WYKŁADOWCA

- Wiesława Wawrzyniak-Kosz
- www.prac.im.pwr.wroc.pl/~wkosz

ZALICZENIE

- 2 kolosy x 3 zadania x 5pkt
- min. 15pkt na zaliczenie
- max 34 pkt

- 1 kolos: 17.04
- 2 kolos: 12.06

- max 20% nieobecności
- max 4 pkt za aktywność

- poprawka w sesji

RACHUNEK PRAWDOPODOBIEŃSTWA

Temat wykładu: Prawdopodobieństwo

PRAWDOPODOBIEŃSTWO

- doświadczenie losowe – doświadczenie /zjawisko /eksperyment, które kończy się jakimś wynikiem, dopóki się nie zakończy nie można podać wyniku
    - jak trzymam monetę i puszczam, to nie dośw. losowe
    - awers vs. rewers → dośw. losowe
    - trasa dojazdu to losowe doświadczenie
    - wszystkie czasy pracy
- stabilność częstości – jeśli coś powtarzamy wiele razy, to relacja wystąpienia danej wartości do liczby powtórzeń, dążąc do nieskończoności, zbliża się do pewnej liczby
    - jak rzucimy 10 razy to może wyjść 0 orłów, zwiększając liczbę powtórzeń zmieniamy częstość wystąpienia orła → okazuje się, że granica dąży do $\frac{1}{2}$.
- opisujemy doświadczenie losowe podając zbiór wszystkich możliwych wyników Ω

$\omega \in \Omega$

- ω - zdarzenie elementarne
- podzbiór Ω to zdarzenie losowe A, $A \subset \Omega$
- zdarzenie losowe A zaszło, gdy doświadczenie zakończyło się wynikiem z A
- przykład: rzut kostką

  $\Omega = \{1, 2, 3, 4, 5, 6\}$

  A - wypadła parzysta liczba oczek $A = \{2, 4, 6\}$
- zdarzenie przeciwne do A:

  $\bar{A} \underset{def}{=} \Omega - A$
- operacje na zbiorach

  $A \cup B$ – suma zdarzeń

  $A \cap B$ – iloczyn – wszystkie wyniki, które są i w A i w B

  $A \setminus B = A \cap \bar{B}$ – różnica – wszystko, co jest w A i nie jest w B

  $B \setminus A = B \cap \bar{A}$ – różnica - to co w B, a nie ma w A
- prawo de Morgana:

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

  chcesz kawę albo herbatę? nie chcę ani tego, ani tego

  $\overline{A \cap B} = \bar{A} \cup \bar{B}$ ← zaprzeczenie iloczynu to suma zaprzeczeń

DEFINICJA AKSJOMETRYCZNA PRAWDOPODOBIEŃSTWA

- niech $\mathcal{A}$ będzie zbiorem zdarzeń z Ω spełniającym warunki:

  (1) $\mathcal{A} \neq \emptyset$

  (2) $A \in \mathcal{A}$ to $\bar{A} \in \mathcal{A}$

  (3) $A_1, A_2 \ldots$ $A_i \in \mathcal{A}$ to $\bigcup_i A_i \in \mathcal{A}$
- Funkcję P określoną na $\mathcal{A}$ o wartościach w $[0,1]$, spełniającą warunki:

① $0 \leqslant P(A) \leqslant 1$ dla każdego $A \in \mathcal{A}$

② $P(\Omega) = 1$

③ dla zdarzeń $A_1, A_2 \ldots$ parami rozłącznych $(A_i \cap A_j = \emptyset ;\ i \neq j)$

$$P\left(\bigcup_{i}^{\infty} A_i\right) = \sum_{i=1}^{\infty} P(A_i)$$

nazywamy prawdopodobieństwem

- $\bigcup_{i=1}^{\infty} A_i = A_1 \cup A_2 \cup A_3 \ldots$ — suma zbiorów

WŁASNOŚCI PRAWDOPODOBIEŃSTWA

① $P(\bar{A}) = 1 - P(A)$ bo $A \cup \bar{A} = \Omega$ $\quad A \cap \bar{A} = \emptyset$ – rozłączne

$P(A \cup \bar{A}) = P(\Omega)$

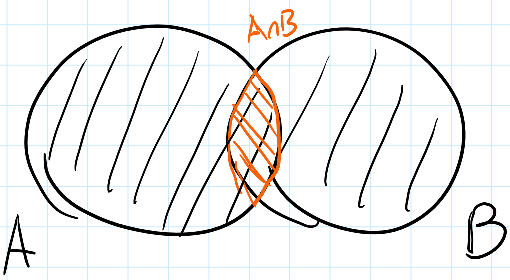
$P(A) + P(\bar{A}) = 1$

$P(\emptyset) = 1 - \underbrace{P(\Omega)}_{1 \text{ – aksjomat } 2} = 0$

$\emptyset$ – zbiór pusty, zdarzenie niemożliwe

② $\underset{\text{podzbiór}}{A \subset B} \rightarrow P(A) \leqslant P(B)$

* ③ $P(A \cup B) = P(A) + P(B) - P(A \cap B)$



$A \cup B = A \cup (B \setminus A)$

$A,\ B \cap \bar{A}$ rozłączne

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

$B = B \cap \Omega = B \cap (A \cup \bar{A}) = \underbrace{(B \cap A) \cup (B \cap \bar{A})}_{\text{rozłączne}}$

- Z powyższego : $P(B) = P((B \cap A) \cup (B \cap \bar{A})) = P(B \cap A) + P(B \cap \bar{A})$

$P(B \cap \bar{A}) = P(B) - P(B \cap A)$

$$P(B \cap \bar{A}) = P(B) - P(B \cap A)$$
$(B \setminus A)$

TWIERDZENIE

- niech Ω ma N elementów lub elementy Ω można ponumerować liczbami naturalnymi (ustawić je w ciąg) oraz niech liczby

$p_i = P(\omega_i)$ $\quad\quad \Omega = \{\omega_1, \omega_2 \dots\}$

spełniają następujące warunki

1. $p_i \geqslant 0 \quad\quad i = 1, 2 \dots$
2. $\sum_{i=1}^{\infty} p_i = 1$

to funkcja $P(A) = \sum_{i: \omega_i \in A} p_i$ jest prawdopodobieństwem na Ω.

$A \subset \Omega \quad\quad A = \{\omega_1, \omega_2 \dots\}$

WNIOSEK

- jeśli Ω ($\Omega = \{\omega_1, \omega_2, \dots\}$) ma N elementów oraz $p_i(\omega_i) = \frac{1}{N}$

wtedy $P(A) = \frac{\text{liczba elementów } A}{N}$ ← tzw. prawdopodobieństwo klasyczne

PRZYKŁAD

- opisać Ω i P dla:

1. rzutu 2 monetami rozróżnialnymi:

$$\Omega = \{\underset{\frac{1}{4}}{(0,0)}, \underset{\frac{1}{4}}{(r,r)}, \underset{\frac{1}{4}}{(r,0)}, \underset{\frac{1}{4}}{(0,r)}\}$$

2. rzut 2 monetami nierozróżnialnymi:

$$\Omega = \{\underset{\frac{1}{4}}{(0,0)}, \underset{\frac{1}{4}}{(r,r)}, \underset{\frac{1}{2}}{(0,r)}\}$$

PRAWDOPODOBIEŃSTWO - PRZYKŁADY

Przykłady

- W produkcji firmy A jest 2% braków, w produkcji firmy B jest 3% braków. Losujemy jeden produkt z A oraz jeden z B. Jakie jest prawdopodobieństwo, że

  a) przynajmniej jeden z nich jest dobry

  b) tylko jeden z nich jest dobry

$\Omega = \{(d,d), (d,w), (w,d), (w,w)\}$

jakość elementu z A

jakość elementu z B

$P(d,d) = 0{,}98 \cdot 0{,}97$

$P(d,w) = 0{,}98 \cdot 0{,}03$

$P(w,d) = 0{,}02 \cdot 0{,}97$

$P(w,w) = 0{,}02 \cdot 0{,}03$

sumują się do 1 zgodnie z teorią

$A = \{(d,d), (d,w), (w,d)\}$

a) $P(A) = 1 - P(\bar{A}) = 1 - P(w,w) = 1 - 0{,}0006 = 0{,}9994 = 99{,}94\%$

b) $B = \{(d,w), (w,d)\} = 0{,}98 \cdot 0{,}03 + 0{,}02 \cdot 0{,}97 = 0{,}0488 = 4{,}88\%$

- W pudle jest 16 płytek dobrych i 4 wadliwe. Losujemy jednocześnie 3 płytki. Jakie jest prawdopodobieństwo, że wśród nich będą co najmniej 2 dobre?

$\Omega = \{(p_1, p_2, p_3) : p_i \in \{d, w\}\}$ Ω ma $\binom{n}{k} \to \binom{16+4}{3}$ elementów

$\binom{n}{k} = \frac{n!}{k!(n-k)!}$ $\binom{20}{3}$ = kalkulator „nCr" $\to 1140$ elem. jednakowo prawdopodobne

$A = A_2 \cup A_3$ — co najmniej 2 dobre

A_2 - 2 dobre z 3 wylosowanych

A_3 - 3 dobre z 3 wylosowanych

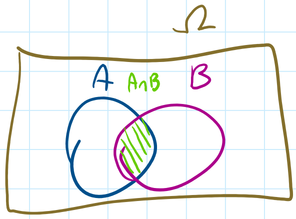
$P(A_2) = \frac{\text{liczba elem. w zdarzeniu } A_2}{\text{liczba elem. } \Omega} = \frac{\binom{16}{2}\binom{4}{1}}{\binom{20}{3}} = \frac{8}{19}$

$P(A_3) = \frac{\text{liczba elem. } A_3}{\text{liczba elem. } \Omega} = \frac{\binom{16}{3}}{\binom{20}{3}} = \frac{28}{57}$

$P(A) = P(A_2 \cup A_3) = P(A_2) + P(A_3) - \underbrace{P(A_2 \cup A_3)}_{= \emptyset \text{ zbiór rozłączny}} = 0{,}91$

- Dane są $P(A) = \frac{1}{3}$, $P(\bar{B}) = \frac{1}{2}$, $P(A \cup B) = \frac{2}{3}$.

- Dane są $P(A)=\frac{1}{3}$, $P(\bar{B})=\frac{1}{2}$, $P(A\cup B)=\frac{2}{3}$.
  Obliczyć:

  a) $P(A\cap B) = P(A)+P(B)-P(A\cup B) = \frac{1}{3}+\frac{1}{2}-\frac{2}{3}=\frac{1}{6}$

  b) $P(A\setminus B) = P(A)-P(A\cap B)=\frac{1}{3}-\frac{1}{6}=\frac{1}{6}$

  $A\cap\Omega = A\cap(B\cup\bar{B}) = (A\cap B)\cup(A\cap\bar{B}) = A\setminus B$

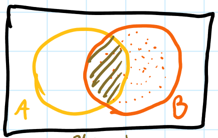
  c) $P(\bar{A}\cup\bar{B}) \overset{I}{=} P(\overline{A\cap B}) = 1-P(A\cap B)=\frac{5}{6}$

  $\overset{II}{=} P(\bar{A})+P(\bar{B})-P(\bar{A}\cap\bar{B}) = \frac{2}{3}+\frac{1}{2}-P(\overline{A\cup B}) =$

  $=\frac{2}{3}+\frac{1}{2}-(1-P(A\cup B)) = \frac{2}{3}+\frac{1}{2}-\frac{1}{3}=\frac{5}{6}$

PRAWDOPODOBIEŃSTWO WARUNKOWE

- Mamy (Ω, P). Niech $P(B)>0$



$P(A\cap B)$

- prawdopodobieństwem warunkowym zdarzenia A przy warunku B, $P(B)>0$, nazywamy liczbę

  $P(A|B) \underset{def}{=} \frac{P(A\cap B)}{P(B)} \rightarrow P(A\cap B) = P(A|B)\cdot P(B)=$

  $P(A)>0 \rightarrow P(B|A)=\frac{P(A\cap B)}{P(A)}$

  sytuacja odwrotna

- PRZYKŁAD:
  Rzucamy 2 razy kostką. Jakie jest prawdopodobieństwo wyrzucenia różnej liczby oczek, jeśli:

  a) suma oczek wynosi 11

  b) suma oczek wynosi 10

  $\Omega = \{(k,l) : k,l\in\{1,2\ldots 6\}\} \rightarrow \Omega$ ma 36 elementów jednakowo prawdopodobnych

  $A = \{(k,l) : k\neq l : k,l\in\{1,2\ldots 6\}\} \rightarrow A$ ma 30 elem. jedn. prawdopodobnych

  $P(A) = \frac{30}{36} = \frac{5}{6}$

  a) $B=\{(5,6),(6,5)\} \subset A$ → podzbiór

  $P(A|B) = \frac{P(A\cap B)}{P(B)} = \frac{P(B)}{P(B)} = 1$ warunek jest podzbiorem zdarzenia, więc $P(A|B)=1$

  b) $B_1 = \{(6,4),(4,6),(5,5)\}$

  $P(A|B_1) = \frac{P(A\cap B_1)}{\ldots} = \frac{\frac{2}{36}}{\ldots} = \frac{2}{\ldots}$

b) $B_1 = \{(6,4), (4,6), (5,5)\}$

$$P(A|B_1) = \frac{P(A\cap B_1)}{P(B_1)} = \frac{\frac{2}{36}}{\frac{3}{36}} = \frac{2}{3}$$

„jaka jest szansa na moje zdarzenie"

$A \cap B_1 = \{(6,4),(4,6)\}$

$P(B_1) = \frac{3}{36}$

- TWIERDZENIE

Jeśli: $P(B) > 0$ to funkcja $P(\cdot|B)$ jest PRAWDOPODOBIEŃSTWEM

$\cdot$ - dowolne zdarzenia z Ω

- zadanie dodatkowe / domowe: uzasadnić aksjomaty na

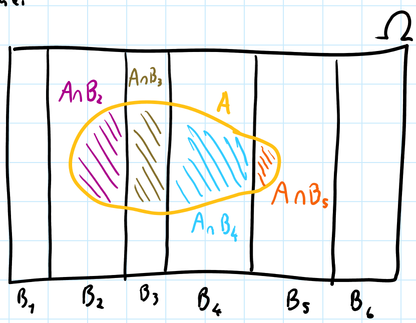
- wniosek: $P(\bar{A}|B) = 1 - P(A|B)$

- TWIERDZENIE O PRAWDOPODOBIEŃSTWIE CAŁKOWITYM

Jeśli zdarzenia $B_1, B_2 \ldots$ spełniają warunki:

1. $\bigcup_{i=1}^{\infty} B_i = \Omega$
2. $B_i \cap B_j = \emptyset \quad i \neq j$
3. $P(B_i) > 0$ dla $i = 1, 2 \ldots$

to $P(A) = \sum_{i=1}^{\infty} P(A|B_i) \cdot P(B_i)$



$$A = A \cap \Omega = A \cap \left(\bigcup_{i=1}^{\infty} B_i\right) = \underbrace{\bigcup_{i=1}^{\infty} (A\cap B_i)}_{\text{zdarzenia parami rozłączne}}$$

$$P(A) = P\left(\bigcup_{i=1}^{\infty} A\cap B_i\right) = \sum_{i=1}^{\infty} P(A\cap B_i) = \sum_{i=1}^{\infty} P(A|B_i)\cdot P(B_i)$$

Często zdarzenia B_i nazywamy czynnikami. Jeśli suma B_i jest Ω, to jeśli znamy $P(A)$ możemy obliczyć prawdopodobieństwa tych zdarzeń warunkowych

- wniosek: WZÓR BAYESA (mamy zdarzenie A, jaka jest szansa że zaszła B_k?)

$$P(B_k|A) = \frac{P(A|B_k)\cdot P(B_k)}{\sum_{i=1}^{\infty} P(A|B_i)\cdot P(B_i)}$$

- PRZYKŁAD:

Bezpiecznik w urządzeniu pochodzi z jednej z 3 firm z prawdopodobieństwami odpowiednio $\frac{1}{4}, \frac{1}{2}, \frac{1}{4}$. Prawdopodobieństwo, że będzie działał przez dany okres T dla poszczególnych firm wynosi 0,9, 0,8, 0,6.

a) jakie jest prawdopodobieństwo, że bezpiecznik będzie działał przez dany czas T?

b) bezpiecznik działa przez okres T. Do której z firm należy z największym prawdopodobieństwem?

c) bezpiecznik przestał działać przed upływem czasu T. Jakie jest prawdopodobieństwo, że pochodził z drugiej firmy?

c) bezpiecznik przestał działać przed upływem czasu T. Jakie jest prawdopodobieństwo, że pochodził z drugiej firmy?

