MODELOWANIE RZECZYWISTOŚCI

Daniel Wójcik
Instytut Biologii Doświadczalnej PAN

d.wojcik@nencki.gov.pl

tel. 022 5892 424

http://www.neuroinf.pl/Members/danek/swps/

Podręcznik

Iwo Białynicki-Birula

Iwona Białynicka-Birula

ISBN: 83-7255-103-0
Data wydania: 6 maja 2002

wkrótce drugie wydanie, rozszerzone

Rzut monetą

Załóżmy, że moneta jest symetryczna. Wtedy prawdopodobieństwo wyrzucenia orła p_o i reszki p_r jest takie samo.

Przyjęto umowę, że suma prawdopodobieństw wynosi 1. Zatem

$$p_o=p_r=1/2$$

Co to jest prawdopodobieństwo

Prawdopodobieństwo:
liczba z przedziału od 0 do 1, przyporządkowana zdarzeniu przypadkowemu.

Liczba ta jest miarą szansy na to, że dane zdarzenie zajdzie.

Jak ustalić prawdopodobieństwo?

- na podstawie rozważań o symetrii
- na podstawie doświadczeń

definicja częstotliwościowa von Misesa:
prawdopodobieństwo p_A zajścia zdarzenia A określamy jako granicę

$$p_A=\lim_{N\to\infty}\frac{N_A}{N}$$

N_A – liczba zdarzeń A podczas przeprowadzenia N prób

Symetryczne rozkłady prawdopodobieństwa

- rzut monetą
- rzut kostką
- wyciągnięcie karty z potasowanej talii

Weryfikacja doświadczalna teorii

- Doświadczenia Buffona
  - 4040 rzutów
  - 2048 razy reszka, 1992 orzełek
  - $p_r = 0.5069$
- Doświadczenia Romanowskiego
  - 80640 rzutów
  - $p_r = 0.4923$
- Program **Buffon**

Chłopcy i dziewczęta

- W badaniach demograficznych stwierdzono, że stosunek liczby urodzeń chłopców do wszystkich urodzeń jest liczbą stałą i wynosi około 22/43 ≈ 0,5116
- Kiedy Laplace odkrył, że w Paryżu w latach 1745-1784 liczba ta spadła do 25/49 ≈ 0,5102 uznał, że coś nadzwyczajnego musiało się wydarzyć.
- Okazało się, że ubodzy mieszkańcy okolic Paryża chętniej podrzucali do stolicy dziewczęta niż chłopców

Definicja prawdopodobieństwa Laplace'a

- Prawdopodobieństwo P(A) jest równe stosunkowi liczby przypadków sprzyjających wystąpieniu zdarzenia A do wszystkich możliwych przypadków
- Jak na podstawie genetyki otrzymać stosunek liczby urodzeń chłopców do liczby wszystkich urodzeń?

Rzut dwoma kostkami

- Jakie jest prawdopodobieństwo wyrzucenia parzystej liczby oczek?

Definicja prawdopodobieństwa Laplace'a

- Prawdopodobieństwo P(A) jest równe stosunkowi liczby przypadków sprzyjających wystąpieniu zdarzenia A do wszystkich możliwych przypadków

- Jak na podstawie genetyki otrzymać stosunek liczby urodzeń chłopców do liczby wszystkich urodzeń?

Statystyczne własności liczb rzeczywistych

Wybierzmy dowolną liczbę rzeczywistą.

Co można powiedzieć o rozkładzie cyfr w zapisie dziesiętnym (dwójkowym, innym) tej liczby?

Przykład: liczby wymierne

$1/5 = 0.2_{10} = 0.001100110011..._2$

$1/7 = 0.142857142857..._{10} = 0.001001..._2$

Statystyczne własności liczb rzeczywistych

- **Przykład**: liczby niewymierne
  - 0.121121112111121111121111112...
  - e = 2,71828182845904523536028747135 27...
  - π = 3,14159265358979323846264338327 95...
  - $\sqrt{2}$ = 1,41421356237309504880168872420 97...
- prawdopodobieństwa wystąpienia cyfr
- prawdopodobieństwa wystąpienia par cyfr
- Program **Poe** i rozwinięcia liczb

Statystyczne własności liczb rzeczywistych

- **Przykład**: Binarna reprezentacja liczby

π = 3,1415926535897932384626433 83...

jest początkowo zdominowana przez zera:
125 zer w pierwszych 204 znakach!

To daje odchylenie prawie 23%
od wartości średniej.

Dopiero po 26 596 znakach
liczby zer i jedynek zrównują się!

