Rachunek prawdopodobieństwa i statystyka

Kurs dla kierunku
Informatyki stosowanej
Uniwersytet Jagielloński
Kraków, 2013/2014

Dr hab. Roman Skibiński

UWAGA:

Slajdy nie zawierają całości materiału przedstawianego na wykładach;
mają jedynie charakter pomocniczy i podsumowujący.

http://koza.if.uj.edu.pl/~rpis/

Warunki zaliczenia

- Ćwiczenia: dwa kolokwia, średnia ≥ 3.0
- Wykład: pięć kartkówek na ćwiczeniach, średnia z czterech najlepszych wyników ≥ 3.0

  Przed kartkówkami (niezapowiedzianymi) dostępne będą zagadnienia do przygotowania.
- Ocena końcowa (do indeksu, z wykładu):

  (2/3*ocena z ćwiczeń + 1/3*ocena z wykładu)*0.9+ 0.5 za programy, min 2.7
- Kolokwium poprawkowe (z wykładu i z ćwiczeń): pisane tylko z części materiału do poprawy

Literatura

- W.Krysicki, J.Bartos i inni, „Rachunek prawdopodobieństwa i statystyka matematyczna w zadaniach" tomy 1 i 2, PWN 2005

Literatura dodatkowa:
- J.Jakubowski, R.Sztencel „Rachunek prawdopodobieństwa dla (prawie) każdego",SCRIPT, W-wa 2006.
- A.Plucińska, E.Pluciński „Probabilistyka", WNT 2000.
- S.Brandt „Analiza danych", PWN (od 1999)
- R.Nowak „Statystyka dla fizyków" PWN 2002.
- V.Rohatgi, „Statistical inference",J.Wiley&Sons, Inc, 1984.

Definiujemy sztukę przewidywania, inaczej sztukę stochastyki, jako sztukę oceniania z największą możliwą dokładnością prawdopodobieństwa zdarzeń, tak żebyśmy w naszych osądach i działaniach zawsze mogli opierać się na tym, co okazało się najlepsze, najodpowiedniejsze, najpewniejsze, najsensowniejsze; jest to jedyny cel mądrości filozofa i roztropności męża stanu.

J.Bernoulli „Ars Conjectandi"
(„Sztuka przewidywania") 1713

Za I.Stewart „Oswajanie nieskończoności. Historia matematyki"

Zakres wykładu

- **Rachunek prawdopodobieństwa** – jak liczyć prawdopodobieństwa zdarzeń i jak je globalnie opisywać.
- **Statystyka matematyczna** – jak wnioskować w sytuacjach, gdy mamy niepełną informację (wniosek o całej grupie na podstawie informacji zebranej na części grupy, np. sondaże przedwyborcze), jak oceniać wiarygodność takiego wnioskowania (hipotezy statystyczne)

Dlaczego ?

- Ma wpływ na nasze życie (gry hazardowe, ubezpieczenia, handel, kryminalistyka, medycyna, polityka)
- Zastosowania w informatyce:
  - symulacje komputerowe
  - metody obliczeniowe
  - modelowanie rzeczywistości (grafika)
  - probabilistyczna (statystyczna) analiza algorytmów
  - algorytmy probabilistyczne

Rozgrzewka

Sprawdzamy równość wielomianów $F(x) \overset{?}{=} G(x)$ o stopniu d w liczbach całkowitych

$$F(x) = (x - x_1)(x - x_2)(x - x_3)...(x - x_d)$$

$$G(x) = a_0 + a_1 x + a_2 x^2 + ... + a_d x^d$$

Mają one co najwyżej d pierwiastków.

Rozgrzewka

Weźmy liczbę losową r całkowitą z przedziału [0,100d) i sprawdzamy czy $F(r) = G(r)$

Nie: wielomiany na pewno są różne (dobra odpowiedź)

Tak: przyjmujemy, że wielomiany są równe

w rzeczywistości albo są równe (i wtedy mamy dobrą odpowiedź), albo trafiliśmy na wspólny pierwiastek (i wtedy mamy złą odpowiedź).