$B_i \to$ bezpiecznik pochodzi z firmy $i = 1, 2, 3$ $C \to$ bezpiecznik działa przez okres T

$P(B_1) = \frac{1}{4}$ $P(C \mid B_1) = 0{,}9$

$P(B_2) = \frac{1}{2}$ $P(C \mid B_2) = 0{,}8$

$P(B_3) = \frac{1}{4}$ $P(C \mid B_3) = 0{,}6$

a) $P(C) = \sum_{i=1}^{3} P(C \mid B_i) \cdot P(B_i) = 0{,}9 \cdot \frac{1}{4} + 0{,}8 \cdot \frac{1}{2} + 0{,}6 \cdot \frac{1}{4} = \frac{31}{40} = 0{,}775$

$B_1, B_2, B_3 \to$ spełniają założenia o prawdopodobieństwie całkowitym

b) ~ która z liczb $P(B_1 \mid C)$ $P(B_2 \mid C)$ $P(B_3 \mid C)$ jest największa?

$P(B_1 \mid C) = \frac{P(C \mid B_1) \cdot P(B_1)}{P(C)} = \frac{0{,}9 \cdot \frac{1}{4}}{\frac{31}{40}} = \frac{9}{31}$

wzór Bayesa

$P(B_2 \mid C) =$

KONSULTACJE

- piątki g. 11-12 C-11, p.206
- kolos poprawkowy → do ustalenia

PRAWDOPODOBIEŃSTWO WARUNKOWE - PRZYKŁADY

Lista 2
zadanie 3

① Wiadomo, że 3% badanych elementów ma skazy. Do ich wykrycia stosuje się test. Jeśli element ma skazę, w 95% test jest pozytywny. W 90% test nie wykazuje wady, gdy element jej nie ma.

a) jakie jest prawdopodobieństwo że elem. ma skazę, jeśli wynik testu jest pozytywny

b) jakie jest prawdopod. że element ma skazę, jeśli test jest dwukrotnie pozytywny?

a) s – element ma skazę $\quad P(s) = 3\% = 0{,}03$

$+$ – test pozytywny

$\bar{s} = d$

$P(+|s) = 0{,}95$

$P(-|d) = 0{,}9 \rightarrow P(+|d) = 1 - 0{,}9 = 0{,}1$

$$P(s|+) = \frac{P(+|s)\cdot P(s)}{P(+)} = \frac{P(+|s)\cdot P(s)}{P(+)} = \frac{0{,}95\cdot 0{,}03}{0{,}95\cdot 0{,}03 + 0{,}1\cdot 0{,}97} = 0{,}227 \quad (22{,}7\%)$$

s, d → rozłączne $\quad \underbrace{P(d) = 0{,}97 \qquad P(s) = 0{,}03}$

spełniają założenia twierdzenia o prawdopodobieństwie całkowitym

$$P(+) = P(+|s)\cdot P(s) + P(+|d)\cdot P(d)$$

b) $$P(s|++) = \frac{P(++|s)\cdot P(s)}{P(++)} = \frac{P(++|s)\cdot P(s)}{P(++|s)\cdot P(s) + P(++|s)\cdot P(d)} = \frac{0{,}95^2\cdot 0{,}03}{0{,}95^2\cdot 0{,}03 + 0{,}1^2\cdot 0{,}97} = 0{,}736 = 73{,}6\%$$

element ma skazę, test 2 razy przeprowadzany

+	−	$P(++|s) = P(+|s)\cdot P(+|s) = (0{,}95)^2$
−	+	$P(+-|s) = 0{,}95\cdot 0{,}05$
+	+	$P(+-|s) = 0{,}05 \cdot 0{,}95$
−	−	$P(--|s) = 0{,}05\cdot 0{,}05$

c) $P(s|+++) = 0{,}964 \qquad (96{,}4\%)$

d) $P(d|-) = 0{,}9983 \qquad (99{,}83\%)$

NIEZALEŻNOŚĆ ZDARZEŃ

- jeśli mielibyśmy $P(A|B) = P(A) = \frac{P(A\cap B)}{P(B)}$

zajście zdarzenia B nie zmienia prawdopodobieństwa zdarzenia A

• jeśli mielibyśmy $P(A|B)=P(A) = \frac{P(A\cap B)}{P(B)}$

zajście zdarzenia B nie zmienia prawdopodobieństwa zdarzenia A

• DEFINICJA: zdarzenia A, B są niezależne $\Leftrightarrow P(A\cap B)=P(A)\cdot P(B)$

• DEFINICJA: Zdarzenia $A_1, A_2 \dots A_n$ są niezależne, jeśli dla danego układu k różnych zdarzeń spośród nich $P(A_{i_1}\cap A_{i_2}\cap \dots A_{i_k}) = P(A_{i_1})\dots$ $k=2,3\dots n$

• WŁASNOŚCI

1. Jeśli zdarzenia A, B są rozłączne i $P(A)\cdot P(B)\neq 0$ to A, B nie są niezależne bo $P(A\cap B) = P(\emptyset)=0$ sprzeczne z $P(A)\cdot P(B)\neq 0$

2. Ω, A są zawsze niezależne, bo $P(\Omega\cdot A)=P(A) = P(A)\cdot 1 = P(A)\cdot P(\Omega)$

3. A, B są niezależne to $A, \bar{B}$ są niezależne
   $\bar{A}, B$ są niezależne
   $\bar{A}, \bar{B}$ są niezależne

4. Własność (3) zachodzi dla większej liczby zdarzeń.

*** ZADANIE DODATKOWE**

Uzasadnić, że jeśli zdarzenia A, B są niezależne, to zdarzenia $\bar{A}, \bar{B}$ są niezależne.

PRZYKŁAD

Trzy moduły programu antywirusowego pracują niezależnie. 1 wykrywa 90% wirusów, 2 – 80%, 3 – 70%.

a) jaki % wirusów wykrywają łącznie

b) jaki % wirusów wykrywa 3 moduł, a nie wykrywa ani 1, ani 2

a) A_i – wirus wykryty przez moduł $i=1,2,3$

$P(A_1)=0{,}9 \quad P(A_2)=0{,}8 \quad P(A_3)=0{,}7$

A_1, A_2, A_3 są niezależne

A - wirus wykryty przez 1 z 3 modułów

$A = A_1\cup A_2\cup A_3$ → przynajmniej 1 moduł wykrywa wirusa

$P(A) = P(A_1\cup A_2\cup A_3) =$ I sposób (dwa razy wzór na sumę zdarzeń, skorzystanie z niezależności)

$= 1 - P(\overline{A_1\cup A_2\cup A_3}) =$ II sposób (prawo de Morgana)

$= 1-P(\bar{A}_1\cap\bar{A}_2\cap\bar{A}_3) = \underbrace{1-P(\bar{A}_1)\cdot P(\bar{A}_2)\cdot P(\bar{A}_3)}_{\text{własność 3}} = 1-0{,}1\cdot 0{,}2\cdot 0{,}3 = 0{,}994$ (99,4%)

b) $P(A_3\cap\bar{A}_1\cap\bar{A}_2) \overset{\text{własność 3}}{=} P(A_3)\cdot P(\bar{A}_1)\cdot P(\bar{A}_2) = 0{,}7\cdot 0{,}1\cdot 0{,}2 = 0{,}014 = 1{,}4\%$

ZMIENNA LOSOWA

DYSTRYBUANTA ZMIENNEJ LOSOWEJ

Ω, A, P

DEFINICJA:

- zmienną losową X nazywamy funkcję $X: \Omega \to R$, że $\bigwedge_{t \in R} \{\omega \in \Omega : X(\omega) < t\}$ jest zdarzeniem

- dystrybuantą zmiennej losowej X nazywamy funkcję $F: R \to [0,1]$ że $F(t) = P(X<t) = P(\omega : X(\omega) < t)$, $t \in R$

PRZYKŁAD

Z talii 52 kart (do brydża) losujemy jednocześnie 2 karty. Niech X określa liczbę pików wśród 2 wylosowanych kart.

$$\Omega = \{\{k_1, k_2\}, k_1 \neq k_2, k_i \in \{1 \dots 52\}\}$$

Ω ma $\binom{52}{2}$ wyników

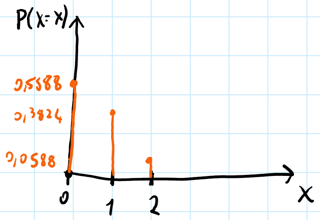
liczba wylosowanych pików

$X: \Omega \to R$

$X \in \{0, 1, 2\}$

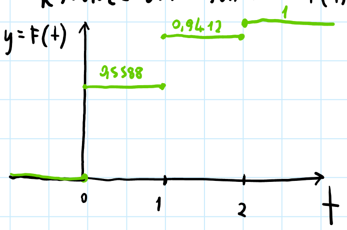
$$P(X=0) = \frac{\binom{39}{2}}{\binom{52}{2}} = 0{,}5588 \qquad P(X=1) = \frac{\binom{13}{1}\cdot\binom{39}{1}}{\binom{52}{2}} = 0{,}3824$$

$$P(X=2) = \frac{\binom{13}{2}}{\binom{52}{2}} = 0{,}0588$$



$$F(t) = P(X<t) = \begin{cases} P(\emptyset) = 0 & t \leq 0 \\ P(X=0) = 0{,}5588 & 0 < t \leq 1 \\ P(X=0 \cup X=1) = 0{,}5588 + 0{,}3824 = 0{,}9412 & 1 < t \leq 2 \\ P(X=0 \cup X=1 \cup X=2) = 1 & 2 < t \end{cases}$$

RYSUNEK DYSTRYBUANTY $F(t) = P(X<t)$



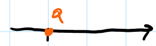
$F(0) = P(X<0) = P(\emptyset) = 0$

$F(1$

FAKT:

Niech $F(t)$ będzie dystrybuantą zmiennej losowej X. Uzasadnić, że:

① $P(X\geq a)=1-F(a)$ bo $P(X\geq a)=1-\underbrace{P(X<a)}_{F(a)}=1-F(a)$

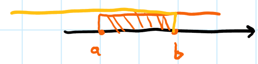


② $a<b$, $P(a\leq X<b)=F(b)-F(a)$

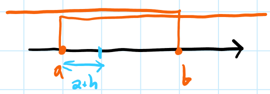
$P(a\leq X<b)=P(\underbrace{X\geq a}_{A}\cap\underbrace{X<b}_{B})=P(X\geq a)+P(X<b)-P(X\geq a\cup X<b)=$

prawd. iloczynu zdarzeń

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



$=1-F(a)+F(b)-1=$



③ $P(X=a)=\underbrace{F(a^+)}_{\lim_{t\to a^+}F(t)}-F(a)$

$P(X=a)=P(\lim_{h\to 0^+}(a\leq X<a+h))=\lim_{h\to 0^+}\left[F(a+h)-F(a)\right]=F(a^+)-F(a)$

ZMIENNE LOSOWE CD.

- zmienne losowe dzielimy na:
  - dyskretne
  - ciągłe
  - mieszane

- zmienna losowa X jest DYSKRETNA (skokowa), jeśli zbiór wartości X jest skończony lub równoliczny z $\mathbb{N}$ ← zbiór liczb naturalnych
  - zbiór par $(x_i, P(X=x_i))$ nazywamy ROZKŁADEM zmiennej losowej X.
  - przykład:

x_i	-3	-1	1	4
$P(X=x_i)$	$0,1$	$0,1$	$0,6$	$0,2$

  - WARTOŚCIĄ OCZEKIWANĄ (średnią) dyskretnej zmiennej losowej X nazywamy liczbę EX

$$EX = \sum_{i=1}^{\infty} x_i P(X=x_i)$$ ← jeśli ten szereg jest rozbieżny, wartość oczekiwana nie istnieje

  - WARIANCJĄ zmiennej losowej dyskretnej X nazywamy liczbę $Var X$

$$Var X = \sum_{i=1}^{\infty} (x_i - EX)^2 \cdot P(X=x_i)$$

  - przykład: dla zmiennej losowej X, Y obliczyć $EX, EY, Var X, Var Y$

$P(X=49) = P(X=51) = \frac{1}{2}$
$P(Y=-100) = \frac{1}{4}$ $\quad P(Y=100) = \frac{3}{4}$

$EX = 49 \cdot \frac{1}{2} + 51 \cdot \frac{1}{2} = 50$
$EY = -100 \cdot \frac{1}{4} + 100 \cdot \frac{3}{4} = 50$

„przeciętna wartość zmiennej"
$Var X = (49-50)^2 \cdot \frac{1}{2} + (51-50)^2 \cdot \frac{1}{2} = 1$
$Var Y = (-100-50)^2 \cdot \frac{1}{4} + (100-50)^2 \cdot \frac{3}{4} = 7500$

„jak bardzo wartości zmiennej odbiegają od wartości przeciętnej?"
„wariancja powinna być podawana wraz z wartościami średnimi zarobków"

  - przykład dodatkowy:

$EX = (-3) \cdot 0,1 + (-1) \cdot 0,1 + 1 \cdot 0,6 + 4 \cdot 0,2 = 1$
$Var X = (-3-1)^2 \cdot 0,1 + (-1-1)^2 \cdot 0,1 + (1-1)^2 \cdot 0,6 + (4-1)^2 \cdot 0,2 = 0,38$

  - TYPOWE ZMIENNE LOSOWE DYSKRETNE:

① Zmienna losowa zerojedynkowa X

$X \in \{0, 1\}$ — porażka (0), sukces (1)

• test na obecność wirusa

$P(X=1) = p \qquad P(X=0) = q$
$EX = p$

tekst
istotne
równania, wzory
przykłady
komentarz
komentarz
komentarz
komentarz
definicje
inne ważne

$P(X=1)=p \qquad P(X=0)=q$

$EX = p$

$VarX = p(1-p)$

(2) Rozkład Bernoulliego $\quad X \sim B(n,p)$

- powtarzamy n razy niezależnie doświadczenie ze zmienną losową zerojedynkową

X - liczba „1" w n niezależnych powtórzeniach

$X \in \{0,1 \ldots, n\} \ni k$

$P(X=k) = \binom{n}{k} p^k q^{n-k} \qquad \sum_{k=0}^{n} \binom{n}{k} p^k q^{n-k} = (p+q)^n = 1^n = 1$

$EX = np$

$VarX = np(1-p)$

- musimy znać liczby n i p (powtórzeń 1 i 0)
- rzut monetą

- liczbę k_0 nazywamy NAJBARDZIEJ PRAWDOPODOBNĄ WARTOŚCIĄ zmiennej losowej X jeżeli $P(X=k_0) \geqslant P(X=k)$

$X \sim B(n,p)$

$$k_0 \begin{cases} \text{część całkowita } (n+1)p, & (n+1)p \notin \mathbb{N} \\ (n+1)p \text{ lub } (n+1)p-1, & (n+1)p \in \mathbb{N} \end{cases}$$

- przykład: firma składa oferty na wykonanie każdego z 4 niezależnych projektów. Każda oferta ma prawdopodobieństwo przyjęcia $\frac{2}{3}$. Jaka jest najbardziej prawdopodobna liczba przyjętych ofert i ile wynosi jej prawdopodobieństwo?