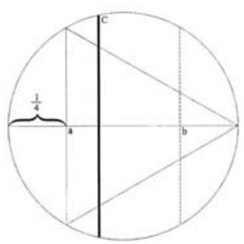
Nieskończona liczba przypadków

- Jakie jest prawdopodobieństwo zastania na stacji stojącego pociągu, jeżeli wiemy, że pociągi wjeżdżają na stację co 10 minut i stoją na niej minutę?
- Paradoks Bertranda:

  jakie jest prawdopodobieństwo tego, że na chybił trafił wybrana cięciwa koła będzie dłuższa od ramienia trójkąta równobocznego wpisanego w to koło?
- Program **Bertrand**

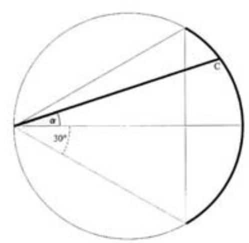
Paradoks Bertranda – rozwiązanie 1

Ustalmy kierunek cięciwy,
np. pionowy. Przesuwając
cięciwę od lewa do prawa
widzimy, że tylko pomiędzy
punktami *a* i *b* długość cięciwy
jest większa od połowy
średnicy. Długość *ab* jest równa
połowie średnicy, zatem
prawdopodobieństwo jest 1/2.



Paradoks Bertranda – rozwiązanie 2

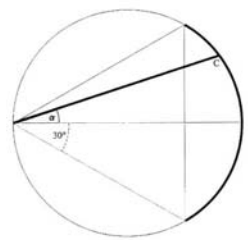
Rozważmy cięciwę
zaczepioną w jednym punkcie.
Zmieniając jej kąt nachylenia
względem średnicy
od -90 do +90 stopni dostajemy
wszystkie możliwe cięciwy.
Te z nich, które ze średnicą
tworzą kąt od -30 do +30 stopni
są dłuższe od połowy średnicy.
Zatem szukane prawdopodobieństwo
wynosi 60/180 = 1/3.



Paradoks Bertranda – rozwiązanie 3

Wybierzmy losowo
dwa punkty na okręgu
łącząc je cięciwą.
Pierwszy punkt jest dowolny,
drugi musi leżeć
na jednej trzeciej okręgu
naprzeciwko pierwszego punktu,
żeby cięciwa była
odpowiednio długa.

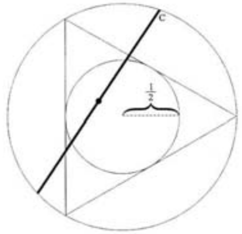
Zatem prawdopodobieństwo
wynosi również 1/3.



Paradoks Bertranda – rozwiązanie 4

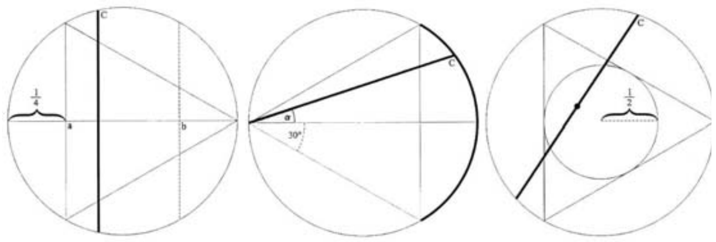
Środek każdej cięciwy leży wewnątrz koła i wyznacza jednoznacznie jej położenie.

Ta część koła, w której leżą środki cięciw spełniających zadany warunek, jest również kołem o promieniu równym połowie długości promienia dużego koła.



Stosunek pola tej części, do pola całego koła wynosi 1/4.

Paradoks Bertranda – podsumowanie



Które rozwiązanie jest poprawne?

Metoda Monte Carlo

- Metoda Monte Carlo polega na wykonaniu wielu eksperymentów losowych w celu oszacowania wyniku.
- Program **Ulam:**

pole prostokąta: funkcja stała
pole trójkąta: x
pole koła i liczba π: $\sqrt{(1-x^2)}$

Ciągi liczb losowych

- Czy ciąg 0,1,0,1,0,1,... jest losowy?
- Napisz na kartce ciąg zer i jedynek o długości 100 znaków
- Policz ile jest w tym ciągu podciągów złożonych z trzech, czterech i pięciu jedynek
- Porównaj z wynikami wygenerowanymi przez program **Bernoulli**

Prawdopodobieństwo wylosowania ciągów samych jedynek

- Aby uzyskać samotną jedynkę w binarnym ciągu trzeba wyrzucić po kolei 0,1,0 – prawdopodobieństwo ½ * ½ * ½ = 1/8
- Aby uzyskać n jedynek w binarnym ciągu trzeba wyrzucić po kolei 0,1, ... ,1,0 – to daje prawdopodobieństwo

$$\frac{1}{2}*\underbrace{\frac{1}{2}*...*\frac{1}{2}}_{n\ razy}*\frac{1}{2}=\frac{1}{2}^{(n+2)}$$

