$. Matematycznie ściśle można powiedzieć, że (10) zachodzi dla tych zbiorów (borelowskich), których brzeg ∂A (brzeg w sensie topologii jednostajnej) spełnia warunek $P_1(\partial A) = 0$. Warunek ten był spełniony we wszystkich trzech przykładach, a nie zachodził, gdy A było zbiorem funkcji o wahaniu nieskończonym. Szczegółowy dowód tego twierdzenia wymaga zastosowania teorii prawdopodobieństwa i analizy funkcjonalnej; szczegóły można znaleźć w [1].

3. Dzielniki pierwsze. Zgodnie z podstawowym twierdzeniem arytmetyki, każda liczba całkowita rozkłada się na czynniki pierwsze i rozkład ten jest jednoznaczny z dokładnością do uporządkowania czynników (patrz np. [5]). Niech $f(n)$ oznacza liczbę różnych czynników pierwszych występujących w rozkładzie liczby n; nie uwzględniamy potęg: $f(3^4 \cdot 5^2)$ jest równe 2, a nie 6. Tabela 1 pokazuje kilka wartości funkcji f. Funkcja f przybiera większe wartości bardzo wolno.

Tabela 1

n	2	3	4	5	6	7	...	29	30	31	...	209	210	211	...
$f(n)$	1	1	1	1	2	1	...	1	3	1	...	2	4	1	...

Najmniejsze liczby n, dla których wartości funkcji f wynoszą odpowiednio 2, 3 i 4, są równe $2 \cdot 3 = 6$, $2 \cdot 3 \cdot 5 = 30$, $2 \cdot 3 \cdot 5 \cdot 7 = 210$. Jednakże nieskończoność zbioru liczb pierwszych implikuje fakt, że f przyjmuje wartości dowolnie duże; z tego samego wynika, że f opada w dół do 1 nieskończenie wiele razy, ponieważ $f(p) = 1$ dla liczb pierwszych p.

Naturalne jest pytać o średnie zachowanie się funkcji f, ponieważ zmienia się ona w dość nieregularny sposób. Można np. pokazać, że

$$\frac{1}{N} \sum_{n=1}^{N} f(n) \approx \log \log N \tag{12}$$

(patrz uwagi następujące po wzorze (17)). Ponieważ $\log \log 10^{70} \approx 5$, typowa liczba całkowita mniejsza niż 10^{70} ma pięć dzielników pierwszych. Pytania delikatniejszej natury wiążą się z rozkładem wartości funkcji f.

Niech S oznacza podzbiór zbioru liczb naturalnych, a $P_N(S)$ frakcję tych liczb całkowitych spośród $1, 2, \ldots, N$, które należą do S:

$$P_N(S) = \frac{1}{N} \times \text{card}[n\colon 1 \leqslant n \leqslant N \text{ i } n \in S]. \tag{13}$$

Zadanie polega na uzyskaniu informacji o wielkościach typu $P_N[n\colon a \leqslant f(n) \leqslant b]$.

Otóż na (13) można patrzeć jak na prawdopodobieństwo. Wybieramy losowo liczbę całkowitą spośród liczb $1 \leqslant n \leqslant N$, a $P_N(S)$ oznacza wówczas prawdopodobieństwo, że liczba ta należy do S. Fakt, że $P_N[n\colon a \leqslant f(n) \leqslant b]$ można traktować jako prawdopodobieństwo, sam przez się nie zapewnia (choć być może trudno w to uwierzyć), że teoria prawdopodobieństwa będzie pomocna w obliczeniach. W rzeczywistości jest on bardzo pomocny ze względu na możliwość wykorzystania pojęcia niezależności. Jeśli $\delta_p(n)$ jest równe 1 lub 0 w zależności od tego, czy liczba pierwsza p jest dzielnikiem n czy nie, to $f(n) = \sum_p \delta_p(n)$. Zrozumiemy co oznacza dystrybuanta funkcji $f(n)$, jeśli będziemy znali łączne zachowanie się funkcji $\delta_p(n)$ jako wielkości losowych.