X - liczba przyjętych ofert spośród 4 złożonych

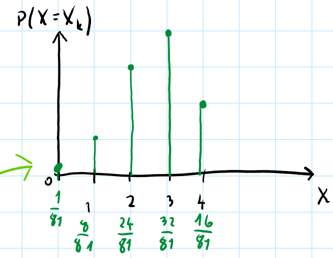
$X \sim B(4, \frac{2}{3}) \qquad k_0 = ?$

$(n+1)p = (4+1)\frac{2}{3} = \frac{10}{3} \notin \mathbb{N}$

$k_0 = 3$

$P(X=3) = \binom{4}{3}\left(\frac{2}{3}\right)^3\left(\frac{1}{3}\right)^1 = \frac{32}{81} = 0{,}395$

$P(X=4) = \ldots$



(3) Rozkład geometryczny z parametrem $p \in (0,1)$

- powtarzamy doświadczenie ze zmienną losową zerojedynkową dopóki sukces pojawi się po raz pierwszy

$X \in \{1,2,3 \ldots\} \ni k$

$P(X=k) = P(\underbrace{0,0,\ldots,}_{k-1}\underbrace{1}_{k}) = q^{k-1} \cdot p$

$EX = \frac{1}{p}$

$VarX = \frac{1-p}{p^2}$

$EX = \frac{1}{p}$

$Var\, X = \frac{1-p}{p^2}$

- przykład: wadliwość wyrobów wynosi 0,02. Jakie jest prawdopodobieństwo, że wadliwy wyrób wylosujemy po raz pierwszy za piątym razem?

X - numer losowania, w którym wylosujemy po raz pierwszy wadliwy wyrób
X ma rozkład geometryczny

$p = 0{,}02$

$P(X=5) = (1-0{,}02)^4 \cdot 0{,}02 = 0{,}0184$

$k_0 = 1$

$EX = \frac{1}{0{,}02} = 50$

④ Rozkład Poissona z parametrem λ

$\lambda > 0$ gdy $P(X=k) = e^{-\lambda} \frac{\lambda^k}{k!}$ $\quad k \in \{0,1,2\dots\}$

$EX = \lambda$
$Var\, X = \lambda$

$$k_0 = \begin{cases} \text{część całkowita } \lambda & \text{gdy } \lambda \notin \mathbb{N} \\ \lambda \text{ lub } \lambda - 1 & \text{gdy } \lambda \in \mathbb{N} \end{cases}$$

- przykład dla $\lambda = 4{,}0$ (odczyt z tabeli)

$P(X \geq k)$ $\qquad P(X \geq 6) = 0{,}2149$

- przykład: zadanie 6 /lista 3

$\lambda = 8$ $\qquad p = 0{,}95$

X - liczba samochodów do naprawy w warsztacie

X ma rozkład Poissona z param. $\lambda = 8$

$EX = 8$ lub 7 $\qquad \infty$ możliwe rozbite samochody

ile powinno być miejsc w warsztacie, aby dla każdego uszkodzonego auta było miejsce w warsztacie?

l - liczba miejsc w warsztacie

$X \leq l$ liczba aut do naprawy $\leq$ liczba miejsc

$P(X \leq l) \geq 0{,}95$

$(X > l) = (X \geq l+1)$

$1 - P(X > l) \geq 0{,}95$

$P(X \geq l+1) \leq 0{,}05$

$P(X > l) \leq 0{,}05$

Z tablicy: $p \to 0{,}0342 \to k = 14$

$l + 1 \geqslant 14 \to l \geqslant 13$

ze wzoru

$\lambda = 8$

$P(X=7) = e^{-8} \frac{8^7}{7!}$

$P(X=7) = P(X \geqslant 7) - P(X \geqslant 8) = 0{,}6866 - 0{,}5470$

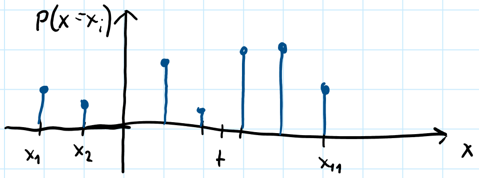
z tablic

Wykład 5

środa, 27 marca 2019 07:33

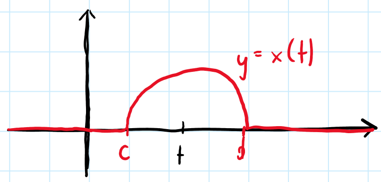
ZMIENNE LOSOWE CIĄGŁE

- przypomnienie: zmienna losowa dyskretna



$$F(t) = P(X<t) = \sum_{i: x_i < t} P(X=x_i)$$

- zmienna losowa



$$F(t) = P(X<t) = \int_{-\infty}^{t} f(x)\,dx$$

- zmienna losowa X jest ciągła, jeśli istnieje nieujemna, całkowalna funkcja $f(x)$, zwana GĘSTOŚCIĄ X, że

$$F(t) = P(X<t) = \int_{-\infty}^{t} f(x)\,dt$$

- FAKTY:

1. Dystrybuanta $F(t)$ zmiennej losowej X ciągłej jest funkcją ciągłą
2. Jeśli $f(x)$ jest ciągła, to $F'(x) = f(x)$
3. X ciągła: $P(X=a) = F(a^+) - F(a) = 0$

   $F(t)$ ciągła

   $$P(a \leqslant X < b) = F(b) - F(a) = \int_a^b f(x)\,dx$$

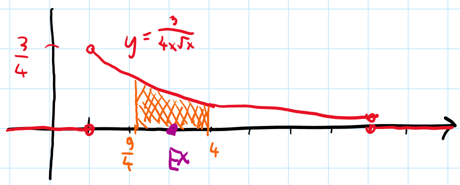
- TWIERDZENIE: $f(x)$ jest gęstością pewnej zmiennej losowej X, gdy:

1. $\bigwedge_{x \in \mathbb{R}} f(x) \geqslant 0$
2. $\int_{-\infty}^{\infty} f(x)\,dx = 1$

- PRZYKŁAD: Czy można dobrać stałą c, aby funkcja

$$f(x) = \begin{cases} \frac{c}{x\sqrt{x}} & 1 \leqslant x \leqslant 9 \\ 0 & \text{poza} \end{cases}$$

była gęstością?



① $c \geqslant 0$

② $1 = \int_{-\infty}^{\infty} f(x)\,dx = \int_{-\infty}^{1} 0\,dx + \int_{1}^{9} \frac{c}{x\sqrt{x}}\,dx + \int_{9}^{\infty} 0\,dx = c \cdot \left(-2x^{-\frac{1}{2}}\right)\Big|_1^9 =$

$= -2c\left(\frac{1}{3} - 1\right) = \frac{4}{3}c$

$c = \frac{3}{4}$ Odpowiedź: $c = \frac{3}{4}$

- obliczyć: $P\left(\frac{9}{4} < X < 4\right) = \int_{\frac{9}{4}}^{4} \frac{3\,dx}{4x\sqrt{x}} = \frac{3}{4}\left(-2x^{-\frac{1}{2}}\right)\Big|_{\frac{9}{4}}^{4} = -\frac{3}{2}\left(\frac{1}{2} - \frac{2}{3}\right) = \frac{1}{4}$
- wartość oczekiwana: $EX = \int_{-\infty}^{\infty} x f(x)\,dx = \int_{1}^{9} x \frac{3}{4x\sqrt{x}}\,dx = \frac{3}{4} \cdot 2\sqrt{x}\,\Big|_1^9 = \frac{3}{2}(3-1) = 3$
- wariancja zmiennej: $Var\,X = E(X^2) - (EX)^2 = \int_{-\infty}^{\infty} x^2 f(x)\,dx - 3^2 = \int_{1}^{9} x^2 \cdot \frac{3}{4} \cdot \frac{1}{x\sqrt{x}}\,dx - 9 =$

  własność 4

  $= \frac{3}{4} \cdot \frac{2}{3}\, x^{\frac{3}{2}}\Big|_1^9 - 9 = \frac{1}{2}(27-1) - 9 = 4$
- odchylenie standardowe: $DX = \sqrt{Var X} = 2$

- WARTOŚCIĄ OCZEKIWANĄ EX zmiennej losowej ciągłej X nazywamy liczbę

$$EX = \int_{-\infty}^{\infty} x f(x)\,dx$$

- własności wartości oczekiwanej:

① $E(cX + d) = cEX + d$

② $g(x)$ – funkcja całkowalna

$$E g(X) = \int_{-\infty}^{\infty} g(x) f(x)\,dx$$

gęstość

③ $E(X+Y) = EX + EY$

- WARIANCJĄ zmiennej losowej X nazywamy liczbę $Var\,X$

$$Var\,X = \int_{-\infty}^{\infty} (x - EX)^2 f(x)\,dx$$

- DYSPERSJA (odchylenie standardowe) X to $DX = \sqrt{Var\,X}$

- własności $Var\,X$

① $\operatorname{Var} X \geq 0$

② $\operatorname{Var} X = 0 \iff P(X = c) = 1$

× skoncentrowane w jednym punkcie

③ $\operatorname{Var}(aX + b) = a^2 \operatorname{Var} X$

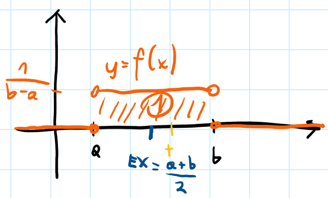
④ $\operatorname{Var} X = E(X^2) - (EX)^2 = \int_{-\infty}^{\infty} x^2 f(x)\,dx - (EX)^2$

⑤ X, Y - niezależne $\Rightarrow$ $\operatorname{Var}(X+Y) = \operatorname{Var} X + \operatorname{Var} Y$

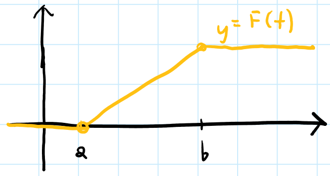
TYPOWE ROZKŁADY CIĄGŁE:

① Rozkład jednostajny na przedziale $[a,b]$ ma gęstość $f(x)$

$$f(x) = \begin{cases} \frac{1}{b-a}, & a \leq x \leq b \\ 0 & \text{poza} \end{cases}$$



$$F(t) = \int_{-\infty}^{t} f(x)\,dx = \begin{cases} 0, & t < a \\ \frac{t-a}{b-a} & a \leq x \leq b \\ 0, & t > b \end{cases}$$

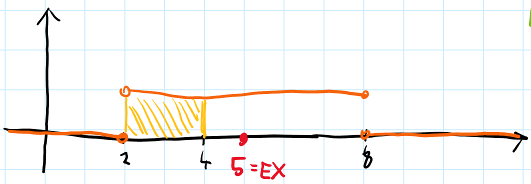


$EX = \frac{a+b}{2}$ $\qquad$ $\operatorname{Var} X = \frac{(b-a)^2}{12}$

- PRZYKŁAD: Poziom cieczy X w zbiorniku ma rozkład jednostajny na $[2,8]$. Jakie jest prawdopodobieństwo, że poziom cieczy jest większy niż 4? Podaj wariancję i tej cieczy

$$P(X<4) = \int_{-\infty}^{2} f(x)\,dx = \int_{-\infty}^{2} 0\,dx + \int_{2}^{4} \frac{1}{6}\,dx = \frac{1}{6}x\Big|_2^4 = \frac{1}{3}$$

$\frac{1}{b-a}$

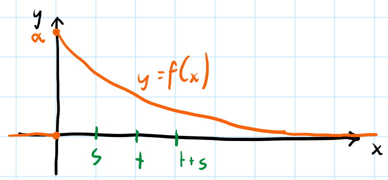


$EX = \frac{2+8}{2} = 5$ $\qquad$ $\operatorname{Var} = \frac{(8-2)^2}{12} = 3$

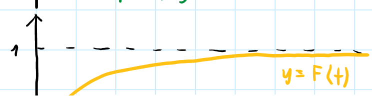
② ROZKŁAD WYKŁADNICZY Z PARAMETREM α, $\alpha > 0$

ma gęstość:

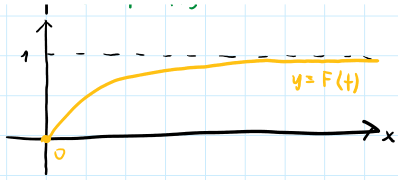
$$f(x) = \begin{cases} 0, & x < 0 \\ \alpha e^{-\alpha x}, & x \geq 0 \end{cases}$$



$$F(t) = \begin{cases} 0, & t < 0 \\ \quad -\alpha t \end{cases}$$



$$F(t)=\begin{cases}0, & t<0\\ 1-e^{-\alpha t}, & t\geq 0\end{cases}$$



$EX=\frac{1}{\alpha}$ $\qquad$ $VarX=\frac{1}{\alpha^2}$

- rozkład wykładniczy ma tzw. własność „braku pamięci", tzn.

$$\bigwedge_{s,t>0} P(X>s+t\,|\,X>s)=P(X>t)$$

„jeżeli urządzenie pracuje dłużej niż do chwili s, to zaczynamy liczenie od chwili s"

bo $P(X>s+t\,|\,X>s)=\frac{P((X>s+t)\cap(X>s))}{P(X>s)}=\frac{P(X>t+s)}{P(X>s)}=\frac{1-F(t+s)}{1-F(s)}=$

$$=\frac{1-[1-e^{-\alpha(s+t)}]}{1-[1-e^{-\alpha s}]}=\frac{e^{-\alpha(s+t)}}{e^{-\alpha s}}=e^{-\alpha t}=1-[1-e^{-\alpha t}]=1-F(t)=P(X>t)$$

- PRZYKŁAD: Czas pracy diody ma rozkład wykładniczy z parametrem $\alpha=10^{-3}$. Wiadomo, że dioda działała przez 1000 dób. Jakie jest prawdopodobieństwo, że będzie działać co najmniej przez 3000 dób? Podać wartość oczekiwaną i wariancję diody.

