Jak odróżnić ciąg losowy od nielosowego?

Dr Rafał SZTENCEL

Słowo „losowy" będzie tu miało znaczenie „powstały w wyniku niezależnych rzutów symetryczną monetą". Przypatrzmy się poniższym trzem ciągom orłów i reszek:

(1) RRRORORROO OORRORRORR OOORRRRROO RORROROORO RRROROOROO OROORROOOR OROOROOROR RRORORROOO ORORRORROR ROROORORRO.

(2) RRORORORRR RORRORORRO RRORROOOOR ORROORORRR ORORORRORO OOORRORORR RRROROORRR OROOORROOR OROORORRRO ROOORORROR.

(3) OOORORRROR OROOOOROOO ORORORooRO RORORROROR OOORORROOO OOOOORRORR ROORRORROR OORRORRROR OOORROOOOR RROOORRRRO.

Dwa z nich zostały napisane przez człowieka, jeden powstał w wyniku komputerowej symulacji rzutów symetryczną monetą. Czy można go rozpoznać?

Istnieje zaskakująco prosty sposób. Znajdźmy w każdym ciągu najdłuższą serię jednakowych symboli. Pierwszy ciąg zawiera serię pięciu reszek (w trzeciej dziesiątce), drugi – czterech reszek (pierwsza i druga dziesiątka), wreszcie trzeci – aż ośmiu orłów (piąta i szósta dziesiątka). Otóż człowiek będzie miał pewne opory przed umieszczeniem w ciągu długiej serii orłów czy reszek. Natura takich oporów nie ma, bo nie ma pamięci. Dlatego całkiem niezły test losowości dla ciągów o długości 100 polega na uznaniu za losowe ciągów z najdłuższą serią liczącą 6 lub więcej symboli. Ciąg (3) zostanie wtedy słusznie zakwalifikowany jako losowy.

Podany sposób nie jest, oczywiście, doskonały. Może się zdarzyć, że ciąg losowy zostanie uznany za nielosowy i odwrotnie: ciąg nielosowy uznany za losowy. Jak zobaczymy, można nietrudno oszacować prawdopodobieństwo błędu pierwszego rodzaju. Z błędem drugiego rodzaju jest nieco gorzej – trzeba by znać probabilistyczną charakterystykę źródła. W każdym razie niechęć do pisania długich serii gwarantuje, że błąd drugiego rodzaju zdarza się rzadko.

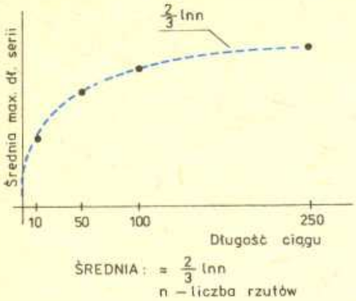
Teraz zastąpimy trudne rachunki symulacją komputerową. Jeśli ktoś wykona 10 000 eksperymentów, polegających na 100-krotnym rzucie monetą, to otrzyma pewnie wyniki zbliżone do poniższych:

maksymalna długość serii	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20
liczba eksperymentów	1	255	1658	2629	2221	1517	837	443	239	92	52	28	17	7	1	2	0	1

Widać teraz, że błąd pierwszego rodzaju zdarza się w około 19% przypadków (dla maksymalnej długości serii nie przekraczającej 5). Można by zmniejszyć o 1 krytyczną wartość długości serii – szansa błędu pierwszego rodzaju zmalałaby do 2,5%, ale wzrosłaby znacznie szansa błędu drugiego rodzaju.

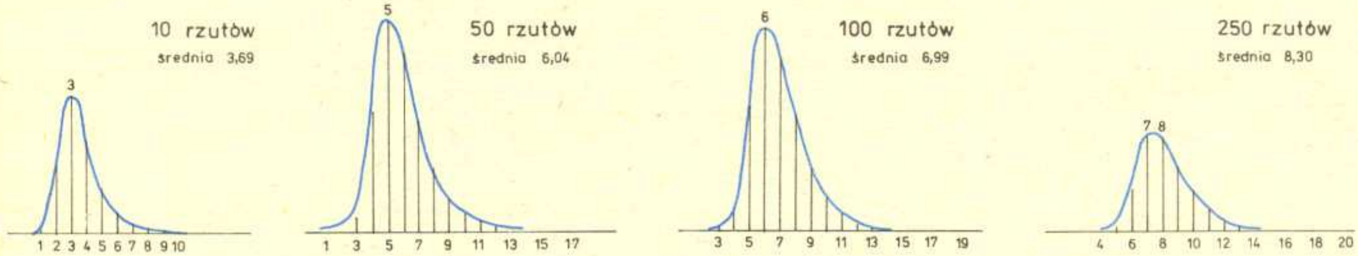
Popatrzmy teraz na sytuację bardziej ogólnie. Mówiąc uczenie, mamy hipotezę dotyczącą rozkładu prawdopodobieństwa na przestrzeni próbek. Mówi ona, że szanse otrzymania każdego z 2^{100} możliwych ciągów orłów i reszek są równe.