Maksymalnie wspólnych pierwiastków jest d; Szansa, że trafiliśmy w taki jest d/(100*d)=0.01=1%

Czyli w 99% dostaniemy poprawną odpowiedź.

Rozgrzewka

- Można to poprawić: większy zakres [0,1000d)
  (ale możemy mieć problem z dużymi liczbami)
- Można też powtarzać całe sprawdzanie kilka razy

Tak można sprawdzać mnożenie macierzy

A=BC $\rightarrow$ w_1=Ar w_2=B(Cr) w_1=w_2

Liczba operacji: n^3+n^2 $\rightarrow$ $3n^2+n$

Historia

- Starożytność, Średniowiecze – gry losowe
- XVI w G.Cardano (1501-1576), „Księga o grach losowych" – podstawy prawdopodobieństwa (gry w kości i w karty, dodatkowo rozdział o skutecznym oszukiwaniu)
- A.Gombaud (Chevalier de Méré, 1607-1684) – korespondencja pomiędzy B.Pascalem (1623-1662; 1654,1655 – trójkąt Pascala) a P.de Fermat (1601-1665), problem podziału puli przy przerwaniu gry („problem of points")

Rozwiązanie biorące pod uwagę tylko dotychczasowe wyniki jest błędne, należy uwzględnić możliwe zdarzenia do zaplanowanego końca gry.

Historia

- w różnych odmianach istnieje paradoks de Méré, np. dlaczego częściej wypada „6" w 4-ech rzutach jedną kostką niż dwie „6" w 24-ech rzutach dwoma kostkami.

Rzut jedną kostką: 4*1/6=4/6

Rzut dwoma kostkami: 24*1/6*1/6=24/36=4/6

Ale naprawdę interesuje nas prawdopodobieństwo otrzymania przynajmniej raz szóstki w 4-ech rzutach = 1-prawdopodobieństwo nie otrzymania żadnej szóstki=1-(5/6)^4=671/1296=0.5177 i prawdopodobieństwo otrzymania przynajmniej raz dwóch szóstek w 24-ech rzutach dwoma kostkami = 1-prawdopodobieństwo nie otrzymania dwóch szóstek ani razu)=1-(35/36)^24=0.4914

Pokazuje to, że:

1) trzeba precyzyjnie definiować czego prawdopodobieństwo liczymy
2) de Méré dużo czasu spędzał grając w kości

Historia

- Ch.Huygens(1629-1695), J.Bernoulli (1654-1705; 1713 – „Sztuka przewidywania", białe i czarne kamyki w urnie),
- problemy typu „rzut uczciwą monetą". Ale co to znaczy „uczciwa moneta"?
- T.Bayes (1701-1761): analiza bayesowska
- P.Laplace(1749-1827), K.Gauss(1777-1855)
- teoria miary
- A.Quetelet (1796-1874); 1835 statystyka społeczeństwa
- F.Galton (1822-1911); 1865 dziedziczenie, regresja
- XX w: A.N.Kołomogorow (1903-1987): nowoczesna tzw. aksjomatyczna teoria prawdopodobieństwa
- Wykorzystanie komputerów

Pojęcia wstępne

- **Eksperyment deterministyczny** – warunki wyznaczają wynik (np. tylko białe kule w urnie)
- **Eksperyment przypadkowy (zdarzenie losowe)** to taki eksperyment, którego wyniku nie potrafimy przewidzieć, mimo, że powtarzamy go w takich samych warunkach.

  (np. białe i czarne kule w urnie)

  Jedyne co możemy zrobić to zebrać możliwe wyniki i określić ich prawdopodobieństwo.