X – czas pracy diody

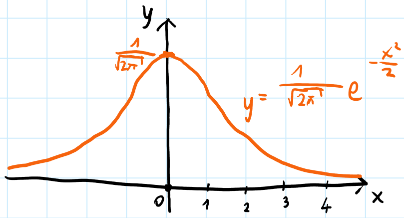
$$P(X>\underbrace{3000}_{t+s}\,|\,X>\underbrace{1000}_{s})\underset{\substack{\text{własność}\\ \text{braku pamięci}}}{=}P(X>2000)=1-F(2000)=1-[1-e^{-10^{-3}\cdot 2000}]=e^{-2}=\frac{1}{e^2}$$

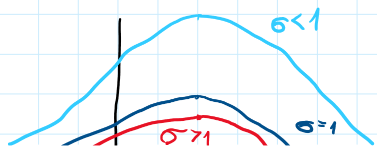
$EX=\frac{1}{\alpha}=1000$

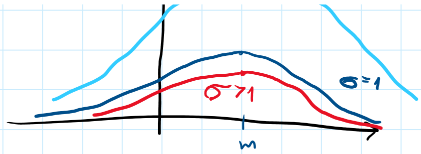
$VarX=\frac{1}{\alpha^2}=10^6$

3 ROZKŁAD NORMALNY z parametrami $m\in\mathbb{R}$, $\sigma>0$

- oznaczenie: $N(m,\sigma)$
- gęstość: $f(x)=\frac{1}{\sqrt{2\pi}\,\sigma}e^{-\frac{(x-m)^2}{2\sigma^2}}$, $x\in\mathbb{R}$
- narysować gęstość $N(0,1)$, $f(x)=\frac{1}{\sqrt{2\pi}}e^{-\frac{x^2}{2}}$







$EX = m$ $\quad$ $VarX = \sigma^2$

ROZKŁAD NORMALNY

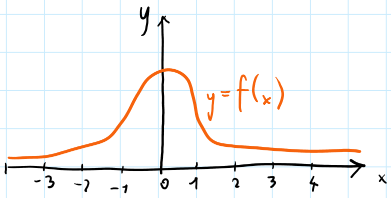
(twierdzenie graniczne)

$X \sim N(0,1)$ $\qquad f(x) = \frac{1}{\sqrt{2\pi}} e^{\frac{-x^2}{2}}$

$\Phi(t)$ – dystrybuanta z.l. $N(0,1)$

$$\Phi(t) = P(x<t) = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{t} e^{-\frac{x^2}{2}} dx$$

- wartości $\Phi(t)$ podawane w tablicach



- PRZYKŁAD: Niech $X \sim N(0,1)$. Obliczyć, korzystając z tablic,

a) $P(X < 1{,}21) = \int_{-\infty}^{1{,}21} \frac{1}{\sqrt{2\pi}} e^{-\frac{x^2}{2}} dx = \Phi(1{,}21) = 0{,}8869$

b) $P(X < -1{,}21) = P(X > 1{,}21) = 1 - P(X \leqslant 1{,}21) = 1 - \Phi(1{,}21) =$

$= 1 - 0{,}8869 = 0{,}1131$

funkcja parzysta
$\Phi(-a) = 1 - \Phi(a)$

c) $P(-3{,}5 < X < 2{,}34) = \Phi(2{,}34) - \Phi(-3{,}5) = 0{,}99358 - (1 - 0{,}99774) =$

$= 0{,}9933474$

zapisać na tablicach

- **FAKT I:** $Y \sim N(m, \sigma)$ to $\frac{Y-m}{\sigma} \sim N(0,1)$

standaryzacja Y

$$P(Y<t) = P\Big(\underbrace{\frac{Y-m}{\sigma}}_{N(0,1)} < \frac{t-m}{\sigma}\Big) = \Phi\Big(\frac{t-m}{\sigma}\Big)$$

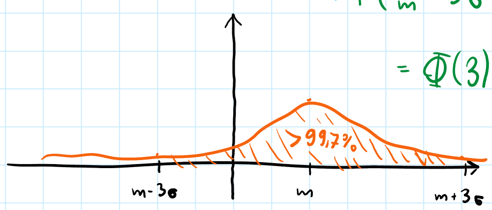
- **FAKT II:**

Jeśli $X \sim N(m,\sigma)$ to $P(m - 3\sigma < X < m + 3\sigma) > 0{,}997$

$P(m - 3\sigma < X < m + 3\sigma) = P\Big(-3 < \underbrace{\frac{X-m}{\sigma}}_{\sim N(0,1)} < 3\Big) =$

$= \Phi(3) - \Phi(-3) = 2 \cdot \Phi(3) - 1 = 2 \cdot 0{,}99865 - 1 =$

$= 0{,}9973$



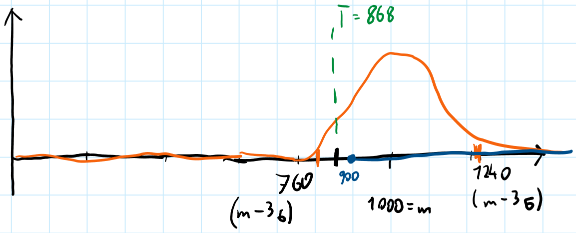
- PRZYKŁAD: statystyki np. w medycynie

- jeżeli $t > 3{,}5 \rightarrow \Phi(t) \approx 1$
  $t < -3{,}5 \rightarrow \Phi(t) \approx 0$

- PRZYKŁAD:
  Czas sprawnej pracy silników pewnego typu ma rozkład normalny z parametrami 1000, 80. Jakie jest prawdopodobieństwo, że:

  a) silnik pracuje krócej niż 400 dni?
  b) silnik pracuje dłużej niż 900 dni?
  c) jaki powinien być okres gwarancji, aby na 95% silnik pracował przez cały okres gwarancji?

X - czas pracy silnika $\quad X \sim N(1000, 80)$



a) $P(X<400) = P\Big(\underbrace{\frac{X-1000}{80}}_{\sim N(0,1)} < \underbrace{\frac{400-1000}{80}}_{-\frac{60}{8}}\Big) = \Phi(-7{,}5) = 0$

b) $P(X>900) = 1 - P(X \leqslant 900) = 1 - P\left(\frac{X-1000}{80} \leqslant \frac{900-1000}{80}\right) = 1 - \Phi\left(-\frac{5}{4}\right) = 1 - 1 + \Phi(1{,}25) =$
$= 0{,}8944$

c) T - okres gwarancji

$P(X \geqslant T) > 0{,}95$

$1 - P(X<T) > 0{,}95 \rightarrow P(X<T) < 0{,}05$

STANDARYZACJA: $P\Big(\underbrace{\frac{X-1000}{80}}_{\sim N(0,1)} < \frac{T-1000}{80}\Big) = \underbrace{\Phi\left(\frac{T-1000}{80}\right)}_{\Phi(-a) = 1-\Phi(a)} < 0{,}05$

$1 - \Phi\left(-\frac{T-1000}{80}\right) < 0{,}05 \rightarrow \underbrace{\Phi\left(-\frac{T-1000}{80}\right)}_{\text{funkcja rosnąca}} > 0{,}95 = \Phi(1{,}65)$

$-\frac{T-1000}{80} > 1{,}65 \rightarrow T < 868$

Okres gwarancji mniejszy niż 868 dni

NIEZALEŻNOŚĆ ZMIENNYCH LOSOWYCH

- zmienne losowe $X_1, X_2 \dots X_n$ są niezależne $\Leftrightarrow$ dla każdych liczb $t_1, t_2 \dots t_n$
  $P(X_1<t_1) \cap P(X_2<t_2) \cap \dots \cap P(X_n<t_n) = P(X_1<t_1) \cdot P(X_2<t_2) \cdot \dots \cdot P(X_n<t_n)$

- niech $(X, Y) \rightarrow$ dwie zmienne
  $P(X<t \cap Y<u) \underset{def.}{=} F(t, u)$

$P(X<t \cap Y<u) \underset{def.}{=} F(t,u)$

$X, Y \to$ wektory niezależne $\to P(X<t)P(Y<u)$

- WŁASNOŚCI:

1. $F(t,u) = F_x(t) \cdot F_y(u)$
2. $\frac{\partial^2 F(t,u)}{\partial u \partial t} \underset{def}{=} \underbrace{f(t,u)}_{\text{gęstość } (X,Y)}$
3. $X, Y \to$ niezależne

$f(t,u) = f_x(t) \cdot f_y(u)$

- TWIERDZENIE:

Jeżeli $X_1, X_2 \dots X_n$ są niezależnymi zmiennymi losowymi $X_k \sim N(m_k, \sigma_k)$ to

$\sum_{k=1}^{n} X_k \sim N\left(\sum_{k=1}^{n} m_k, \underbrace{\sqrt{\sigma_1^2 + \sigma_2^2 + \dots \sigma_n^2}}\right)$ $\sqrt{\sum_{k=1}^{n} \sigma_k^2}$

- TWIERDZENIA GRANICZNE:

1. SŁABE PRAWO WIELKICH LICZB

Jeśli $X_1, X_2 \dots X_n$ jest ciągiem niezależnych zmiennych losowych o jednakowym rozkładzie i skończonej wariancji $(Var X_i = c^2 < \infty \to EX_i = m < \infty)$ $(i = 1, 2 \dots)$ to

$$\bigwedge_{\varepsilon > 0} \lim_{n \to \infty} P\Big(\big|\underbrace{\tfrac{1}{n} \sum_{k=1}^{n} X_k}_{\text{średnia sumy zmiennych losowych}} - m\big| < \varepsilon\Big) = 1$$

- WŁASNOŚCI

$E\left(\frac{1}{n} \sum_{k=1}^{n} X_k\right) = \frac{1}{n} \sum_{k=1}^{n} EX_k = \frac{1}{n} \sum_{k=1}^{n} m = \frac{1}{n} \cdot n \cdot m = m$

$Var\left(\frac{1}{n} \sum_{k=1}^{n} X_k\right) = \left(\frac{1}{n}\right)^2 \sum_{k=1}^{m} Var X_k = \frac{1}{n^2} \sum_{k=1}^{n} c^2 = \frac{1}{n^2} \cdot n \cdot c^2 = \frac{c^2}{n}$

- PRZYKŁAD:

a) $X_k \sim B(1, \frac{1}{4})$ $k = 1, 2 \dots$ $EX_k = \frac{1}{4}$

$\frac{1}{n} \sum_{k=1}^{n} X_k \xrightarrow[\text{prawdopodobieństwa}]{\text{wg.}} \frac{1}{4}$

b) $X_k \sim B(1, p)$ to $\frac{1}{n} \sum_{k=1}^{n} X_k \to p$

2. CENTRALNE TWIERDZENIE GRANICZNE

Jeśli $X_1, X_2 \dots$ jest ciągiem niezależnych zmiennych losowych o skończonej wariancji (wartości oczekiwane też są skończone) to

Jeśli $X_1, X_2 \ldots$ jest ciągiem niezależnych zmiennych losowych o skończonej wariancji (wartości oczekiwane też są skończone) to

$$\bigwedge_{t\in\mathbb{R}} \lim_{n\to\infty} P\left(\frac{\sum_{k=1}^{n} X_k - E\left(\sum_{k=1}^{n} X_k\right)}{\sqrt{Var \sum_{k=1}^{n} X_k}} < t\right) = \Phi(t)$$

- WNIOSEK:

Jeśli $X_1, X_2 \ldots$ to ciąg niezależnych zmiennych losowych o jednakowym rozkładzie i skończonej wariancji $Var\, X_k = c^2$ $(EX_k = a)$ $k = 1, 2 \ldots$, to

$$\bigwedge_{t\in\mathbb{R}} \lim_{n\to\infty} P\left(\frac{\sum_{k=1}^{n} X_k - na}{\sqrt{nc^2}} < t\right) = \Phi(t)$$

jakaś stała

- PRZYKŁAD:

Czas obsługi klienta jest zmienną losową o rozkładzie wykładniczym z parametrem $\alpha = 0{,}5$ min. Jakie jest prawdopodobieństwo że czas obsługi 200 takich klientów przekroczy 420 minut?

$k = 1, 2 \ldots 200$ $\qquad$ X_k – czas obsługi klienta

X_k – rozkład wykładniczy z $\alpha = 0{,}5$ min

$EX_k = \frac{1}{\alpha} = 2$ $\qquad$ $Var\, X_k = \frac{1}{\alpha^2} = 4$

$$P\left(\sum_{k=1}^{200} X_k > 420\right) = 1 - P\left(\sum_{k=1}^{200} X_k \leqslant 420\right) = 1 - P\left(\frac{\sum_{k=1}^{200} X_k - 200\cdot 2}{\sqrt{200\cdot 4}} < \frac{420 - 200\cdot 2}{\sqrt{200\cdot 4}}\right) =$$

$$= 1 - \Phi(0{,}71) = 1 - 0{,}7611 = 0{,}2389$$

Wykład 7

środa, 10 kwietnia 2019 07:33

- zadania do listy 5 włącznie
- kolokwium za tydzień, 45 minut + wykład, zadania przez edukację dostaniemy
- wziąć tablice, wziąć dodatkowe kartki,
- wydruk ze strony: typowe rozkłady prawdopodobieństwa
- kalkulator

PRZYKŁAD:

Prawdopodobieństwo, że wyprodukowany element będzie dobry, wynosi 0,9.

a) jakie jest prawdopodobieństwo, że wśród 150 takich elementów dobrych będzie przynajmniej 130

b) ile elementów należy wyprodukować, aby prawdopodobieństwo, że będzie wśród nich przynajmniej 100 dobrych, było większe niż 0,95?

POPRAWKA

X_k o wart. całkowitych $\qquad a, b$ całkowite

$$\underbrace{P\left(a \leqslant \sum_{k=1}^{n} X_k \leqslant b\right)} = \underbrace{P\left(a - \tfrac{1}{2} \leqslant \sum_{k=1}^{n} X_k \leqslant b + \tfrac{1}{2}\right)}$$

$$P\left(\sum_{k=1}^{n} X_k = b\right) = P\left(b - \tfrac{1}{2} \leqslant \sum_{k=1}^{n} X_k \leqslant b + \tfrac{1}{2}\right)$$

ROZWIĄZANIE:

X_k - jakość k-tego elementu

$k = 1 \ldots 150$ $\qquad EX_k = 1 \cdot 0{,}9 = 0{,}9$

$X \sim B(1; 0{,}9)$, $\qquad Var\, X_k = 0{,}9 \cdot 0{,}1 = 0{,}09$

a) $$P\Big(\underbrace{\sum_{k=1}^{150} X_k}_{\text{liczba dobrych elementów wśród wszystkich}} \geqslant 130\Big) \overset{\text{poprawka}}{=} P\Big(\sum_{k=1}^{150} X_k \geqslant \underbrace{129{,}5}_{\text{chcemy ująć też wartość 130}}\Big) = 1 - P\Big(\sum_{k=1}^{150} X_k \leqslant 129{,}5\Big) =$$

$$\overset{\text{centralne twierdzenie graniczne}}{=} 1 - P\left(\frac{\sum_{k=1}^{150} X_k - 150 \cdot 0{,}9}{\sqrt{150 \cdot 0{,}09}} < \frac{129{,}5 - 150 \cdot 0{,}9}{\sqrt{150 \cdot 0{,}09}}\right) =$$

$\rightarrow N(0,1)$ $\qquad \rightarrow = -1{,}22$

$$= 1 - \Phi(-1{,}22) = 1 - 1 + \Phi(1{,}22) = 0{,}8888$$

b) „dla jakiego n liczba dobrych elementów jest większa niż 100?"

$$P\Big(\underbrace{\sum_{k=1}^{n} X_k}_{\text{liczba dobrych elementów}} \geqslant 100\Big) > 0{,}95$$

liczba dobrych elementów

$$P\left(\sum_{k=1}^{n} X_k > 99{,}5\right) = 1 - P\left(\sum_{k=1}^{n} X_k \leq 99{,}5\right) = 1 - P\left(\underbrace{\frac{\sum_{k=1}^{n} - n\cdot 0{,}9}{\sqrt{n\cdot 0{,}09}}}_{\to N(0,1)} \leq \frac{99{,}5 - n\cdot 0{,}9}{\sqrt{n\cdot 0{,}09}}\right) =$$

$$= 1 - \Phi\left(\frac{99{,}5 - n\cdot 0{,}9}{\sqrt{n\cdot 0{,}09}}\right) = \underbrace{\Phi\left(-\frac{99{,}5 - n\cdot 0{,}9}{\sqrt{n\cdot 0{,}09}}\right)}_{1-\Phi(a) = \Phi(-a)} > 0{,}95$$

$$-\frac{99{,}5 - n\cdot 0{,}9}{\sqrt{n\cdot 0{,}09}} > 1{,}64$$

$$n\cdot 0{,}9 - 1{,}64\cdot\sqrt{n}\cdot 0{,}3 - 99{,}5 > 0$$

$$\sqrt{n_1} = 10{,}79 \qquad \sqrt{n} < 0$$

$$n_1 = 116{,}45 \longrightarrow n \geqslant 117$$

PRZYKŁAD ALGEBRAICZNY

- niech $X_1, X_2, \dots, X_n$ to niezależne zmienne losowe o jednakowym rozkładzie, opisanym dystrybuantą $F(t)$.