Eksperyment może obalić hipotezę, może też nie być podstaw do jej odrzucenia. Prowadzi to do podziału przestrzeni próbek na dwa zbiory. W naszym przypadku podział nastąpił ze względu na maksymalną długość serii; należało przy tym zadbać o to, by prawdopodobieństwo błędu nie było zbyt duże.



Jest teraz oczywiste, że testów losowości można wymyślić mnóstwo. Pierwszy z brzegu przykład to test oparty na liczbie serii. Za mała lub za duża liczba serii jest podejrzana i świadczy o braku losowości. Trzeba by ułożyć tabelę analogiczną do zamieszczonej wyżej (tu zresztą można się obejść bez komputera).

Interesujące byłoby zbadanie rozkładu maksymalnej długości serii w zależności od długości ciągu. Na rysunkach pokazano wyniki symulacji dla ciągów o długościach 10, 50, 100, 250. Widać, że średnie są z grubsza proporcjonalne do logarytmu długości ciągu. Jak to uzasadnić teoretycznie? A z jakim typem rozkładu prawdopodobieństwa mamy do czynienia?

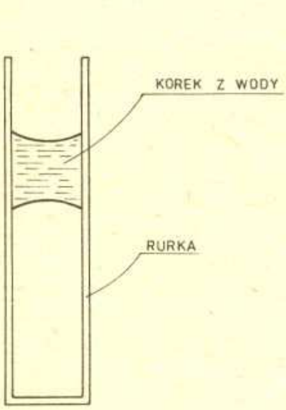


Wydział Fizyki Uniwersytetu Warszawskiego

KORESPONDENCYJNY KLUB FIZYKÓW

Drodzy Członkowie i Sympatycy Klubu!

Przypominamy, że co miesiąc przyznajemy nagrodę książkową dla autora najciekawiej opracowanego rozwiązania postawionych zagadnień. A oto nowa seria propozycji:



Zbadaj, jak zależy objętość gazu od ciśnienia, pod jakim gaz się znajduje, i od jego temperatury. Badania wykonaj ustalając temperaturę i zmieniając ciśnienie, a następnie ustalając ciśnienie i zmieniając temperaturę. W tym celu musisz zaopatrzyć się w szklaną rurkę o średnicy wewnętrznej kilku milimetrów. Jeden koniec rurki należy zatkać (korkiem, plasteliną itp.). Do drugiego końca wprowadź ostrożnie nieco wody, może być zabarwiona, tak, aby tworzyła ruchomy korek zamykający słup powietrza w rurce. Rurkę można ustawiać w różnych położeniach: słupek wody jest u góry i ciśnie na słup powietrza, słupek wody jest u dołu i zmniejsza ciśnienie wywierane na słup powietrza lub w położeniach pośrednich. Znając długość słupka wody i jego położenie możemy obliczyć, jakie ciśnienie panuje w rurce: ciśnienie atmosferyczne + ciśnienie słupa wody, ciśnienie atmosferyczne – ciśnienie słupa itd. Długość słupa powietrza pozwala wyznaczyć objętość powietrza zamkniętego ruchomym słupkiem wody. Dokonując pomiarów przy różnych położeniach rurki możemy wykreślić zależność objętości gazu od ciśnienia, pod jakim gaz się znajduje. Narysuj wykres. Jak nazywa się prawo, którego ilustracją jest otrzymany wykres?

Podobne pomiary możesz wykonać wstawiając rurkę do naczynia z wodą, do której wstawiono również termometr. Podgrzewaj wodę notując temperaturę i wysokość słupa powietrza. Powietrze jest pod stałym ciśnieniem. Będziesz mógł wykreślić zależność objętości gazu pod stałym ciśnieniem od temperatury. Jak nazywa się prawo, którego ilustracją jest otrzymany wykres?

Redaguje doc. dr Tomasz HOFMOKL

Listy prosimy przysyłać pod adresem:
Korespondencyjny Klub Fizyków
Wydział Fizyki Uniwersytetu Warszawskiego
ul. Hoża 69, 00-681 Warszawa.