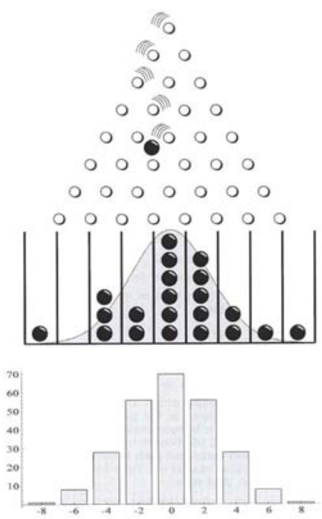
- Zatem prawdopodobieństwo wyrzucenia 6,7,8 jedynek pod rząd to odpowiednio 1/256, 1/512 i 1/1024

Prawdopodobieństwo wylosowania ciągów samych jedynek

Jeżeli prawdopodobieństwo wyrzucenia 8 jedynek pod rząd to 1/1024, ile razy trzeba rzucić monetą, żeby średnio zaobserwować jeden ciąg 8 jedynek w serii?

Deska Galtona

Pochylona deska z wbitymi gwoździami ułożonymi w trójkąt. Można jej użyć do wizualizacji wielokrotnego rzucania monetą



Deska Galtona

- Jeżeli prawdopodobieństwo skoku w prawo lub w lewo na każdym gwoździu jest takie samo, to prawdopodobieństwo rozkładu na ostatnim poziomie dane jest przez trójkąt Pascala
- Program **Galton**

n \ s	-9	-8	-7	-6	-5	-4	-3	-2	-1	0	1	2	3	4	5	6	7	8	9
0										1									
1									1		1								
2								1		2		1							
3							1		3		3		1						
4						1		4		6		4		1					
5					1		5		10		10		5		1				
6				1		6		15		20		15		6		1			
7			1		7		21		35		35		21		7		1		
8		1		8		28		56		70		56		28		8		1	
9	1		9		36		84		126		126		84		36		9		1

Trójkąt Pascala

Trójkąt Pascala powstaje przy obliczaniu n-tej potęgi dwumianu. Każdy współczynnik w trójkącie Pascala równy jest liczbie dróg jakimi można do niego dojść

n \ s	-9	-8	-7	-6	-5	-4	-3	-2	-1	0	1	2	3	4	5	6	7	8	9
0										1									
1									1		1								
2								1		2		1							
3							1		3		3		1						
4						1		4		6		4		1					
5					1		5		10		10		5		1				
6				1		6		15		20		15		6		1			
7			1		7		21		35		35		21		7		1		
8		1		8		28		56		70		56		28		8		1	
9	1		9		36		84		126		126		84		36		9		1

Współczynniki dwumianu

- Liczby w n-tym wierszu trójkąta Pascala, zwane współczynnikami dwumianowymi, oznaczane są symbolem Newtona i dane są wzorem:

$$\binom{n}{k}=\frac{n!}{k!(n-k)!}$$

- Zatem prawdopodobieństwa trafienia do odpowiedniej przegródki wynoszą

$$p(n,k)=\frac{n!}{2^n k!(n-k)!}$$

- Zachodzi

$$\sum_{k=0}^{n} p(n,k)=1$$

Błądzenie przypadkowe

- Ruch punktu na prostej lub w przestrzeni o dowolnym wymiarze polegający na wykonywaniu losowych kroków o stałej długości w jednym z kilku wybranych kierunków.
- Deska Galtona jest równoważna błądzeniu przypadkowemu na prostej.
- Błądzenie przypadkowe jest modelem ruchu cząstki Browna

Spacery losowe – model dyfuzji

- Błądzenie przypadkowe (spacer losowy) w przestrzeni stanowi model dyfuzji i ruchów Browna – rozprzestrzeniania się cząsteczek w danym środowisku.
- Średnia odległość od punktu początkowego rośnie z czasem t jak $\sqrt{t}$
- Program **Smoluchowski**