Liczba wielokrotności liczby p mniejszych od N jest równa części całkowitej $[N/p]$ liczby N/p. Prawdopodobieństwo tego, że $\delta_p(n) = 1$, czyli że $p|n$ jest więc równe

$$P_N[n\colon p|n] = \frac{1}{N}\left[\frac{N}{p}\right] \approx \frac{1}{p}. \tag{14}$$

Przybliżenie w powyższym wzorze jest zadowalające dla dużych N: ponieważ $[N/p]$ różni się od N/p o mniej niż 1, błąd we wzorze (14) jest mniejszy niż $1/N$. Wzór (14) stanowi odbicie faktu, że p dzieli co p-tą liczbę całkowitą, i w żadnej mierze nie wymaga, aby p było liczbą pierwszą.

Podstawowe twierdzenie arytmetyki głosi, że jeśli liczby całkowite a i b są względnie pierwsze (nie mają żadnych wspólnych dzielników pierwszych), wówczas każda z nich z osobna dzieli n wtedy i tylko wtedy, gdy ich iloczyn ab jest dzielnikiem n. Fakt ten dobrze ilustruje przykład, o którym legenda mówi, że został spożytkowany w praktyce przez Turinga. Koło zębate w jego rowerze miało wadliwy ząb, a łańcuch złe ogniwo, i jeśli tylko nie pedałował on bardzo szybko, to kiedy te wadliwe części szczepiały się, łańcuch spadał. Policzył on wówczas liczbę, powiedzmy a, zębów na kole i liczbę, powiedzmy b, ogniw w łańcuchu i stwierdził, że a i b były względnie pierwsze, co nie było zresztą dla niego niespodzianką. Między kolejnym spotkaniem wadliwego zęba ze złym ogniwem koło zębate wykonywało więc b cykli, podczas gdy łańcuch obracał się przez a cykli. Turing opowiadał, że pedałował równocześnie licząc i dodawał nagle prędkości przy każdym b-tym cyklu zębatki, co pozwalało mu na przeskoczenie niebezpiecznego miejsca.

Jako szczególny przypadek powyższych uwag, różne liczby pierwsze p i q dzielą z osobna n wtedy i tylko wtedy, gdy dzieli je pq. Stąd i z (14), gdzie p zastąpiono przez

pq, wynika, że

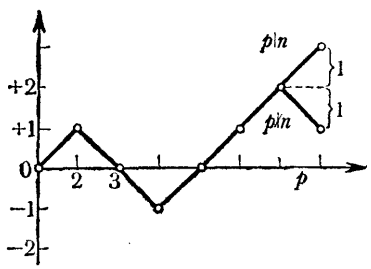
$$P_N[n\colon p|n \text{ i } q|n] = P_N[n\colon pq|n] = \frac{1}{N}\left[\frac{N}{pq}\right] \approx \frac{1}{pq} = \frac{1}{p}\cdot\frac{1}{q}.$$

Ponieważ ze względu na (14) ułamki $1/p$ i $1/q$ aproksymują odpowiednio $P_N[n\colon p|n]$ i $P_N[n\colon q|n]$ dla dużych N, otrzymujemy stąd

$$P_N[n\colon p|n \text{ i } q|n] \approx P_N[n\colon p|n]\cdot P_N[n\colon q|n]. \tag{15}$$

Zatem zdarzenia $[n\colon p|n]$ i $[n\colon q|n]$ spełniają w sposób przybliżony definicję niezależności, o ile n jest losowe, $1 \leqslant n \leqslant N$, a N dostatecznie duże. Wzór (15) można rozszerzyć z dwóch do trzech lub więcej liczb pierwszych.

