

MOMENTY ZMIENNEJ LOSOWEJ X

Moment rzędu k (k - liczba naturalna)

$$m_k = E(X^k)$$

Uwaga

$$D^2 X = m_2 - m^2.$$

Własność. Jeśli istnieje m_k to istnieje m_s dla każdego $s < k$.

Moment centralny rzędu k (k - liczba naturalna)

$$\mu_k = E\left((X - EX)^k\right)$$

Zauważmy, że w szczególności $\mu_1 = 0$,
 $\mu_2 = D^2X$.

współczynnik asymetrii (skośności)

$$a = \frac{\mu_3}{\sigma^3}$$

współczynnik skupienia (kurtoza)

$$k = \frac{\mu_4}{\sigma^4}$$

Wskaźnik kurtozy

$$k = \frac{\mu_4}{\sigma^4} - 3$$

Zależności między momentami zwykłymi i centralnymi.

$$m_k = \sum_{i=0}^k \binom{k}{i} \mu_{k-i} m_1^i$$

w szczególności

$$m_2 = \mu_2 + m_1^2,$$

$$m_3 = \mu_3 + 3\mu_2 m_1 + m_1^3,$$

$$m_4 = \mu_4 + 4\mu_3 m_1 + 6\mu_2 m_1^2 + m_1^4,$$

$$\mu_k = \sum_{i=0}^k (-1)^i \binom{k}{i} m_{k-i} m_1^i$$

w szczególności

$$\mu_2 = m_2 - m_1^2,$$

$$\mu_3 = m_3 - 3m_2m_1 + 2m_1^3,$$

$$\mu_4 = m_4 - 4m_3m_1 + 6m_2m_1^2 - 3m_1^4,$$

Parametry dla wybranych rozkładów

Wybrane rozkłady skokowe.

NAZWA ROZKŁADU	FUNKCJA ROZKŁADU PRAWDOPODOBIENSTWA WŁASNOŚCI	WART. OCZEKIWANA WARIANCJA INNE PARAMETRY
Rozkład jednostajny dyskretny c, n - całkowite; $n > 0$	$P(X = k) = \frac{1}{n} \quad k = c, c + 1, c + 2, \dots, c + n - 1$ (gdy $n = 1$ to rozkład jednopunktowy)	$EX = c + (n - 1)/2;$ $D^2X = (n^2 - 1)/12$ $a = 0$ $k = 1,8 - 2,4/(n^2 - 1)$
Rozkład zerojedynkowy $p \in (0, 1)$	$P(X = 0) = q \quad P(X = 1) = p; \quad q = 1 - p$	$EX = p; \quad D^2X = pq$ $a = \frac{q - p}{\sqrt{pq}} \quad k = \frac{1}{pq} - 3$
Rozkład dwumianowy $p \in (0, 1), \quad n \in \mathbb{N}$	$P(X = k) = \binom{n}{k} p^k q^{n-k} \quad q = 1 - p$ $k = 0, 1, 2, \dots, n$ X - liczba sukcesów w n próbach B.	$EX = np; \quad D^2X = npq$ $a = \frac{q - p}{\sqrt{npq}}$ $k = \frac{1 - 6pq}{npq} + 3$
Rozkład Poissona $\lambda > 0$	$P(X = k) = \frac{\lambda^k}{k!} e^{-\lambda}$ $k = 0, 1, 2, \dots$	$EX = \lambda;$ $D^2X = \lambda$ $a = \frac{1}{\sqrt{\lambda}}$ $k = \frac{1}{\lambda} + 3$ $m_3 = \lambda + 3\lambda^2 + \lambda^3,$ $m_4 = \lambda + 7\lambda^2 + 6\lambda^3 + \lambda^4$ $\mu_3 = \lambda,$ $\mu_4 = \lambda + 3\lambda^2$

Wybrane rozkłady ciągłe.

NAZWA ROZKŁADU	GĘSTOŚĆ WŁASNOŚCI	WART. OCZEKIWANA WARIANCJA INNE PARAMETRY
Rozkład jednostajny $a, b \in R \quad a < b$	$f(x) = \begin{cases} \frac{1}{b-a} & x \in (a; b) \\ 0 & x \notin (a; b) \end{cases}$	$\begin{aligned} EX &= (a+b)/2 \\ D^2X &= (b-a)^2/12 \\ a &= 0 \\ k &= 1,8 \end{aligned}$
Rozkład normalny $m \in R, \sigma \in (0, +\infty)$	$f(x) = \frac{1}{\sigma\sqrt{2\pi}} e^{-\frac{(x-m)^2}{2\sigma^2}} \quad x \in R$	$\begin{aligned} EX &= m; \\ D^2X &= \sigma^2 \\ a &= 0 \\ k &= 3 \\ m_k &= m \cdot m_{k-1} + (k-1)\sigma^2 m_{k-2} \\ \mu_k &= \begin{cases} 0 & \text{gdy } k \text{ – nieparzyste} \\ \sigma^k (k-1)!! & \text{gdy } k \text{ – parzyste} \end{cases} \end{aligned}$
Rozkład wykładniczy $a \in (0, +\infty)$	$f(x) = \begin{cases} ae^{-ax} & x > 0 \\ 0 & x \leq 0 \end{cases}$	$\begin{aligned} EX &= 1/a; \\ D^2X &= 1/a^2 \\ a &= 2 \\ k &= 9 \\ m_k &= \frac{k!}{a^k} \\ \mu_k &= \frac{k!}{a^k} \sum_{j=0}^k \frac{(-1)^j}{j!} \end{aligned}$