Wyznaczyć dystrybuantę (i gęstość, jeśli istnieje) dla zmiennych losowych:

a) $Y = \max\{X_1, X_2 \dots X_n\}$

b) $Z = \min\{X_1 \dots X_n\}$

X_k - czas pracy k-tego elementu $\quad k = 1 \dots n$

Y - maksimum $\quad \max\{X_1 \dots X_n\}$
„czas pracy n elementów połączonych w układ równoległym" → gdy co najmniej jeden pracuje

Z - minimum $\quad \min\{X_1 \dots X_n\}$
„czas pracy elementów połączonych szeregowo" → gdy wszystkie pracują

$$F(t) = P(X_k < t) \qquad k = 1 \dots n$$

a) $F_Y(t) = P(Y < t) = P(\max(X_1 \dots X_n) < t) =$

$$= P(X_1 < t \cap X_2 < t \cap \dots \cap X_k < t) = P(X_1 < t)\cdot P(X_2 < t)\cdot \dots \cdot P(X_k < t) =$$

$$= (F(t))^n$$

$$f_Y(t) = \frac{\partial}{\partial t} F_Y(t) = n\cdot F(t)^{n-1} \cdot \underbrace{\frac{\partial}{\partial t} F(t)}_{f(t)}$$

b) $F_Z(t) = P(Z < t) = P(\min(X_1 \ldots X_n) < t) = 1 - P(\min(X_1 \ldots X_n) \geq t) =$

$= 1 - P(X_1 \geq t \cap \ldots \cap X_k \geq t) = 1 - \underbrace{P(X_1 \geq t)}_{1-F(t)} \cdot \ldots \cdot P(X_k \geq t) =$

$= 1 - [1 - F(t)]^n$

$f_Z(t) = \frac{\partial}{\partial t} F_Z(t) = n(1 - F(t))^{n-1} \cdot \frac{\partial}{\partial t} F(t)$

PRZYKŁAD

Czas działania każdego elementu jest zmienną losową o rozkładzie jednostajnym na przedziale $[0, 10]$

Obliczyć wartość oczekiwaną czasu działania 5 takich elementów podłączonych

a) równolegle
b) szeregowo

X_k - czas działania k-tego elementu $k = 1 \ldots 5$

$EX_k = 5$

$F_{X_k}(t) = \begin{cases} 0 & t \leq 0 \\ \frac{t}{10} & 0 < t \leq 10 \\ 1 & t > 10 \end{cases}$ $f_{X_k}(t) = \begin{cases} \frac{1}{10} & 0 < t < 10 \\ 0 & \text{poza zakresem} \end{cases}$

a) $Y_{max}\{X_1, \ldots X_5\}$

$EY = \int_{-\infty}^{\infty} t \cdot f_Y(t)\, dt = \int_0^{10} t \cdot 5 \frac{t^4}{10^5} dt = \frac{5}{10^5} \cdot \frac{t^6}{6}\Big|_0^{10} = \frac{50}{6} = 8\frac{1}{3}$

$f_Y(t) = \begin{cases} 5 \cdot \left(\frac{t}{10}\right)^4 \cdot \frac{1}{10} & t \in (0; 10) \\ 0 & \text{poza przedziałem} \end{cases}$

b) zrobić sb. w domu, wynik: $\frac{5}{3}$

LISTA 3 / ZADANIE 2

Spośród 6 płytek dobrych i 4 wadliwych losujemy jednocześnie 3 płytki.

X - liczba płytek dobrych wśród wylosowanych

Obliczyć dystrybuantę X, EX, VarX

$X \in \{0, 1, 2, 3\}$

$P(X=0) = \frac{\binom{4}{3}}{\binom{10}{3}} = \frac{4}{120} = \frac{1}{30}$

$$P(X=0)=\frac{\binom{4}{3}}{\binom{10}{3}}=\frac{4}{120}=\frac{1}{30}$$

$$P(X=1)=\frac{\binom{6}{1}\binom{4}{2}}{\binom{10}{3}}=\frac{36}{120}=\frac{3}{10}$$

$$P(X=2)=\frac{\binom{6}{2}\binom{4}{1}}{120}=\frac{1}{2}$$

$$P(X=3)=\frac{\binom{6}{3}}{120}=\frac{1}{6}$$

$$F(t)=P(X<t)=\begin{cases} 0 & t\leqslant 0\\ \frac{1}{30} & 0<t\leqslant 1\\ \frac{1}{30}+\frac{3}{10}=\frac{1}{3} & 1<t\leqslant 2\\ \frac{1}{3}+\frac{1}{2}=\frac{5}{6} & 2<t\leqslant 3\\ 1 & t>3\end{cases}$$

$$EX=\sum_{k=0}^{3}x_k\cdot P(X=x_k)=0\cdot\frac{1}{30}+1\cdot\frac{3}{10}+2\cdot\frac{1}{2}+3\cdot\frac{1}{6}=1{,}8$$

$$VarX=E(X^2)-(EX)^2=\sum_{k=0}^{3}x_k^2\,P(X=x_k)-(EX)^2=$$

$$=0^2\cdot\frac{1}{30}+1^2\cdot\frac{3}{10}+2^2\cdot\frac{1}{2}+3^2\cdot\frac{1}{6}-(1{,}8)^2=0{,}56$$

LISTA 4 / ZADANIE 3

Czas produkcji wyrobu jest zmienną losową X o funkcji gęstości

$$f(x)=\begin{cases}\frac{1}{2\sqrt{x}} & 1<x<4\\ 0 & x\leqslant 1,\ x\geqslant 4\end{cases}$$

Oblicz EX oraz Var czasu produkcji wyrobu.
Oblicz P, że wśród 15 wyrobów 8 ma czas produkcji krótszy niż $\frac{9}{4}$ s

X – czas produkcji elementu

$$EX=\int_{-\infty}^{\infty}x f(x)\,dx=\int_1^4 x\,\frac{1}{2\sqrt{x}}\,dx=\frac{1}{2}\cdot\frac{2}{3}\,x^{\frac{3}{2}}\Big|_1^4=\frac{1}{3}(8-1)=\frac{7}{3}$$

$$VarX=E(X)^2-(EX)^2=\int_{-\infty}^{\infty}x^2 f(x)\,dx-(EX)^2=\int_1^4 x^2\cdot\frac{1}{2\sqrt{x}}\,dx-\left(\frac{7}{3}\right)^2=$$

$$=\frac{1}{2}\cdot\frac{2}{5}\,x^{\frac{5}{2}}\Big|_1^4-\frac{49}{9}=\frac{1}{5}(32-1)-\frac{49}{9}=\frac{34}{45}$$

Y – liczba wyrobów z czasem produkcji krótszym niż $\frac{9}{4}$ s

$$Y\sim B(\underbrace{15}_{\text{liczba prób}},\ p)\qquad p=P\left(X<\frac{9}{4}\right)=\int_{-\infty}^{\frac{9}{4}}\frac{1}{2\sqrt{x}}\,dx=\sqrt{x}\Big|_1^{\frac{9}{4}}=\frac{3}{2}-1=\frac{1}{2}$$

$$p=\frac{1}{2}$$

$$P(Y=8)=\binom{15}{8}\left(\frac{1}{2}\right)^8\left(\frac{1}{2}\right)^7=0{,}1964$$

* na kolosa: rozkład normalny

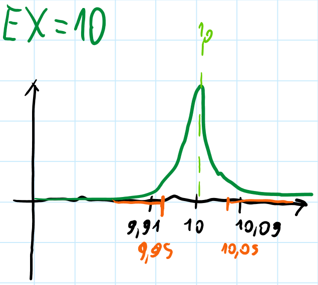
* na kolosa: rozkład normalny

LISTA 5 / ZADANIE 4

Średnica prod. elementów ma rozkład normalny $N(10;\ 0{,}03)$

Norma: $10 \pm 0{,}05$ mm

$EX = 10$



a) jaki % wyrobów nie spełnia wymogów normy?

$$P(X \geq 10{,}05 \cup X \leq 9{,}95) = P(|X-10| \geq 0{,}05) =$$

$$= 1 - P(|X-10| < 0{,}05) = 1 - P\left(\frac{-0{,}05}{0{,}03} < \underbrace{\frac{X-10}{0{,}03}}_{N(0,1)} < \frac{0{,}05}{0{,}03}\right) =$$

$$= 1 - \left(\Phi\left(\tfrac{5}{3}\right) - \Phi\left(-\tfrac{5}{3}\right) = 2\Phi\left(\tfrac{5}{3}\right) - 1\right) = 2 - 2\Phi(1{,}67) =$$

$$= 2 - 2 \cdot 0{,}95252 = 0{,}09492 \rightarrow 9{,}49\ \%$$

b) dopuszczalne σ, aby wyrobów niespełniających wymogów było co najwyżej $0{,}1\%$

$$P(|X-10| \geq 0{,}05) \leq 0{,}001$$

$$1 - P(|X-10| < 0{,}05) \leq 0{,}001$$

$$P\Big(\underbrace{\frac{(x-10)}{\sigma}}_{N(0,1)} < \frac{0{,}05}{\sigma}\Big) \geq 0{,}999$$

$$\Phi\left(\frac{0{,}05}{\sigma}\right) - \Phi\left(\frac{-0{,}05}{\sigma}\right) = 2\Phi\left(\frac{0{,}05}{\sigma}\right) - 1 \geq 0{,}999$$

$$\Phi\left(\frac{0{,}05}{\sigma}\right) \geq 0{,}9995 \geq \Phi(3{,}29)$$ → dystrybuanta z tablic

$$\frac{0{,}05}{\sigma} \geq 3{,}29 \qquad \sigma \leq 0{,}015$$

Wykład 8

środa, 17 kwietnia 2019 08:31

WEKTORY LOSOWE DWUWYMIAROWE

(X, Y) $\quad X, Y \rightarrow$ zmienne losowe

- jeśli X, Y są dyskretne, to rozkład wektora (X, Y) określa

$$\big(\underbrace{x_i, y_j ; P(X = x_i ; Y = y_j)}_{\text{rozkład łączny}}, \; i = 1 \ldots n, \; j = 1 \ldots k\big)$$

P. $X \backslash Y$	0	1	rozkład brzegowy X
0	0,8	0,01	0,81
1	0	0,07	0,07
2	0,02	0,1	0,12
rozkład brzegowy Y	0,82	0,18	= 1

- jeśli X, Y są ciągłe, określamy gęstość łączną $f(X, Y)$

$(X, Y) \in \mathbb{R}^2$

$$P((X, Y) \in A) = \iint_A f(x, y)\, dx\, dy$$

$$f_x(x) = \int_{-\infty}^{\infty} f(x, y)\, dy \qquad f_y(y) = \int_{-\infty}^{\infty} f(x, y)\, dx$$

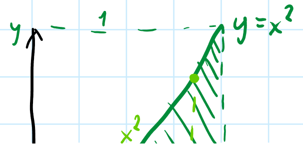
gęstość brzegowa X $\qquad$ gęstość brzegowa Y

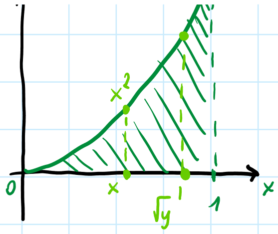
- PRZYKŁAD

Dla wektora (X, Y) o łącznej gęstości

$$f(x, y) = \begin{cases} 12xy & 0 \leq x \leq 1, \; 0 \leq y \leq x^2 \\ 0 & \text{poza} \end{cases}$$

wyznaczyć gęstości brzegowe:





$$f_x(x) = \int_{-\infty}^{\infty} f(x,y)\,dy = \begin{cases} \int_0^{x^2} 12xy\,dy = 12x\left[\frac{y^2}{2}\right]_0^{x^2} = 6x^5 \\ 0 \quad x<0 \cup x>1 \end{cases}$$

$$f_y(y) = \int_{-\infty}^{\infty} f(x,y)\,dx = \begin{cases} \int_{\sqrt{y}}^{1} 12xy\,dx = 12y\frac{x^2}{2}\Big|_{\sqrt{y}}^{1} = 6y(1-y) & y \in \langle 0;1\rangle \\ 0 \quad y<0 \quad y>1 \end{cases}$$

- KOWARIANCJĄ zmiennych losowych X, Y nazywamy liczbę

$$\text{cov}(X,Y) = E\big((X-EX)(Y-EY)\big) = E(XY) - EX \cdot EY$$

- WSPÓŁCZYNNIK KORELACJI zmiennych losowych X, Y to liczba

$$\rho(X,Y) = \frac{\text{cov}(X,Y)}{\sqrt{\text{Var}X \cdot \text{Var}Y}}$$

  - własności współczynnika korelacji:

1. $-1 \leq \rho(X,Y) \leq 1$
2. X, Y niezależne $\rightarrow \rho(X,Y) = 0$
3. $\rho(X,Y) = \pm 1 \rightarrow P(Y = aX + b) = 1$ (zmienne liniowo zależne)
4. $\rho(aX+b, cY+d) = \text{sgn}(a \cdot c) \cdot \rho(X,Y)$

funkcja znaku

- niezależność zmiennych: $f(x,y) = f_x(x) \cdot f_y(y)$

- PRZYKŁAD: wsp. korelacji dla przykładu powyżej

$$f\left(\tfrac{1}{2}, \tfrac{1}{8}\right) = \tfrac{3}{4} \neq f_x\left(\tfrac{1}{2}\right) \cdot f_y\left(\tfrac{1}{8}\right)$$

$EX = \frac{6}{7}$ $\quad EY = \frac{1}{2}$ $\quad \text{Var}X = 0{,}015$ $\quad \text{Var}Y = 0{,}05$

$$\text{cov}(X,Y) = E(X,Y) - EX \cdot EY$$

$$\mathrm{cov}(X,Y) = E(X,Y) - EX \cdot EY$$

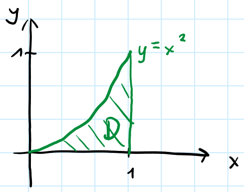
$$E(X,Y) = \iint_R xy\, f(x,y)\,dx\,dy = \int_0^1 dx \int_0^{x^2} xy \cdot 12xy\,dy = \int_0^1 12x^2 \int_0^{x^2} y^2\,dy =$$

$$= \int_0^1 12x^2 \left[\frac{y^3}{3}\right]_0^{x^2} dx = 4\int_0^1 x^8\,dx = \frac{4}{9}$$

$$\rho(X,Y) = \frac{\frac{4}{9} - \frac{6}{7}\cdot\frac{1}{2}}{\sqrt{0{,}05 \cdot 0{,}15}} = 0{,}58$$

PRZYKŁAD DO POWYŻSZEGO

$$f(x,y) = \begin{cases} 12xy, & 0 \leq x \leq 1,\ 0 \leq y \leq x^2 \\ 0 & \text{poza} \end{cases}$$



$$f_X(x) = \begin{cases} 6x^5 & x \in \langle 0,1 \rangle \\ 0 & \text{poza} \end{cases}$$