Można wykorzystać powyższy fakt, aby skonstruować pewien rodzaj spaceru losowego zawierający informacje o rozkładzie liczby n na czynniki pierwsze, a w szczególności o $f(n)$. Wybieramy losowo liczbę całkowitą n spośród liczb $1, 2, \ldots, N$. Na osi pionowej, na której zaznaczono liczby całkowite startujemy z punktu 0 i przesuwamy się o jedną jednostkę w górę, jeśli $2|n$ i jedną jednostkę w dół, jeśli $2\nmid n$. Z nowego położenia ($+1$, lub -1) przesuwamy się o jednostkę w górę, jeśli $3|n$, a o jednostkę w dół jeśli $3\nmid n$. Dalej postępujemy w opisany sposób badając kolejno wszystkie liczby pierwsze. Rysunek 10 opisuje spacer losowy związany z roz-



Rys. 10

kładem na czynniki pierwsze (który będziemy nazywać dalej *faktoryzacyjnym spacerem losowym* — przyp. tłum.) podobnie jak rys. 8 opisuje spacer losowy otrzymany przez rzucanie monetą. Każda liczba na osi czasu jest liczbą pierwszą odpowiadającą jednemu krokowi w opisanym spacerze losowym. Zastanowimy się dalej, jak długo kontynuować opisany wyżej proces.

Powyższa trajektoria jest przypadkowa, ponieważ n wybrano losowo. Skoro jednak losowość objawia się tylko w wyborze liczby n zanim jeszcze spacer się zacznie, faktoryzacyjny spacer losowy może wydawać się mniej losowy niż spacer losowy otrzymany przy rzutach monetą. Wrażenie to jest błędne. Można wyobrazić sobie proces rzucania monetą T razy przed spacerem, zakodowania ciągu reszek i orłów i wykonania dopiero wtedy odpowiedniego spaceru. Ponieważ widzielibyśmy

całą jego historię zarejestrowaną przed odbyciem, spacer byłby nieciekawy. Wyobraźmy sobie zatem przyjaciela, który rzuca monetę T razy i notuje wyniki przed podróżą, ale zamiast pokazać nam wszystkie wyniki równocześnie, ujawnia je nam jeden po drugim, w miarę jak wykonujemy spacer. Przywraca to stan niepewności. Dla spaceru losowego związanego z faktoryzacją możemy wyobrazić sobie przyjaciela, który losowo wybiera liczbę n, $1 \leqslant n \leqslant N$, rozkłada ją na czynniki pierwsze i w każdym kroku spaceru ujawnia nam, czy odpowiednia liczba p dzieli n czy nie.

Przyrosty losowej trajektorii z rys. 10 na odcinku zależą od tego, jak wiele liczb pierwszych z odpowiedniego zbioru dzieli liczbę n. Przyrosty na rozłącznych odcinkach zależą od rozłącznych zbiorów liczb pierwszych, a stąd na mocy (15) — czy raczej (15) razem z jej uogólnieniem na trzy i więcej liczb pierwszych — przyrosty te będą w przybliżeniu niezależne dla dużych N. Jednak w przeciwieństwie do ruchu Browna w faktoryzacyjnym spacerze losowym występuje silna bezwładność ku dołowi. Ze względu na (14), prawdopodobieństwo przemieszczenia się w dół przy kroku odpowiadającym liczbie p jest w przybliżeniu równe $1-1/p$, co dla dużych p jest prawie równe 1. Lekarstwem na to jest przesuwać się w górę o $1-1/p$ jeśli $p|n$ i przesuwać się w dół zaledwie o $1/p$ jeśli $p\nmid n$. Średnie przesunięcie jest teraz równe

$$(1-p^{-1})\mathrm{P}_N[n\colon p|n]+(-p^{-1})\mathrm{P}_N[n\colon p\nmid n],$$

co na mocy (14) jest równe w przybliżeniu

$$\left(1-\frac{1}{p}\right)\frac{1}{p}+\left(-\frac{1}{p}\right)\left(1-\frac{1}{p}\right)=0.$$

Jest to odpowiednik równania (8), które pokazywało, że spacer losowy otrzymany za pomocą rzutów monetą jest pozbawiony bezwładności.