Definicja częstościowa

Powtarzamy eksperyment *n* razy

$N_k(n)$ – liczba wystąpienia wyniku *k* w *n* eksperymentach

$f_k(n)$ – względna częstość wyniku k $f_k(n)=\frac{N_k(n)}{n}$

spełnia z def. $0 \le f_k(n) \le 1$

$$\sum_k f_k(n) = 1$$

Częstościowa definicja prawdopodobieństwa:

$$P(k) \equiv \lim_{n\to\infty} f_k(n)$$

-trudne w praktyce (nieskończona liczba eksperymentów, powtarzalność doświadczeń, definicja eksperymentu (prawdopodobieństwo urodzenia dziewczynki/chłopca),

-wynika z aksjomatycznej teorii prawdopodobieństwa,

-przykład: aplet „Fałszywa kostka"

Pojęcia wstępne

- **Przestrzeń próbek eksperymentu przypadkowego** to zbiór Ω wszystkich możliwych wyników tego eksperymentu.
- **Zdarzenie elementarne** – każdy możliwy wynik eksperymentu przypadkowego.

  Powtarzając eksperyment przypadkowy jako wynik otrzymujemy jedno i tylko jedno zdarzenie elementarne; zdarzenia elementarne wykluczają się wzajemnie
- **Zdarzenie** to podzbiór przestrzeni próbek.

Pojęcia wstępne

- Przestrzeń próbek eksperymentu przypadkowego może być

  -skończona (liczba oczek w rzucie kostką),

  -nieskończona w sposób przeliczany (ilość rzutów aż wypadnie „6"),

  -nieskończona w sposób nieprzeliczalny (odległość na jaką rzucimy kostkę)
- inny podział: dyskretna i ciągła
- mogą istnieć typy mieszane
- Szczególne zdarzenia:

  - **zdarzenie niemożliwe** – pusty podzbiór przestrzeni Ω

  - **zdarzenie pewne** – cała przestrzeń Ω

Działania na zbiorach (przypomnienie)

- Operacje na zbiorach
  (= podzbiorach przestrzeni próbek = zdarzeniach)

Suma $A \cup B: \quad x \in A \vee x \in B$

Iloczyn $A \cap B: \quad x \in A \wedge x \in B$

Dopełnienie $\bar{A}: \quad A + \bar{A} = \Omega$

Rozłączność $A \cap B = \phi$

Zawartość $A \subset B: \quad x \in A \Rightarrow x \in B$

Równość $A = B: \quad A \subset B \wedge B \subset A$

Własności działań na zdarzeniach:

Przemienność $A \cup B = B \cup A \qquad A \cap B = B \cap A$

Łączność $A \cup (B \cup C) = (A \cup B) \cup C \qquad A \cap (B \cap C) = (A \cap B) \cap C$

Dystrybutywność (rozłączność sumy i iloczynu)

$$A \cup (B \cap C) = (A \cup B) \cap (A \cup C) \qquad A \cap (B \cup C) = (A \cap B) \cup (A \cap C)$$

Prawa de'Morgana

$$\overline{A \cap B} = \bar{A} \cup \bar{B} \qquad \overline{A \cup B} = \bar{A} \cap \bar{B}$$

Aksjomatyczna definicja prawdopodobieństwa

Każdemu zdarzeniu A w przestrzeni próbek Ω przyporządkowujemy liczbę rzeczywistą *P(A)* zwaną prawdopodobieństwem, tak by miała ona następujące własności:

I: $\forall A \subset \Omega \quad P(A) \geq 0$

II: $P(\Omega)=1$

III: Jeżeli $A_1, A_2, \ldots$ jest ciągiem rozłącznych zdarzeń to $P(\bigcup_k A_k) = \sum_k P(A_k)$

Wnioski z aksjomatów

1. $P(\overline{A}) = 1 - P(A)$
2. $P(A) \leq 1$
3. $P(\phi) = 0$
4. $A \subset B \Rightarrow P(A) \leq P(B)$
5. Dla dwóch zdarzeń rozłącznych:

$$P(A \cup B) = P(A) + P(B)$$

6. Dla dwóch dowolnych zdarzeń:

$$P(A \cup B) = P(A) + P(B) - P(A \cap B)$$