KRZYWA REGRESJI

- (X,Y) – X,Y zmienne losowe niezależne
- $f(X,Y)$ – gęstość łączna
- gęstością warunkową zmiennej losowej Y gdy $X = x$ nazywamy funkcję
  $$f_{(Y|X)}(y|x) = \frac{f(x,y)}{f_X(x)}$$
- WARUNKOWĄ WARTOŚCIĄ OCZEKIWANĄ zmiennej losowej Y względem X, gdy $X = x$ nazywamy liczbę
  $$E(Y|X=x) = \int_{-\infty}^{\infty} y f_{(Y|X)}(y|x)\, dy$$
  zależną od X
- FUNKCJĄ REGRESJI (zmiennej losowej Y względem X) nazywamy funkcję
  $$m_y(x) = E(Y|X=x)$$
  - jeżeli X,Y są niezależne, to $E(Y|X=x) = EY$ dla każdego x
  - dla dowolnej funkcji $h(x)$ zachodzi
    $$\min E(Y-h(X))^2 = E(Y-E(Y|X))^2$$
  - PRZYKŁAD: Wyznaczyć funkcję regresji zmiennej losowej Y względem zmiennej losowej X dla danych z ostatniego przykładu

$$f_{(Y|X)}(y|x) = \frac{f(x,y)}{f_X(x)} = \frac{12xy}{6x^5} = \frac{2y}{x^4} \qquad (x,y) \in D \qquad x \neq 0$$

$$E(Y|X=x) = \int_{-\infty}^{\infty} y f_{(Y|X)}(y|x)\, dy = \int_0^{x^2} y \frac{2y}{x^4}\, dy = \frac{2}{x^4} \cdot \frac{y^3}{3}\Big|_{y=0}^{x^2} = \frac{2}{x^4}\left(\frac{x^6}{3} - 0\right) = \frac{2}{3}x^2$$

$$m_y(x) = \begin{cases} \frac{2}{3}x^2 & x \in (0,1) \\ 0 & \text{poza} \end{cases}$$

$$m_y(x) = \begin{cases} \frac{2}{3}x^2 & x \in (0,1) \\ 0 & \text{poza} \end{cases}$$

np. dla $x = \frac{1}{2}$ $\rightarrow$ $m_y(x) = m_y\left(\frac{1}{2}\right) = \frac{2}{3}\left(\frac{1}{2}\right)^2 = \frac{1}{6}$

WSTĘPNA ANALIZA DANYCH STATYSTYCZNYCH

- PRÓBA PROSTA LOSOWA długości n $(n \in \mathbb{N})$ to układ n niezależnych zmiennych losowych o jednakowym rozkładzie

$(X_1, X_2 \dots X_n)$, X_k $k=1,\dots,n$ – taki sam rozkład prawdopodobieństwa

$(x_1, x_2 \dots x_n)$ x_n – obserwacja (realizacja) próby losowej $(X_1 \dots X_n)$

- PRZYKŁAD: w jednakowych 27 próbkach dielektryka określono zawartość tytanianu baru i otrzymano następujące wyniki (w mg):

x_1 8.31, 8.33, 8.36, 8.41, 8.42, 8.42, 8.44, 8.46, 8.47, 8.48, 8.48, 8.48, 8.48, 8.49, 8.50, 8.50, 8.50, 8.51, 8.55, 8.56, 8.58, 8.58, 8.62, 8.62, 8.65, 8.70, 8.70 x_{27}

X_k – zawartość tytanianu baru w k-tej próbce, $k = 1 \dots 27$

- SZEREG ROZDZIELCZY

x_{min}, x_{max} $\qquad x_{min} = 8{,}31 \qquad x_{max} = 8{,}70$

$r = x_{max} - x_{min}$ $\qquad r = 0{,}39 \qquad h = \frac{0{,}39}{5} \approx 0{,}08$

k – liczba klas dla $n \leq 50$:
$\frac{\sqrt{n}}{2} \leq k \leq \sqrt{n}$

k powinno być ≥ 5 i ≤ 30

h – szerokość klasy

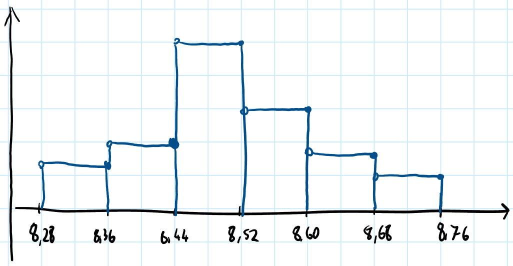
$h = \frac{r}{k}$

dla $n = 27$, $k = 5$ ← w rzeczywistości 1 klasa więcej, bo rozszerzyliśmy przedziały szersze niż zakładane

chcemy, żeby 1. obserwacja była gdzieś kawałek za środkiem klasy, nie na samym początku

klasa (i)	$(8.28, 8.36]$	$(8.36, 8.44]$	$(8.44, 8.52]$	$(8.52, 8.60]$	$(8.60, 8.68]$	$(8.68, 8.76]$
liczebność klasy n_i	3	4	11	4	3	2
częstość klasy $\frac{n_i}{n}$	$\frac{3}{27}$	$\frac{4}{27}$	$\frac{11}{27}$	$\frac{4}{27}$	$\frac{3}{27}$	$\frac{2}{27}$

- HISTOGRAM – graficzne przedstawienie szeregu rozdzielczego



- funkcję określoną na próbie losowej $(X_1, X_2 \dots X_n)$ nazywamy STATYSTYKĄ

- WAŻNIEJSZE STATYSTYKI:

① Średnia z próby

$\bar{X} = \frac{1}{n}\sum_{i=1}^{n} X_i$ $\qquad \bar{x} =$

② Wariancja z próby

$S^2 = \frac{1}{n}\sum_{i=1}^{n}(X_i - \bar{X})^2 = \left(\frac{1}{n}\sum_{i=1}^{n} X_i^2\right) - (\bar{X})^2$

$S_1^2 = \frac{1}{n-1}\sum_{i=1}^{n}(X_i - \bar{X})^2$

} zależne od autora pakietu

③ Dyspersja z próby (odchylenie standardowe)

$S = \sqrt{S^2} \qquad S_1 = \sqrt{S_1^2}$

④ Moment rzędu k

$M_k = \frac{1}{n}\sum_{i=1}^{n} X_i^k$

moment centralny rzędu k : $C_k = \frac{1}{n}\sum_{i=1}^{n}(X_i - \bar{X})^k$

⑤ Współczynnik zmienności

$v = \frac{S}{\bar{X}} \cdot 100\%$

⑥ Mediana z próby

ustawiamy próbę $(X_1, X_2 \ldots X_n)$ w ciąg niemalejący $X_n \leq X_{n+1} \leq X_{n+2} \ldots$

$Me = \begin{cases} X_{\left(\frac{n+1}{2}\right)} & n \text{ nieparzyste} \\ \frac{X_{\frac{n}{2}} + X_{\left(\frac{n}{2}\right)+1}}{2} & n \text{ parzyste} \end{cases}$

⑦ Kwartyl dolny Q_1

Q_1 – mediana z obserwacji $\leq Me$

⑧ Kwartyl górny Q_3

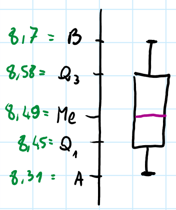
Q_3 – środkowa (mediana) z obserwacji $> Me$

w przedziale $[Q_1, Q_3]$ jest około połowy obserwacji

⑨ Odstęp międzykwartylowy

$IRQ = Q_3 - Q_1$

⑩ Wykres ramkowy



$B = \min\{X_{max}, Q_3 + \frac{3}{2} IRQ\}$

$A = \max\{X_{min}, Q_1 - \frac{3}{2} IRQ\}$

Obserwacje odstające – obserwacje poza przedziałem $[A, B]$

• obliczenia dla przykładu:

$\bar{x} = \frac{1}{27}(8{,}31 \ldots + 8{,}70) = 8.503$

- obliczenia dla przykładu:

$\bar{x} = \frac{1}{27}(8{,}31 \ldots + 8{,}70) = 8.503$

$S^2 = \frac{1}{27}\left((8{,}31)^2 + \ldots + (8{,}70)^2\right) - (8.503)^2 = 0{,}011$

$S = 0{,}104$

$V = \frac{S}{\bar{x}} \cdot 100\% = \frac{0{,}104}{8{,}508} \cdot 100\% = 1{,}2\%$

$Me = X_{14} = 8.49$

$Q_1 = \frac{X_7 + X_8}{2} = 8{,}45$

$Q_3 = X_{21} = 8{,}58$

$A = \max\{8.31,\ 8.45 - \frac{3}{2} \cdot 0{,}13\} = 8{,}31$

$B = \min\{8.7,\ 8.58 + \frac{3}{2} \cdot 0{,}13\} = 8{,}7$

Wykład 10

środa, 8 maja 2019 07:33

KOLOKWIUM POPRAWKOWE

- 27.06 czwartek

- reguły matematyczne w statystyce – sprawdzają się, ale nie ma dowodów

ESTYMACJA PARAMETRÓW ROZKŁADU PRAWDOPODOBIEŃSTWA ZMIENNEJ LOSOWEJ

- niech (X_1, X_2, X_n) jest próbą losową, a X_k ma nieznany rozkład $F(x)$
    - aproksymacja – nielosowa; estymacja – losowość
    - estymacja parametrów tego rozkładu zmiennej losowej w oparciu o obserwacje to temat naszych zajęć

$X_k \rightarrow$ rozkład $F(x)$ nieznany

$\Theta \rightarrow$ parametr rozkładu

a) $X_k \sim B(1, p)$ $\quad \Theta = p$

b) $X_k \sim N(m, \sigma)$ $\quad \Theta = (m, \sigma)$ $\quad$ (p, m, σ nieznane)

c) X_k ma rozkład $F(x)$ $\quad \Theta = EX_k$ dla $\Theta = Var X_k$ $\quad \Theta = Me X_k$ ....

- ESTYMATOREM PARAMETRU Θ związanego z rozkładem prawdopodobieństwa $F(x)$ nazywamy dowolną funkcję (którą będziemy oznaczać $\hat{\Theta}_n$) próby losowej $(X_1, X_2 \dots X_n)$

$$\hat{\Theta}_n = f(X_1, \dots, X_n)$$

- estymator $\hat{\Theta}_n$ jest ZGODNY dla parametru Θ jeśli spełnia warunek

$$\bigwedge_{\varepsilon > 0} \lim_{n \to \infty} P(|\hat{\Theta}_n - \Theta| < \varepsilon) = 1$$

- jeśli $(X_1 \dots X_n)$ jest próbą prostą i $\Theta = EX_k$, $k = 1 \dots n$, to:

1. $X_k \sim B(1, p)$ to $\bar{X}$ jest estymatorem zgodnym dla p
2. X_k ma rozkład Poissona to $\bar{X}$ jest estymatorem zgodnym dla $\lambda = EX_k$
3. $X_k \sim N(m, \sigma)$ to $\bar{X}$ jest estymatorem zgodnym dla m
4. X_k ma rozkład o $EX_k = \Theta$ to $\bar{X}$ jest estymatorem zgodnym dla Θ

Fakt ten wynika ze słabego prawa wielkich liczb

- ZADANIE DOMOWE:

Sprawdzić, że jeśli $(X_1 \dots X_n)$ to próba prosta o $X_k \sim N(m, \sigma)$ $k = 1 \dots n$, m – znane, to

- ZADANIE DOMOWE:
Sprawdzić, że jeśli $(X_1 \dots X_n)$ to próba prosta o $X_k \sim N(m, \sigma)$ $k = 1 \dots n$, m - znane, to
$$S^2 = \frac{1}{n} \sum_{k=1}^{n} (X_k - m)^2$$ jest estymatorem zgodnym dla σ^2

- estymator $\hat{\Theta}_n$ jest NIEOBCIĄŻONY dla parametru Θ jeśli
$$E\hat{\Theta}_n = \Theta$$

- PRZYKŁAD:
Niech $(X_1 \dots X_n)$ - próba prosta, $X_k \sim N(\overset{\text{nieznane}}{m}, \sigma)$. Który z estymatorów

① $\hat{\Theta}_1 = \bar{X}$ $\quad E\hat{\Theta}_1 = E\bar{X} = m$ nieobciążony

② $\hat{\Theta}_2 = \frac{1}{n-2} \sum_{k=1}^{n} X_k$ $\quad E\hat{\Theta}_2 = E\left(\frac{1}{n-2} \sum_{k=1}^{n} X_k\right) = \frac{1}{n-2} \sum_{k=1}^{n} \underbrace{EX_k}_{=m} = \frac{1}{n-2} \sum_{k=1}^{n} m = \frac{1}{n-2} \cdot n \cdot m \neq m$ obciążony

③ $\hat{\Theta}_3 = X_1$ $\quad E\hat{\Theta}_3 = EX_1 = m$ nieobciążony

④ $\hat{\Theta}_4 = \frac{1}{n-1} \sum_{k=1}^{n-1} X_k$ $\quad E\hat{\Theta}_3 = E\left(\frac{1}{n-1} \sum_{k=1}^{n-1} X_k\right) = \frac{1}{n-1} \sum_{k=1}^{n-1} EX_k = \frac{1}{n-1} \cdot (n-1) \cdot m = m$ nieobciążony

jest nieobciążony dla m?

- jeśli $\hat{\Theta}$, $\tilde{\Theta}$ są nieobciążone dla parametru Θ, to ten jest LEPSZY, który ma mniejszą wariancję

- PRZYKŁAD (POPRZEDNI)

① $Var\,\hat{\Theta}_1 = Var\,\bar{X} = Var\left(\frac{1}{n} \sum_{k=1}^{n} \overset{\text{niezależne}}{X_k}\right) = \frac{1}{n^2} \sum_{k=1}^{n} \underbrace{Var\,X_k}_{\sigma^2} = \frac{1}{n^2} \cdot n\,\sigma^2 = \frac{\sigma^2}{n}$

② —

③ $Var\,\hat{\Theta}_3 = Var\,X_1 = \sigma^2$

④ $Var\,\hat{\Theta}_4 = \frac{1}{(n-1)^2} (n-1)\,\sigma^2 = \frac{\sigma^2}{n-1}$

$$\frac{\sigma^2}{n} < \frac{\sigma^2}{n-1} < \sigma^2$$

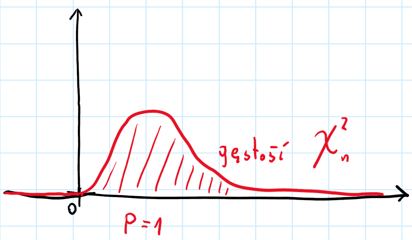
- BŁĄD ŚREDNIOKWADRATOWY estymatora $\hat{\Theta}_n$ to liczba
$$E(\hat{\Theta}_n - \Theta)^2 = Var\,\hat{\Theta}_n + \underbrace{(E\hat{\Theta}_n - \Theta)^2}_{\text{kwadrat obciążenia } \hat{\Theta}_n}$$

- jeśli $(X_1 \dots X_n)$ - niezależne zmienne losowe, $X_k \sim N(0,1)$, $k = 1 \dots n$ to rozkład sumy kwadratów tych zmiennych $X_1^2 + X_2^2 + \dots X_n^2$ nazywamy rozkładem CHI-KWADRAT Z N STOPNIAMI SWOBODY

$$\chi_n^2 = \sum_{k=1}^{n} X_k^2$$

$$E\chi_n^2 = n$$

$$Var\,\chi_n^2 = 2n$$



- niech X, Y - niezależne, $X \sim N(0,1)$, $Y \sim \chi^2_n$ - wtedy rozkład zmiennej losowej t_n

$$t_n = \frac{X}{\sqrt{Y}} \sqrt{n}$$

nazywamy ROZKŁADEM T-STUDENTA Z N STOPNIAMI SWOBODY

$$E t_n = 0$$

$$\operatorname{Var} t_n = \frac{n}{n-2}, \quad n > 2$$

- TWIERDZENIE FISHERA: jeśli $(X_1, X_2 \dots X_n)$ to próba prosta taka, że $X_k \sim N(m, \sigma)$ $(m, \sigma$ - nieznane$)$ to:

$$\bar{X} \;;\; S^2 = \frac{1}{n} \sum_{k=1}^{n} (X_k - \bar{X})^2$$

są niezależne.