Ponieważ średnia odległość przebywana w kroku odpowiadającym liczbie p jest równa w przybliżeniu 0, wariancja jest w granicy równa

$$(1-p^{-1})^2\mathrm{P}_N[n\colon p|n]+(-p^{-1})^2\mathrm{P}_N[n\colon p\nmid n],$$

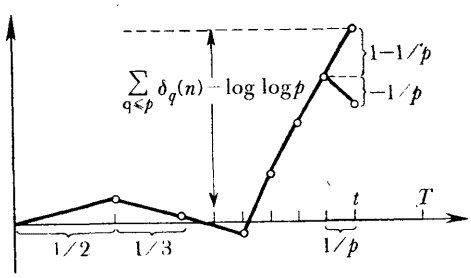
co na mocy (14) jest w przybliżeniu równe

$$\left(1-\frac{1}{p}\right)^2\frac{1}{p}+\left(-\frac{1}{p}\right)^2\left(1-\frac{1}{p}\right)=\frac{1}{p}\left(1-\frac{1}{p}\right)\approx\frac{1}{p}.$$

Przebywana odległość staje się więc bardzo mała dla dużych p, w przeciwieństwie do spaceru losowego generowanego przez rzuty monetą, który na mocy (9) rozwija się z impetem nie malejącym nigdy. Tym razem lekarstwem jest poświęcenie tylko $1/p$ ilości czasu na wykonanie kroku odpowiadającego liczbie p. Po uwzględnieniu obu powyższych modyfikacji trajektoria przypomina tę z rys. 11.

Podsumowując stwierdzamy, że odcinek czasowy odpowiadający liczbie pierwszej p ma długość $1/p$. Na tym przedziale trajektoria zmienia się o wielkość równą $\delta_p(n)-1/p$; znaczy to, że zmienia się ona o $1-1/p$ jeśli $p|n$ (przybliżone prawdopodobieństwo tego faktu jest równe $1/p$), i zmienia się o $0-1/p$, jeśli $p\nmid n$ (prawdopodobieństwo tego zdarzenia jest równe około $1-1/p$).

Punkt t na rys. 11 (prawy koniec przedziału odpowiadającego liczbie p) jest równy $\sum_{q\leqslant p} 1/q$ (sumowanie prowadzimy po liczbach pierwszych q nie przewyższających p). Odległość przebywana w kroku odpowiadającym q ma wariancję równą około $1/q$, a stąd na mocy przybliżonej niezależności poszczególnych kroków (znów (15)), wariancja położenia w chwili t jest równa w przybliżeniu $\sum_{q\leqslant p} 1/q$, lub po prostu t. Powyższa modyfikacja faktoryzacyjnego spaceru losowego nie tylko wyeliminowała bezwładność, ale i poprawiła tak skalę czasową, że wariancje są mniej więcej takie, jak dla spaceru losowego generowanego przez rzuty monetą i dla ruchu Browna.



Rys. 11

Można pokazać, że

$$\sum_{q\leqslant u}\frac{1}{q}\approx \log\log u \tag{16}$$

dla dużych u (oba wyrażenia dążą do nieskończoności wraz z u, a ich różnica jest ograniczona; patrz [5], str. 351). Dla dalszego ciągu jest nieistotne, że suma w (16), zamiast wzrastać w sposób nieco nieregularny, jest asympotycznie równoważna standardowej funkcji w rodzaju $\log\log u$; wzory staną się jednak prostsze (i pozostaną prawdziwe), jeśli przy każdym pojawieniu się sumy z lewej strony wzoru (16) zastąpimy ją prawą stroną (16).