Wzór Stirlinga

- Prawdopodobieństwo powrotu do miejsca startu wynosi

$$p(2n,n)=\frac{(2n)!}{2^{2n}n!n!}$$

- Dla n = 2,4,6,8,10,12... otrzymujemy 1/2, 3/8, 5/16, 35/128, 63/256, 231/1024...
- Ze wzoru Stirlinga

$$n!\approx\sqrt{2\pi n}n^n e^{-n}$$

otrzymujemy przybliżenie:

$$p(2n,n)\approx\frac{\sqrt{2}}{\sqrt{\pi n}}$$

Asymptotyka rozkładu dwumianowego

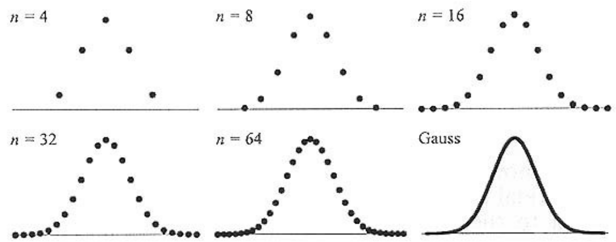
- Widzimy, że prawdopodobieństwo powrotu cząstki do punktu startu maleje do 0
- Zauważmy, że także liczba możliwych wyników rośnie. Żeby badać zachowanie rozkładu dla dużych *n* musimy go znormalizować.
- Mnożąc prawdopodobieństwa przez $\sqrt{n/2}$ i tak samo zwiększając skalę na osi *x* otrzymujemy dla dużych *n* rozkład Gaussa:

$$p(x)=\sqrt{1/\pi}\exp(-x^2)$$

Asymptotyka rozkładu dwumianowego

- Mnożąc prawdopodobieństwa przez $\sqrt{n/2}$ i tak samo zwiększając skalę na osi *x* otrzymujemy dla dużych *n* rozkład Gaussa:

$$p(x)=\sqrt{1/\pi}\exp(-x^2)$$



Generatory liczb losowych

- Komputery generują liczby pseudolosowe, nie losowe.

- Wygenerowane liczby powtarzają się po pewnym czasie. Im dłuższy okres, tym lepszy generator. Im lepieje „potasowane” liczby, tym lepszy generator.

Inne źródła liczb losowych

- Tabele liczb losowych, np. rozwinięcia liczb normalnych

- Pomiary fizyczne odpowiednich układów

- Do czego potrzebujemy losowych ciągów? Do każdego praktycznego zastosowania teorii prawdopodobieństwa:
  - Wybór próbki statystycznej
  - Gry matematyczne (np. negocjacje)
  - Obliczenia metodą Monte Carlo

Liczby normalne ponownie

- Liczba normalna – liczba, w której rozwinięciu w danym układzie każdy blok cyfr jest tak samo prawdopodobny jak każdy inny blok tej samej długości

- Rozwijają liczbę normalną w bazie o podstawie 36 możemy generować losowe teksty: 10 cyfr + 26 liter łacińskich, albo 35 polskich liter i spacja.

- Przykład rozwinięcia liczby pi po polsku.

Prawdopodobieństwo wystąpienia dowolnego tekstu

- Trzydziestotomowa encyklopedia Brytannica zawiera około 30 milionów znaków. Prawdopodobieństwo wystąpienia jej tekstu w przypadkowym tekście wynosi

$$\left(\frac{1}{36}\right)^{30\,000\,000} \approx \left(\frac{1}{10}\right)^{45\,000\,000}$$

- Prawdopodobieństwo wystąpienia krótkich słów, jak „Budda" wynosi

$$\left(\frac{1}{36}\right)^{5} \approx \left(\frac{1}{60\,000\,000}\right)$$

czyli w tekście długości 60 mln znaków oba te słowa powinny wystąpić przynajmniej raz

Kodowanie tekstów w rozwinięciach

- Znajdowanie tekstów w rozwinięciach liczb normalnych nie tylko zależy od podstawy ale i od kodowania, tj. od tego co przypiszemy danemu symbolowi w rozwinięciu

Grafy i sieci

Mosty Królewca: własności topologiczne

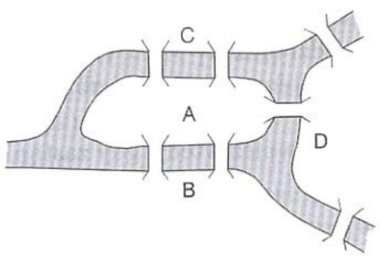
Problem komiwojażera: własności metryczne

Języki naturalne:
opis obiektów nieskończonych

Małe światy i inne sieci

Mosty Królewca

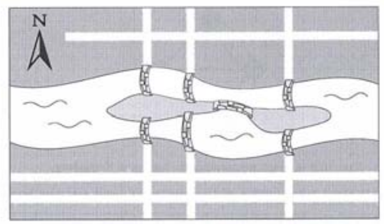
- Czy można przejść Królewiec przechodząc każdy most dokładnie raz?