- ponadto $\bar{X} \sim N(m, \frac{\sigma}{\sqrt{m}})$ a $\frac{nS^2}{\sigma^2}$ ma rozkład χ^2_{n-1}

$$\sum_{k=1}^{n} (X_k - \bar{X}) = 0$$

- PRZYKŁAD:
$(X_1 \dots X_n)$ próba prosta $X_n \sim N(m, \sigma)$, m, σ - niezależne.
Który z estymatorów S^2 czy $S_1^2 = \frac{1}{n-1} \sum_{k=1}^{n} (X_k - \bar{X})^2$ jest nieobciążony dla σ^2?

$\frac{nS^2}{\sigma^2}$ ma rozkład χ^2_{n-1}

$$E \frac{nS^2}{\sigma^2} = E\chi_{n-1} = n-1$$

$$\frac{n}{\sigma^2} ES^2 = n-1$$

$$ES^2 = \frac{n-1}{n} \sigma^2 \neq \sigma^2$$ (nie nieobciążony)

$$nS^2 = (n-1) S^2$$

$$ES_1^2 = \frac{n}{n-1} ES^2 = \frac{n}{n-1} \cdot \frac{n-1}{n} \cdot \sigma^2 = \sigma^2$$ (nieobciążony)

- METODY WYZNACZANIA ESTYMATORÓW:

① METODA NAJWIĘKSZEJ WIARYGODNOŚCI (MLE)

Określamy funkcję parametru Θ

$$L(x_1, x_2 \dots x_n, \Theta) = \begin{cases} \prod_{k=1}^{n} P_\Theta (X_k = x_k) & X_k - \text{dyskretna} \\ \prod_{k=1}^{n} f_\Theta (x_k) & X_k - \text{ciągła} \end{cases}$$

x_k - zaobserwowana wartość X_k

$$\prod_{i=1}^{n} a_i = a_1 \cdot a_2 \cdot \ldots \cdot a_n$$

którą nazywamy FUNKCJA WIARYGODNOŚCI

- estymatorem metody największej wiarygodności dla parametru Θ nazywamy taką wartość $\hat{\Theta}$, że

$$\max L(x_1 \dots x_n, \Theta) = L(x_1 \dots x_n, \hat{\Theta})$$

- uwagi:
  - wyznaczamy $\max \ln L(x_1 \dots x_n, \Theta)$
  - estymatory MLE są nieobciążone, zgodne, o minimalnej wariancji przynajmniej asymptotycznie, czyli gdy $n \to \infty$

PRZYKŁAD

Wyznaczyć estymatory metody największej wiarygodności, gdy $(X_1, X_2 \dots X_n)$ to próba prosta dla parametru Θ

a) $X_k \sim B(l, p)$ $p \in (0,1)$ p nieznane $\Theta = p$

$$L(x_1, x_2 \dots x_n, p) = \prod_{k=1}^{n} P_p(X_k = x_k) = \prod_{k=1}^{n} \binom{l}{x_k} p^{x_k} (1-p)^{l-x_k}$$

$x_1, x_2 \dots x_n \to$ obserwacje próby losowej $(X_1, X_2 \dots X_n)$

$$P(X_k = x_k) = \binom{l}{x_k} p^{x_k} (1-p)^{l-x_k}$$

$$\ln L(X_1 \dots X_n, p) = \sum_{k=1}^{n} \left(\ln \binom{l}{x_k} + x_k \ln p + (l - x_k) \ln(1-p) \right)$$

$$\frac{\partial}{\partial p} \ln L(x_1 \dots x_n, p) = \sum_{k=1}^{n} \left(x_k \frac{1}{p} + (l - x_k) \frac{(-1)}{1-p} \right) = 0 \qquad / \cdot p(1-p) \neq 0$$

$$\sum_{k=1}^{n} \left(x_k (1-p) + (l - x_k)(-1) p \right) = 0$$

$$\sum_{k=1}^{n} \left(x_k - x_k p - lp + x_k p \right) = 0$$

$$\left(\sum_{k=1}^{n} x_k \right) - nlp = 0$$

$$p = \frac{\sum_{k=1}^{n} x_k}{nl} = \frac{\overline{x}}{l}$$

lokalne maksimum jest największą wartością funkcji

$$\frac{\partial^2 \ln L}{\partial p^2} = \sum_{k=1}^{n} \left(x_k \frac{-1}{p^2} + (l - x_k) \frac{-1}{(1-p)^2} \right) < 0$$

$x_k \in \{0, 1 \dots l\}$

- niech $X_k \sim B(4, p)$, p nieznane. Obliczyć estymator metody MLE dla obserwacji: 2, 1, 3, 0, 2

$n = 5$ $l = 4$

$$\hat{p} = \frac{\overline{x}}{l} = \frac{\frac{8}{5}}{4} = \frac{2}{5} = 0{,}4$$

b) LISTA 8, zadanie 6

$L(x_1, x_2 \dots x_n, p)$ X_k ma gęstość $f_c(x) = \begin{cases} \frac{c}{x^{c+1}} & x > 1 \\ 0 & x \leq 1 \end{cases}$

$\Theta = c$

$L(\dots) = \prod_{k=1}^{n} f(\dots) = \begin{cases} \prod_{k=1}^{n} \frac{c}{x_k^{c+1}} & x_k > 1 \text{ dla } k = 1, 2 \dots \Theta = c \end{cases}$

$\Theta = c$

$$L(x_1, x_2 \dots x_n, p) = \prod_{k=1}^{n} f_c(x_k) = \begin{cases} \prod_{k=1}^{n} \frac{c}{x_k^{c+1}} & x_k > 1 \quad \text{dla } k = 1, 2 \dots \quad \Theta = c \\ 0 & \text{przynajmniej jedno } x_k \leq 1 \end{cases}$$

$$\ln L(x_1 \dots x_n) = \sum_{k=1}^{n} (\ln c - (c+1) \ln x_k) \qquad x_k > 1 \qquad k = 1 \dots n$$

$$\frac{\partial}{\partial c} \ln L(x_1 \dots x_n, c) = \sum_{k=1}^{n} \left(\frac{1}{c} - \ln x_k \right) = \frac{n}{c} - \sum_{k=1}^{n} \ln x_k = 0$$

$$c = \frac{n}{\sum_{k=1}^{n} \ln x_k}$$ ← lokalne maksimum w którym jest największa wartość funkcji $L(X_1 \dots X_n, c)$

$$\frac{\partial^2}{\partial c^2} \ln L = \sum_{k=1}^{n} \frac{-1}{c^2} < 0$$

- niech $(X_1 \dots X_n)$ to próba prosta, X_k ma gęstość $f_c(x)$

Dla obserwacji 2, 4, 6, 8, 7 wyznaczyć estymator MLE dla parametru c.

$n = 5$

$$c = \frac{n}{\sum_{k=1}^{n} \ln(x_k)} = \frac{5}{\ln 2 + \ln 4 + \ln 6 + \ln 8 + \ln 7} = \frac{5}{\ln(2 \cdot 4 \cdot 6 \cdot 8 \cdot 7)} = \frac{5}{7,9} = 0,63$$

ZADANIE DOMOWE

Niech $(X_1 \dots X_n)$ to próba prosta, że $X_k \sim N(m, \sigma)$

Wyznaczyć estymator metody największej wiarygodności dla $\Theta = (m, \sigma)$.

WSTĘP DO ZADANIA:

- funkcja wiarygodności: $L(x_1 \dots x_n, m, \sigma) = \prod_{k=1}^{n} \frac{1}{\sqrt{2\pi}\,\sigma} e^{-\frac{(x_k - m)^2}{2\sigma^2}}$
- szukamy max wartości funkcji

② METODA MOMENTÓW

Polega na przyrównaniu momentów rozkładu zmiennej losowej z jej momentem empirycznym, tzn.

np. $EX_k = \bar{X}$

$$\mathrm{Var}\, X_k = S^2$$

$$EX_k^l = \frac{1}{n} \sum_{k=1}^{n} X_k^l$$

- PRZYKŁAD: niech $X_1, X_2 \dots X_n$ → próba prosta z rozkładu jednostajnego na przedziale $[a, b]$. Wyznaczyć estymatory metody momentów dla parametrów a, b.

$$\begin{cases} EX_k = \frac{a+b}{2} = \bar{X} \\ VarX_k = \frac{(b-a)^2}{12} = S^2 \end{cases} \longrightarrow \begin{cases} b = 2\bar{X} - a \\ \frac{(2\bar{x} - a)^2}{12} = S^2 \end{cases}$$

$$\frac{(2\bar{x}-a-a)^2}{12} = S^2 \qquad \frac{(\bar{x}-a)^2}{3} = S^2 \qquad (\bar{x}-a)^2 = 3S^2$$

$$\bar{x} - a = \sqrt{3}s \quad \cup \quad \bar{x} - a = -\sqrt{3}s$$

$$\begin{cases} \hat{a} = \bar{x} - \sqrt{3}s \\ \hat{b} = \bar{x} + \sqrt{3}s \end{cases} \cup \begin{cases} \hat{a} = \bar{x} + \sqrt{3}s \\ \hat{b} = \bar{x} - \sqrt{3}s \end{cases}$$

- PRZYKŁAD: Wyznaczyć estymatory metody momentów gdy $(X_1 \dots X_n) \rightarrow$ próba prosta, X_k ma rozkład jednostajny na $[a, b]$, $a, b \rightarrow$ nieznane, dla obserwacji $(1.2, 2.1, 1.5, 1.8)$.

$$\bar{x} = \frac{1}{4}(1,2+2,1+1,5+1,8) = 1,7$$

$$S^2 = 0,115$$

$$\hat{a} = 1,7 - \sqrt{3} \cdot \sqrt{0,115} = 1,11$$

$$\hat{b} = 1,7 + \sqrt{3} \cdot \sqrt{0,115} = 2,29$$

ESTYMACJA PRZEDZIAŁOWA PARAMETRÓW ROZKŁADU ZMIENNEJ LOSOWEJ

- $(X_1, X_2 \dots X_n) \longrightarrow$ próba prosta

- nie chcemy przybliżyć nieznanego parametru liczbami, chcemy podać przedział liczbowy, który z zadanym przez nas dużym prawdopodobieństwem nakryje nieznany parametr liczbowy

- PRZEDZIAŁEM UFNOŚCI dla parametru Θ na poziomie ufności $1-\alpha$ (α jest mała liczba, najczęściej $0.05, 0.02, 0.1, 0.01$) nazywamy przedział liczbowy (Θ_1, Θ_2) taki, że:

  1. $\Theta_i = f(X_1, X_2 \dots X_n) \qquad i = 1,2$
  2. przedział (Θ_1, Θ_2) nakrywa nieznany parametr Θ z prawdopodobieństwem

$$P(\Theta_1 < \Theta < \Theta_2) = \underbrace{1-\alpha}_{}$$ poziom ufności

- PRZYKŁAD: niech $(X_1, X_2 \dots X_n) \rightarrow$ próba prosta z rozkładu $N(m, \sigma)$, $\sigma \rightarrow$ znane. Wyznaczyć przedział ufności dla parametru m na poziomie ufności $1-\alpha$.

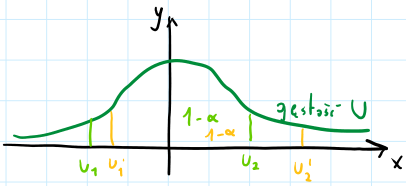
$\bar{X} \sim N\left(m, \frac{\sigma}{\sqrt{n}}\right)$ } standaryzacja

$U = \frac{\bar{X}-m}{\frac{\sigma}{\sqrt{n}}} \sim N(0,1)$

takich przedziałów jest nieskończenie wiele
najkrótszy z nich jest symetryczny wzgl. 0,
$u_1 = -u_2$

znajdźmy u_1, u_2 takie, że $P(u_1 < U < u_2) = 1-\alpha$

dla $1-\alpha > 0,5$



$$P(-u_\alpha < U < u_\alpha) = \Phi(u_\alpha) - \Phi(-u_\alpha) = -1 + 2\underbrace{\Phi(u_\alpha) = 1-\alpha}_{\Phi(u_\alpha) = 1-\frac{\alpha}{2}}$$

$$P\left(-u_\alpha < \frac{\bar{X}-m}{\sigma}\sqrt{n} < u_\alpha\right) = 1-\alpha$$

$$P\left(\bar{X} - \frac{\sigma u_\alpha}{\sqrt{n}} < m < \bar{X} + \frac{\sigma u_\alpha}{\sqrt{n}}\right) = 1-\alpha$$

- PRZYKŁAD: czas analizy ma rozkład normalny $N(m,2)$.
  Dla obserwacji próby prostej 34, 35, 30, 32, 34, 31, 34, 33, 35.

Wyznaczyć przedział ufności dla wartości oczekiwanej czasu analizy (średniej czasu analizy, przeciętnej czasu analizy) na poziomie ufności $1-\alpha = 0,95$

X_k – czas k-tej analizy, $k = 1 \dots 9$ $\quad n = 9$

$X_k \sim N(m,2)$ $\quad \sigma = 2$

$EX = m$

$$\bar{X} - \frac{\sigma u_\alpha}{\sqrt{n}} < m < \bar{X} + \frac{\sigma u_\alpha}{\sqrt{n}}$$

$$\Phi\, u_\alpha = 1 - \frac{\alpha}{2} = 1 - \frac{0,05}{2} = 0,975 \underset{\text{tablice}}{=} \Phi(1,96)$$

$u_\alpha = 1,96$

$\bar{X} = \frac{1}{9}(34 + \dots + 35) = 33,1$

$$31,8 = 33,1 - \frac{2 \cdot 1,96}{\sqrt{9}} < m < 31,1 + \frac{2 \cdot 1,96}{\sqrt{9}} = 34,4$$

ODP:
Przedział $(31,8 \,,\, 34,4)$ zawiera parametr m z prawdopodobieństwem 0,95

- do kolokwium: wydrukować tabele ze strony

- parametr – wielkość związana z badaną wartością
- poziom ufności → przedział, który z dużym prawdop. pokryje szukaną wartość

PRZYKŁAD:

Dla jakiego n długość przedziału jest mniejsza od l?

$$\frac{2\sigma u_\alpha}{\sqrt{n}} < l \qquad n > \left(\frac{2\sigma u_\alpha}{l}\right)^2$$

PRZYKŁAD:

Zmierzono rezystancję 10 jednakowych elementów losowo wybranych z produkcji i otrzymano wielkości:

11.4, 9.8, 10.2, 10.5, 10.9, 9.9, 11.2, 10.0, 10.7, 10.4

Zakładając, że obserwacje pochodzą z rozkładu normalnego, wyznaczyć przedział ufności:

a) dwustronny
b) jednostronny typu $(c, +\infty)$

dla średniej rezystancji elementów na poziomie ufności $1-\alpha = 0.9$

przedział ufności dla parametru „m"

X_k - rezystancja elementu $\quad k = 1 \dots 10$

$X_k \sim N(m, \sigma) \quad m, \sigma$ – nieznane

a) $\bar{X} - \frac{s t_\alpha}{\sqrt{n-1}} < m < \bar{X} + \frac{s t_\alpha}{\sqrt{n-1}}$

$\bar{x} = \frac{1}{10}(11.4 + \dots + 10.4) = 10.50$

$s^2 = \frac{1}{10}\sum_{i=1}^{10}(x_i - 10{,}5)^2 = 0{,}27$

$s = \sqrt{s^2} = \sqrt{0{,}27} = 0{,}52$

t_α → odczytujemy z tablic t_{n-1}, rozkład Studenta

$t_\alpha \longrightarrow v = n-1 = 9 \longrightarrow p = \frac{\alpha}{2} = 0{,}05$