A zatem t na rys. 11 jest zasadniczo równe $\log\log p$, a wysokość $\sum_{q\leqslant p}[\delta_q(n)-1/q]$ krzywej w punkcie t jest równa w przybliżeniu

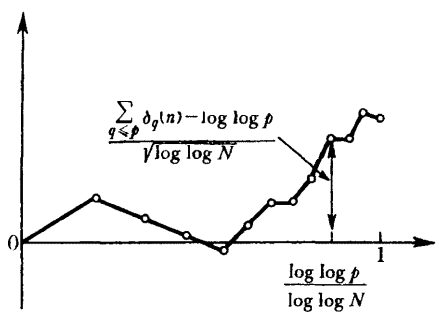
$$\sum_{q\leqslant p}\delta_q(n)-\log\log p. \tag{17}$$

Otóż n ma $\sum_{q\leqslant p}\delta_q(n)$ dzielników pierwszych, które nie przekraczają p, i unormujemy tę wielkość odejmując wartość $\log\log p$ równą liczbie dzielników dla „przeciętnego" n. (Jeśli n jest wybrane losowo, $1\leqslant n\leqslant N$, wówczas $\sum_{q\leqslant p}\delta_q(n)$ ma średnią równą około $\log\log p$ na mocy wzorów (14) i (16); stąd właśnie wynika (12).) Faktoryzacyjny spacer losowy jest właśnie zapisem tych różnic (17). Spacer kontynuujemy do momentu, kiedy zostaną wyczerpane wszystkie liczby pierwsze $p\leqslant N$, a odpowiada-

jący temu punkt na osi czasu jest równy

$$T = \sum_{p \leqslant N} \frac{1}{p} \approx \log\log N.$$

Losowa trajektoria przypomina teraz trajektorię spaceru generowanego przez rzuty monetą tym, że jej przyrosty są prawie niezależne dla dużych N, nie ma w istocie bezwładności, a wariancje są prawie równe tamtym. Podobnie jak poprzednio przeskalowanie osi prowadzi w granicy (dla $N \to \infty$) do ruchu Browna. Aby przeprowadzić punkt T na 1 dzielimy skalę poziomą przez czynnik $T = \log\log N$ i, podobnie jak w przypadku spaceru związanego z rzutami monetą i z tych samych powodów, dzielimy skalę pionową przez $\sqrt{T}$ stosując przekształcenie (3). Punkt t z rys. 11 przechodzi na $\log\log p/\log\log N$, a otrzymana trajektoria jest pokazana na rys. 12.



Rys. 12

Ponieważ trajektoria zależy od wyboru n i N, oznaczmy ją przez $\mathrm{traj}_N(n)$. Skoro n wybieramy losowo ($1 \leqslant n \leqslant N$), trajektoria też jest losowa, a prawdopodobieństwo, że należy ona do danego podzbioru A przestrzeni $C_0[0, 1]$ jest równe $\mathrm{P}_N[n\colon \mathrm{traj}_N(n) \in A]$. Następujące twierdzenie wiąże liczby pierwsze z ruchem Browna: *jeśli A jest podzbiorem (borelowskim) przestrzeni $C_0[0, 1]$ spełniającym warunek $P_1(\partial A) = 0$, wówczas*

$$\mathrm{P}_N[n\colon \mathrm{traj}_N(n) \in A] \to P_1(A) \quad (N \to \infty). \tag{18}$$

gdzie $P_1(A)$ oznacza miarę Wienera. Dowód wzoru (18) wykorzystuje metody teorii prawdopodobieństwa, analizy funkcjonalnej i teorii liczb. *Implicite* twierdzenie to znajduje się w pracy [8], str. 122, *explicite* w książkowym wydaniu [1], a w postaci bardziej ogólnej w [9]. (Ogólne omówienie probabilistycznych metod w teorii liczb można znaleźć w [6], [8] i wykładach autora z 1973 r., które ukażą się drukiem w Annals of Probability).

