Rachunek prawdopodobieństwa i statystyka

Kurs dla kierunku
Informatyki stosowanej
Uniwersytet Jagielloński
Kraków, 2020/2021

Dr hab. Roman Skibiński, prof. UJ

UWAGA:

Slajdy zawierają (prawie) całość materiału przedstawianego na wykładach;
Powinny być uzupełnione wiadomościami z ćwiczeń
http://koza.if.uj.edu.pl/~rpis/
http://koza.if.uj.edu.pl/~aplety/

Warunki zaliczenia (1)

- Ćwiczenia: ocena będzie średnią z trzech ocen cząstkowych (w skali 2.0-5.0): dwóch z kolokwiów z zadań i oceny z aktywności na ćwiczeniach. Średnia ta zostanie przeliczona na ocenę z ćwiczeń, wpisywaną do indeksu, w następujący sposób:
  [2.66-3.20] – ocena dostateczna,
  (3.20-3.70] – ocena dostateczna plus,
  (3.70-4.20] – ocena dobra,
  (4.20-4.50] – ocena dobra plus,
  (4.51-5.0] – ocena bardzo dobra.
- Warunkiem zaliczenia ćwiczeń jest zaliczenie obu kolokwiów co najmniej na ocenę dostateczną.
- Niezaliczone kolokwia będzie można poprawiać na kolokwium zaliczeniowym, poprawa dotyczy tylko niezaliczonej części materiału. W przypadku poprawiania jednego kolokwium ocena z ćwiczeń będzie liczona jako średnia z czterech ocen (2 kolokwia + aktywność + kolokwium poprawkowe) przeliczona jak wyżej. Analogicznie w przypadku poprawiania obu kolokwiów średnia będzie liczona z pięciu ocen.

Warunki zaliczenia (2)

- Wykład: cztery kartkówki na ćwiczeniach, oceniane w skali 0-5 pkt. Suma punktów prowadzi do oceny:
  [12-13.5) – ocena dostateczna,
  [13.5-15.0) – ocena dostateczna plus,
  [15.0-16.5) – ocena dobra,
  [16.5-18.0) – ocena dobra plus,
  [18.0-20.0] – ocena bardzo dobra.
  - Przed kartkówkami (niezapowiedzianymi) dostępne będą zagadnienia do przygotowania
  - W trakcie semestru można poprawić jedną, najgorzej napisaną kartkówkę + te z usprawiedliwionymi nieobecnościami
  - Osoby, które nie uzyskają 12 punktów piszą kolokwium poprawkowe z całości wykładu
- **Ocena końcowa (wpisywana do indeksu, z wykładu):**
  (2/3*ocena z ćwiczeń + 1/3*ocena z wykładu)*0.9
  Powyższy algorytm prowadzi do oceny wpisywanej do indeksu:
  do 2.70 – ocena niedostateczna,
  [2.70-3.15) – ocena dostateczna,
  [3.15-3.60) – ocena dostateczna plus,
  [3.60-4.05) – ocena dobra,
  [4.05-4.5] – ocena dobra plus,

  Osoby, które rozwiążą zadania programistyczne będą miały ocenę podwyższoną o pół stopnia np. z dobrej plus na bardzo dobrą. **Zatem warunkiem otrzymania oceny bardzo dobrej w indeksie jest napisanie zadanych programów.**

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)

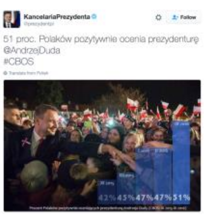
RPiS 2020/2021 7

Dlaczego ?

- Ma wpływ na nasze życie (gry hazardowe, ubezpieczenia, handel, kryminalistyka, medycyna, polityka, manipulacje w ikonografikach)
- Zastosowania w informatyce:
  - symulacje komputerowe
  - metody obliczeniowe
  - modelowanie rzeczywistości (grafika)
  - probabilistyczna (statystyczna) analiza algorytmów
  - algorytmy probabilistyczne (w tym komputer kwantowy)
  - systemy kolejkowe
  - eksploracja danych
  - układy z szumem (np. rozpoznawanie mowy)

RPiS 2020/2021 8

Rozgrzewka 1

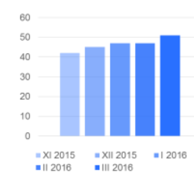


https://twitter.com/prezydentpl/status/710903898567999488
Zaczerpnięte z blogu smarterpoland.pl

RPiS 2020/2021 9

Rozgrzewka 1

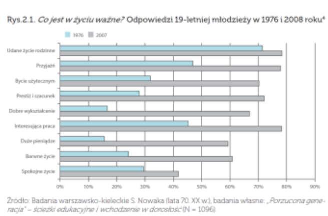
1. Stosunek długości słupków: 1:3.7, stosunek poparcia w procentach 1:1.21. Lie-factor: ~12.6.
2. Słupki nie oddają nawet różnic pomiędzy prezentowanymi liczbami. Trzypunkowa odległość pomiędzy 42% a 45% jest z 5x mniejsza niż 4 punktowa odległość pomiędzy 47% a 51%.
3. Różna jasność słupków może wpływać na odbieraną wielkość prezentowanych liczb.



https://twitter.com/prezydentpl/status/710903898567999488
Zaczerpnięte z blogu smarterpoland.pl

RPiS 2020/2021 10

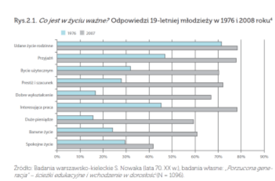
Rozgrzewka 2



Zaczerpnięte z blogu smarterpoland.pl

RPiS 2020/2021 11

Rozgrzewka 2

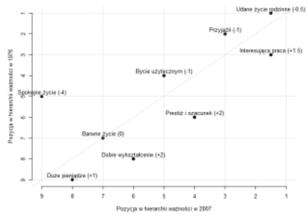


- *„Najprostsze porównywalne dane ukazują niezmiennie wysoką w hierarchii ważności pozycję rodziny – udane życie rodzinne jest podkreślane jako sprawa bardzo ważna zarówno przez dawne, jak i przez nowe młode pokolenie (przez nowe nawet bardziej). ….. Drugie podobieństwo dotyczy relatywnie niskiego wartościowania spokojnego życia. W innych kwestiach charakterystyki dawnej i nowej młodzieży wyraźnie się rozchodzą."*
- A teraz wniosek z prezentacji
  „Widoczna jest mentalna odrębność dzisiejszego młodego pokolenia"

Zaczerpnięte z blogu smarterpoland.pl

RPiS 2020/2021 12

Rozgrzewka 2



Poprawne wnioski:
- Hierarchia ważności jest w przybliżeniu zachowana
- Największa zmiana dotyczy „Spokojnego życia"

Zaczerpnięte z blogu smarterpoland.pl

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 – nowe możliwości i nowe zadania

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 $\overline{A}: \quad A + \overline{A} = \Omega$

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

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