$t_\alpha = 1{,}833$

$10{,}5 - \frac{0{,}52 \cdot 1{,}833}{\sqrt{9}} < m < 10{,}5 + \frac{0{,}52 \cdot 1{,}833}{\sqrt{9}}$

$$10{,}18 < m < 10{,}82$$

ODP: Na poziomie ufności 0,9, przedział ufności ma końce (10,18 ; 10,82)

b) jednostronne dla m

$$m > \bar{X} - \frac{s\, t_\alpha}{\sqrt{n-1}} \qquad t_\alpha \to \text{odczyt z } t_{n-1}$$

$v = n-1$ $\quad p = \alpha$

$p = 0{,}1$ $\quad v = 9$

$t_\alpha = 1{,}385$

$$m > 10{,}5 - \frac{0{,}52 \cdot 1{,}385}{\sqrt{9}} \rightarrow m > 10{,}26$$

- klasa przyrządu → „dyspersja mnożona od" (przy pomiarze)
- klasa przyrządu związana jest z odchyleniem standardowym wyniku otrzymanego przyrządem

PRZYKŁAD

W celu zbadania woltomierza wykonano nim 9 pomiarów wzorcowych napięć i otrzymano wyniki:

11, 12, 12, 11, 10, 11, 12, 13, 12

Przyjmując, że błędy pomiaru mają rozkład normalny, wyznaczyć przedział ufności dla

a) wariancji błędu pomiaru
b) dyspersji błędu pomiaru

na poziomie ufności $1-\alpha = 0{,}95$

X_k - k-ty pomiar woltomierzem, $k = 1 \dots 9$

$X_k \sim N(m, \sigma)$ $\quad m, \sigma$ - nieznane

a) $$\frac{nS^2}{c_2} < \sigma^2 < \frac{nS^2}{c_1}$$

stałe c_1, c_2 bierzemy z tablic rozkładu chi kwadrat

$$\frac{5{,}298}{17{,}53} < \sigma^2 < \frac{5{,}298}{2{,}18}$$

$$0{,}302 < \sigma^2 < 2{,}43$$

b) $$\sqrt{0{,}302} < \sigma < \sqrt{2{,}43}$$

$$0{,}549 < \sigma < 1{,}559$$

$\bar{X} = 11{,}56$
$S^2 = 0{,}5886$
$nS^2 = 5{,}298$
$v = 9-1 = 8$
$p = \frac{\alpha}{2} = 0{,}025$ $\quad c_2 = 17{,}53$
$p = 1 - \frac{\alpha}{2} = 0{,}975$ $\quad c_1 = 2{,}18$

- kolejne uwagi do tablic:
  - jeśli $n > 30$ → $t_\alpha = u_\alpha$ z modelu 1 [model 2]
  - jeśli $n > 30$ → nie musi być założenia, że X_k mają rozkład normalny [model 2,3]

PRZYKŁAD 3:

Wśród 400 mieszkańców przeprowadzono ankietę „czy chcesz, aby w twoim regionie zbudowano elektrownię jądrową?" było 380 odpowiedzi „NIE". Na poziomie ufności $1-\alpha = 0{,}94$ wyznaczyć przedział ufności dla „NIEPOPIERANIA" budowy elektrowni.

X_k – wynik k-tej ankiety $\quad k = 1..400$

$X_k \sim B(1, p) \qquad p = P(X=1) = P(\text{„NIE"})$

$$\frac{m}{n} - u_\alpha \sqrt{\frac{m}{n^2}\left(1-\frac{m}{n}\right)} < p < \frac{m}{n} + u_\alpha \sqrt{\frac{m}{n^2}\left(1-\frac{m}{n}\right)}$$

$n = 400 \qquad m = 380$

$\Phi(u_\alpha) = 1 - \frac{\alpha}{2} = 0{,}97 \qquad u_\alpha = 1{,}88$

$0{,}93 < p < 0{,}97$ → % niepopierających elektrowni

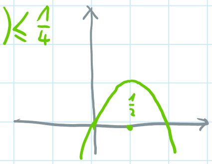
Dodatkowe pytanie: dla ilu obserwacji długość przedziału będzie mniejsza niż 0,03 ?

PRZYPADEK OGÓLNY

- długość przedziału: $2 u_\alpha \frac{1}{\sqrt{n}} \sqrt{\frac{m}{n}\left(1-\frac{m}{n}\right)}$ $\qquad x(1-x) \leq \frac{1}{4}$

$$2u_\alpha \frac{1}{\sqrt{n}} \cdot \frac{1}{2} = \frac{u_\alpha}{\sqrt{n}} \leq l$$

$$n \geq \left(\frac{u_\alpha}{l}\right)^2$$



Zadanie: $n \geq \left(\frac{u_\alpha}{l}\right)^2 = \left(\frac{1{,}88}{0{,}03}\right)^2 = 3927{,}1$

$n \geq 3928$

PRZYKŁAD 4

$X_k \sim N(m_1, \sigma_1) \qquad k = 1 \dots n$

$Y_k \sim N(m_2, \sigma_2) \qquad k = 1 \dots n$

X_i, Y_j → niezależne

SZUKANE: Przedział wartości dla $(m_1 - m_2)$

SZUKANE: Przedział wartości dla $(m_1 - m_2)$

- $\sigma_1, \sigma_2 \rightarrow$ znane $\rightarrow$ $U = \frac{\bar{X} - \bar{Y} - (m_1 - m_2)}{\sqrt{\frac{\sigma_1^2}{n} + \frac{\sigma_2^2}{m}}}$

p.n dla $(m_1 - m_2)$ jak dla m w modelu 1

$$U \sim N(0,1)$$

- $\sigma_1, \sigma_2 \rightarrow$ nieznane $\rightarrow$ $t_{n+m-2} = \frac{\bar{X} - \bar{Y} - (m_1 - m_2)}{\sqrt{nS_x^2 + mS_y^2}} \cdot \sqrt{\frac{n \cdot m (n+m-2)}{n+m}}$

$\sigma_1 = \sigma_2$

p.n dla $(m_1 - m_2)$ jak dla m w modelu 2

TESTOWANIE HIPOTEZ STATYSTYCZNYCH

- HIPOTEZA STATYSTYCZNA – przypuszczenie odnośnie rozkładu obserwowanej zmiennej losowej
- podział hipotez:
  - PARAMETRYCZNE – przypuszczenia odnośnie parametru związanego ze zmienną losową
  - NIEPARAMETRYCZNE – „czy obserwacje z... pochodzą z rozkładu normalnego?"
- HIPOTEZY PARAMETRYCZNE
  - $(X_1, X_2 \dots X_n)$ – próba prosta, każda X_k ma ten sam rozkład $F_\Theta(X)$, $\Theta \in \Theta$ (niezerowe)
  - hipoteza $H: \Theta \in \Theta_1$ — H – hipoteza zerowa

    $K: \Theta \in \Theta_2$ — K – hipoteza alternatywna

    $\Theta_1 \cap \Theta_2 = \emptyset$

    $\Theta_i \subset \Theta$

  - błąd I rodzaju: prawdopodobieństwo odrzucenia H, podczas gdy H jest prawdziwe
  - błąd II rodzaju: nie odrzucamy H, podczas gdy prawdziwa jest któraś z wartości w Θ_2 (zachodzi któraś z alternatyw)
  - hipotezę H weryfikujemy (sprawdzamy, testujemy) w oparciu o obserwacje $(x_1, x_2 \dots x_n)$ próby $(X_1, X_2 \dots X_n)$
    - konstruujemy tzw. zbiór krytyczny Q, żeby przy warunku

      $\bigwedge_{\Theta \in \Theta_1} P(Q) \leqslant \alpha$ — α - najczęściej 0.05, 0.02, 0.01, 0.1

      α - poziom istotności

      $\bigwedge_{\Theta \in \Theta_2} P(Q)$ było jak największe

Wykład 13

poniedziałek, 10 czerwca 2019 14:40

Nowy-Doku
ment-201...

Testowanie Hipotez c.d

Lista 6 zad. 1, 3
-1 7, 8, 9, 10 } zadania 1-6

$H: \Theta \in \Theta_1$
$K: \Theta \in \Theta_2$

Q - zbiór krytyczny

$\bigwedge_{\Theta \in \Theta_1} P(Q) \leq \alpha$

$\bigwedge_{\Theta \in \Theta_2} P(\overline{Q})$, $P(Q)$

$\Rightarrow 1 - P(Q)$

Producent twierdzi, że grubość produkowanych przez niego płytek ma rozkład normalny $N(20, 4)$. Zmierzono grubość 25 płytek wybranych losowo i obliczono, że średnia z próby $\bar{x} = 18.6$. Na poziomie istotności $\alpha = 0{,}05$ zweryfikować hipotezę, że przeciętna grubość płytek jest taka jak podaje producent.

X_k - grubość ktej płytki

X_k - grubość k-tej płytki $X_k \sim N(m, k)$

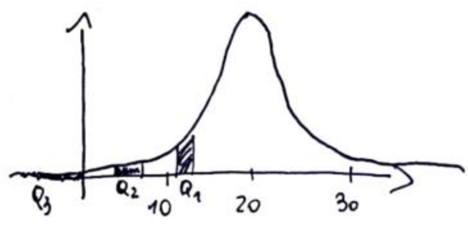
$X_k \sim N(20, 4)$ według producenta

$H: m = 20$
$K: m < 20$

$\bar{X} \sim N\left(m, \frac{4}{\sqrt{25}}\right)$

$\begin{cases} P(Q) \leq \alpha & \text{dla H prawdziwe lub } m=20 \\ P(\bar{Q}) \text{ najmniejsze} & \text{dla } m < 20 \end{cases}$

$Q = (-\infty, c)$



$\alpha = P(Q) = P(\bar{x} < c) = P\left(\frac{\bar{x}-20}{\frac{4}{5}} < \frac{c-20}{\frac{4}{5}}\right) = \Phi\left(\frac{c-20}{\frac{4}{5}}\right)$,

$\alpha = 0{,}05$ gdy $m = 20$

$\Phi\left(-\frac{c-20}{\frac{4}{5}}\right) = 1 - 0{,}05 = 0{,}95 = \Phi(1{,}64)$

$\frac{c-20}{4} \cdot 5 \approx -1{,}64$

$c = 18{,}69$

$Q = (-\infty, 18.69)$

$\bar{x} = 18{,}6 \in Q$ odrzucamy H.

Scanned with CamScanner

Waga wyrobu według metki to 10g. Zważono 10 losowo wybranych wyrobów i otrzymano: 9.2 10.1 10.3 9.7 9.5
10.2 9.3 9.9 10.4 9.4

Przyjmując że waga wyrobu ma rozkład normalny na poziomie istotności a) $\alpha=0.1$
b) $\alpha=0.2$

zweryfikować, że $H: m=10$ $K: m\neq 10$
czyli $m_0=10$

X_k - waga k-tego wyrobu $k=1,\dots,10$
$X_k \sim N(m, \sigma)$ σ - nieznane

$$t_{n-1} = \frac{\bar{x}-m_0}{s}\cdot\sqrt{n-1}$$

$\bar{x}$ = średnia = 9.8 , $m=10$

$s^2 = 0{,}174$

$$t_9 = \frac{9.8-10}{\sqrt{0{,}174}}\cdot 3 = -1{,}44$$

$Q = (-\infty, -t_\alpha)\cup(t_\alpha, +\infty)$ t_α $v=9$

a) $\alpha=0{,}1$ $p=\frac{\alpha}{2}=0{,}05$, $t_\alpha=1{,}833$, $Q=(-\infty,-1{,}833)\cup(1{,}833, +\infty)$

$-1{,}44 \notin Q$

Odp: Nie ma podstaw do odrzucenia H.

b) $\alpha=0{,}2$ $p=\frac{\alpha}{2}=0{,}1$ $t_\alpha=1{,}383$ $Q=(-\infty,-1{,}383)\cup(1{,}383, +\infty)$

$-1{,}44 \in Q$

Odp: Odrzucamy H na poziomie $\alpha=0{,}2$

Wykonano 8 pomiarów wielkości o rozkładzie normalnym otrzymując:

8.17 8.21 8.05 8.19 8.22 8.06

Zweryfikować hipotezę, że wariancja wyników pomiaru jest równa 0,003 na poziomie istotności $\alpha = 0{,}05$ $X_k \sim N(m, \sigma)$

$H: \sigma^2 = 0{,}003$ Statystyka chi-kwadrat

$K: \sigma^2 \neq 0{,}003$ $\chi^2_{n-1} = \frac{ns^2}{\sigma_0^2}$

$\bar{x} = 8.14$, $ns^2 = 0{,}0328$, $\chi^2_7 = \frac{0{,}0328}{0{,}003} = 10{,}93$

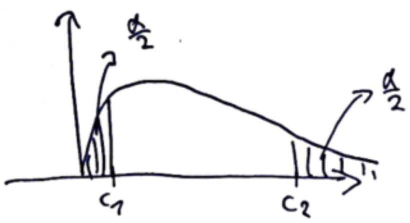
Czyli $Q = (0, c_1) \cup (c_2, +\infty)$

c_1 z tablic χ^2_7 $v = 7$, $\frac{\alpha}{2} = 0{,}025$ $c_1 = 1{,}69$

c_2 $1 - \frac{\alpha}{2} = 0{,}975$ $c_2 = 16{,}01$

$Q = (0, 1{,}69) \cup (16{,}01, +\infty)$

$\chi^2_7 = 10{,}93 \notin Q$



Odp: Nie ma podstaw do odrzucenia H.

Scanned with CamScanner

Sprawdzono symetryczność monetę rzucono 100 razy i 58 razy był orzeł.
Dla $\alpha = 0{,}05$ zweryfikować hipotezę o symetryczności monety

$H: p = \frac{1}{2}$ $\qquad$ X_k - wynik k - rzutu

$K: p \neq \frac{1}{2}$ $\qquad$ $X_k \sim B(1, p)$, $p = P(\text{orzeł})$

↳ (moneta symetryczna gdy szansa na orła/reszkę po 50%)

$$U = \frac{\frac{m}{n} - p_0}{\sqrt{p_0(1-p_0)}} \cdot \sqrt{n} = \frac{\frac{58}{100} - \frac{1}{2}}{\sqrt{\frac{1}{2} \cdot \frac{1}{2}}} \cdot \sqrt{100} = 1.6$$

$$Q = (-\infty, -u_\alpha) \cup (u_\alpha, +\infty)$$

$$\Phi(u_\alpha) = 1 - \frac{\alpha}{2} = 0{,}975 = \Phi(1{,}96)$$

$$Q = (-\infty, -1.96) \cup (1.96, +\infty)$$

$$U = 1{,}6 \notin Q$$

Nie ma podstaw do odrzucenia H.

Def. p-wartość (p-value) jest to najmniejszy poziom istotności przy którym dla zaobserwowanej wartości statystyki odrzucamy H.

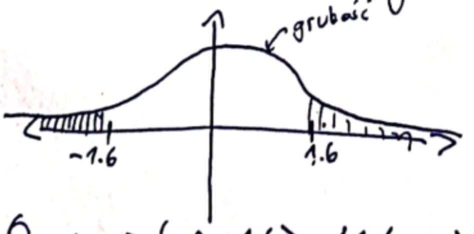
$\alpha < p$ nie ma podstaw do odrzucenia H

$\alpha \geq p$ odrzucamy H.

Przykład. Obliczyć p-wartość dla danych z ostatniego przykładu

$U = 1.6 \quad U \sim N(0,1)$

$$p = P(Q_{najmniejszy}) = P(U < -1.6 \cup U > 1.6) =$$
$$= 2 \cdot P(U < -1.6) = 2 \cdot \Phi(-1.6) = 2 \cdot [1 - \Phi(1.6)] =$$
$$= 0{,}1096$$



$$Q_{najmn} = (-\infty, -1.6\rangle \cup \langle 1.6, +\infty)$$