Z rys. 12 czyli wykresu różnic (17) unormowanych następująco

$$\Big(\sum_{q \leqslant p} \delta_q(n) - \log\log p\Big)\Big/\sqrt{\log\log N} \tag{19}$$

możemy odczytać arytmetyczne własności n, a więc (18) prowadzi do twierdzeń granicznych arytmetyki. Rozważmy trzy zbiory A, do których zastosowaliśmy analogiczny wynik (10). Krzywa na rys. 12 wznosi się w punkcie czasowym 1 na wysokość równą wyrażeniu (19), w którym p zastąpiono przez N; wysokość ta jest równa liczbie $f(n)$ dzielników pierwszych n, unormowanej do

$$\left(f(n)-\log\log N\right)/\sqrt{\log\log N}.$$

Kiedy liczba ta jest największa, n jest w największym stopniu złożone, a kiedy najmniejsza n jest najbardziej „podobna" do liczby pierwszej. Dla $A = [x\colon \alpha \leqslant x(1) \leqslant \beta]$, ze wzorów (18) i (5) wynika, że

$$\text{(20)} \qquad \mathrm{P}_N\left[n\colon \alpha \leqslant \frac{f(n)-\log\log N}{\sqrt{\log\log N}} \leqslant \beta\right] \to \frac{1}{\sqrt{2\pi}} \int_{\alpha}^{\beta} e^{-u^2/2}du.$$

Powyższa relacja to centralne twierdzenie graniczne Erdösa–Kaca dla funkcji f. (Elementarny dowód wzoru (20) można znaleźć w [2].)

Dla $-\alpha = \beta = 0{,}9$, granica we wzorze (20) jest równa około 0,6, a jeśli $N = 10^{70}$ tak, że $\log\log N \approx 5$, podwójna nierówność w (20) jest w granicy równoważna nierówności $-0{,}9 \leqslant [f(n)-5]/\sqrt{5} \leqslant 0{,}9$, co z kolei znaczy w przybliżeniu to samo, co $3 \leqslant f(n) \leqslant 7$. A zatem około 60% liczb całkowitych mniejszych niż 10^{70} ma od 3 do 7 dzielników pierwszych.

Im wyrażenie (17) jest większe przy ustalonym p, tym bardziej złożone wydaje się n przy badaniu jego podzielności przez liczby pierwsze nie przekraczające p; znaczy to, że (17) mierzy względną złożoność n w zależności od p. Maksimum względnej złożoności jest wyznaczone przez

$$\text{(21)} \qquad \max_{p \leqslant N}\left(\sum_{q \leqslant p} \delta_q(n) - \log\log p\right);$$

ponieważ powyższe wyrażenie jest równe iloczynowi $\sqrt{\log\log N}$ i maksymalnej wysokości krzywej na rys. 12, zastosowanie (18) do zbioru występującego w (6) daje w granicy jego dystrybuantę.

Dla $\alpha = 1{,}7$ prawa strona (6) jest równa około 0,1, co oznacza, że dla około 10% liczb całkowitych mniejszych od 10^{70} wyrażenie (21) przewyższa $1{,}7 \times \sqrt{5} \approx 3{,}8$.

Powiemy, że n jest *ekscesywne w punkcie* p, jeśli

$$\text{(22)} \qquad \sum_{q \leqslant p} \delta_q(n) > \operatorname{loglog} p;$$

zachodzi to wówczas, gdy przy badaniu podzielności przez liczby pierwsze mniejsze od p, n jest „bardziej złożona" lub „mniej przypomina liczbę pierwszą" niż przeciętna liczba całkowita. A dokładnie (22) zachodzi wtedy, gdy odpowiedni punkt na krzywej z rys. 12 leży powyżej osi.