Mosty Królewca

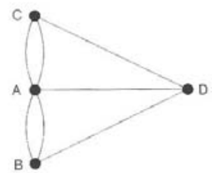
- Czy można usunąć jeden most tak by dało się przejść Królewiec przechodząc każdy most dokładnie raz?

- Czy z każdego miejsca można wtedy wykonać pełen spacer?

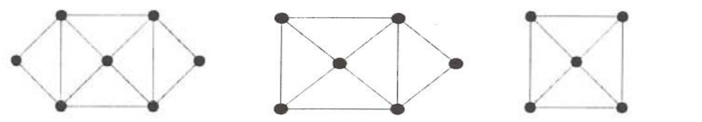


Graf reprezentujący mosty Królewca

- Potrzebujemy informacji o połączeniach poszczególnych obszarów lądu
- Własności obiektu, które nie zmieniają się przy jego odkształceniach (bez rozrywania) to **własności topologiczne**
- Mosty Królewca można reprezentować przy pomocy **grafu**
- **Cykl Eulera**

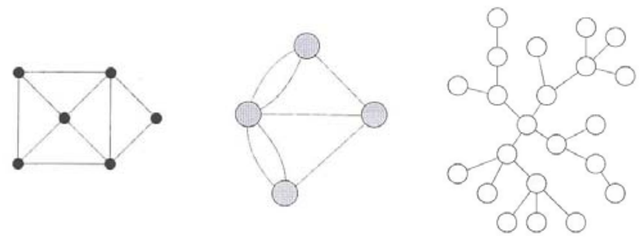


Grafy eulerowskie i półeulerowskie



Graf

- Punkty oznaczające obszary lądu to **wierzchołki grafu**
- Połączenia między obszarami (mosty) to **krawędzie grafu**
- Minimalny model grafu składa się z wierzchołków i krawędzi.

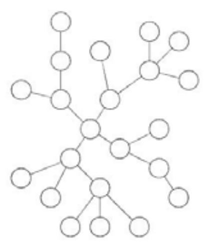


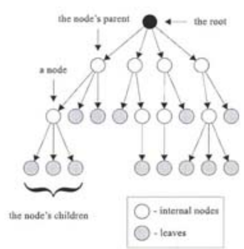
Inne typy grafów

- Minimalny model grafu można rozbudować:
  - Przypisując każdemu z wierzchołków lub krawędzi etykietę (napis) lub wagę (liczbę rzeczywistą)
  - Ustalając kierunek krawędzi (np. ulice jednokierunkowe) – graf skierowany
  - Drzewa to spójne (= niepodzielne bez przecięcia krawędzi) grafy nie zawierające cykli (= zamkniętych dróg)

Przykład drzewa

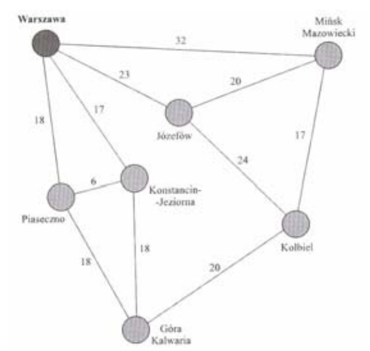
Często wyróżniamy jeden z węzłów drzewa nazywany **korzeniem**. Wtedy każdy węzeł w drzewie ma dokładnie jednego sąsiada, który jest bliżej korzenia od niego. Ten sąsiad nazywa się jego **rodzicem**, pozostali sąsiedzi to **dzieci**. Węzeł bezdzietny nazywa się **liściem**.





Problem komiwojażera

- Która droga jest najlepsza (najkrótsza)?
- minimalny cykl Hamiltona

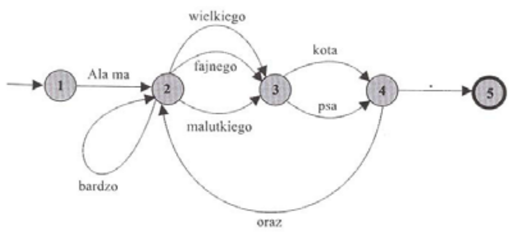


Problem komiwojażera

- Nie istnieje efektywny algorytm rozwiązujący bezbłędnie problem komiwojażera dla dużych grafów w krótkim czasie
- **Problem NP-zupełny**
- Istnieją za to efektywne algorytmy heurystyczne, które szybko przybliżają najkrótszą drogę

Języki naturalne

- Graf generujący podzbiór zdań języka polskiego
- Inne przykłady automatów skończonych (np. sieci boolowskie)



Sieci to też grafy

- Przykłady sieci: Internet, ekosystem, społeczność, znajomi,sieć energetyczna, itd.
- Struktura sieci
- Ewolucja sieci
- Ewolucja na sieci