Rzut na oś poziomą odcinka łamanej odpowiadającego punktowi p ma długość $p^{-1}/\log\log N$, czyli długość czasu, jaką krzywa znajduje się powyżej 0, jest zasadniczo równa

$$\frac{1}{\log\log N}\sum\left[\frac{1}{p}\colon p \leqslant N \text{ i } \sum_{q\leqslant p}\delta_q(n) > \log\log p\right], \tag{23}$$

przy czym sumowanie przebiega po tych p, w których n jest ekscesywne.

Jeśli badamy podzielność n przez kolejne liczby pierwsze, pozostając $1/p$ jednostek czasu w punkcie p ($p \leqslant N$), wówczas (23) jest równe frakcji czasu, podczas którego mamy do czynienia z punktami p, w których n jest ekscesywna. Stosując (18) do zbioru z (7) otrzymujemy, że dla dużych N rozkład (23) pokrywa się w granicy z krzywą gęstości na rys. 5. Dla około 20% liczb całkowitych mniejszych od N wielkość (23) przewyższa 0,9, dla około 20% jest ona mniejsza niż 0,1 i zaledwie dla około 6% leży ona między 0,45 a 0,55.

Pod tym względem dzielniki pierwsze prezentują to samo dziwne zachowanie, co monety. Poniekąd są one nawet bardziej osobliwe. Bardziej naturalną niż (23) wielkością dla rozważań jest, być może,

$$\frac{1}{\pi(N)}\times\operatorname{card}\left[p\colon p \leqslant N \text{ i } \sum_{q\leqslant p}\delta_q(n) > \log\log p\right], \tag{24}$$

to jest *liczba* tych p, dla których n jest ekscesywne w punkcie p, unormowana dzieleniem przez $\pi(N)$ czyli przez ogólną ilość badanych liczb pierwszych. Dla dużych N, przeważająca część punktów załamań krzywej na rys. 12 znajduje się bardzo blisko 1, co oznacza, że w granicy rozkład (24) jest skupiony w 0 z masą $\frac{1}{2}$ i w 1 z masą $\frac{1}{2}$: Jeśli $\varepsilon > 0$ i N przekracza dostatecznie duże N_ε, wówczas (24) jest mniejsze od ε z prawdopodobieństwem leżącym w odcinku od $\frac{1}{2}-\varepsilon$ do $\frac{1}{2}+\varepsilon$, a jest większe niż $1-\varepsilon$ z prawdopodobieństwem leżącym w tej samej dziedzinie. Zatem praktycznie wszystkie liczby naturalne są ekscesywne albo dla praktycznie wszystkich liczb pierwszych, albo dla praktycznie żadnej.

Literatura cytowana

[1] P. Billingsley, *Convergence of probability measures*, Wiley, New York 1968.

[2] — *On the central limit theorem for the prime divisor function*, Amer. Math. Monthly 76 (1969), str. 132–139.

[3] W. Feller, *An introduction to probability theory and its applications*, t. I, wyd. 3, Wiley, New York 1968. (Tłum. polskie PWN, Warszawa 1960).

[4] D. Freedman, *Brownian motion and diffusion*, Holden-Day, San Francisco 1971.

[5] G. H. Hardy i E. M. Wright, *An introduction to the theory of numbers*, wyd. 4, Clarendon Press, Oxford 1960.

[6] M. Kac, *Statistical independence in probability, analysis and number theory*, Carus Math. Monogr. 12, Wiley, New York 1959. (Tłum. rosyjskie Izd. Inostr. Lit., Moskwa 1963.)

[7] G. Karlin, *A first course in stochastic processes*, Academic Press, New York 1966.

[8] J. Kubilius, *Verojatnostnye metody v teorii čisel*, Gos. Izd. Polit. i Naučnoj Liter. Litovskoj SSR, Wilno 1962. (Tłum. angielskie 1964 Amer. Math. Soc. Transl. Math. Monographs, Volume 11).

[9] W. Phillip, *Arithmetic functions and Brownian motion*, Proc. Symp. Pure Math. vol. 24, Amer. Math. Soc. 1973.