Wskazówka do rozkładu wykładniczego:

$$\int_0^{\infty} x^n e^{-ax} dx = \frac{n!}{a^{n+1}}$$

MOMENTY EMPIRYCZNE:

$x_1, x_2, \dots, x_n$ – dane statystyczne (wartości cechy X),

$y_1, y_2, \dots, y_n$ – dane statystyczne (wartości cechy Y),

Momenty zwykłe,

$M_k = \frac{1}{n} \sum x_i^k$ – moment rzędu k cechy X ($M_1 = \bar{X}_n$).

$M_{kl} = \frac{1}{n} \sum x_i^k \cdot y_i^l$ – moment rzędu k, l jednocześnie badanych cech (X, Y).

Momenty centralne,

$\tilde{M}_k = \frac{1}{n} \sum (x_i - \bar{x})^k$ – moment rzędu k cechy X .

$\tilde{M}_{kl} = \frac{1}{n} \sum (x_i - \bar{x})^k (y_i - \bar{y})^l$ – moment rzędu k, l jednocześnie badanych cech (X, Y).

Własności

a) $\tilde{M}_2 = M_2 - M_1^2,$

b) $\tilde{M}_3 = M_3 - 3M_1M_2 + 2M_1^3,$

c) $\tilde{M}_4 = M_4 - 4M_1M_3 + 6M_1^2M_2 - 3M_1^4,$

d) $\tilde{M}_1 = 0,$

e) $\tilde{M}_2 = \frac{1}{n} \sum (x_i - \bar{x})^2 = S^2,$

f) $\tilde{M}_{11} = \frac{1}{n} \sum (x_i - \bar{x})(y_i - \bar{y}) = \frac{1}{n} \sum x_i y_i - \bar{x}\bar{y} = \text{cov}(X, Y),$

$$r_{XY} = \frac{\tilde{M}_{11}}{\sqrt{\tilde{M}_{20}} \cdot \sqrt{\tilde{M}_{02}}} = \frac{M_{11} - M_{10} \cdot M_{01}}{\sqrt{M_{20} - M_{10}^2} \cdot \sqrt{M_{02} - M_{01}^2}} =$$

g)

$$= \frac{\frac{1}{n} \sum x_i \cdot y_i - \frac{1}{n} \sum x_i \cdot \frac{1}{n} \sum y_i}{\sqrt{\frac{1}{n} \sum x_i^2 - \left(\frac{1}{n} \sum x_i\right)^2} \cdot \sqrt{\frac{1}{n} \sum y_i^2 - \left(\frac{1}{n} \sum y_i\right)^2}} = \frac{\frac{1}{n} \sum x_i \cdot y_i - \bar{X} \bar{Y}}{S_X \cdot S_Y}$$

Zadanie 1

Udowodnij własności a), b), c).

Zadanie 2

Jak zmieni się wartość momentów centralnych gdy

- a) pomnożymy wszystkie dane statystyczne przez dodatnią stałą c ,
- b) dodamy do wszystkich danych statystycznych stałą a .

Zadanie 3

Jak zmieni się wartość charakterystyk

Średnia, wariancja, odchylenie standardowe, mediana, dominanta, współczynnik asymetrii, kurioza gdy

- a) pomnożymy wszystkie dane statystyczne przez dodatnią stałą c ,
- b) dodamy do wszystkich danych statystycznych stałą a .

Metoda momentów

Wyznaczanie estymatorów metodą momentów

Nieznane momenty teoretyczne cechy X szacujemy przez momenty empiryczne tego samego rzędu.

Zatem przyjmujemy, że:

$$m_k \cong M_k \quad \text{oraz} \quad m_{kl} \cong M_{kl}$$

$$\mu_k \cong \tilde{M}_k \quad \text{oraz} \quad \mu_{kl} \cong \tilde{M}_{kl}$$

Parametry będące funkcjami momentów teoretycznych szacuje się przez wartości tych funkcji obliczone dla momentów empirycznych.

Przykład

Dla rozkładu wykładniczego z parametrem a mamy wartość oczekiwaną równą

$$EX = m_1 = 1/a.$$

Ponieważ przyjmujemy $m_1 \cong M_1$ to $1/a \cong \bar{X}_n$,

zatem estymatorem parametru a jest $\frac{1}{\bar{X}_n}$.

Uwaga

Jeśli cecha X ma momenty odpowiednio wysokiego rzędu to momenty te mają rozkłady asymptotyczne normalne.

Moment M_k ma asymptotyczny rozkład

$$N\left(m_k, \sqrt{\frac{m_{2k} - m_k^2}{n}}\right)$$

Moment $\tilde{M}_k$ ma asymptotyczny rozkład

$$N\left(\mu_k, \sqrt{\frac{\mu_{2k} - 2k\mu_{k-1}\mu_{k+1} - \mu_k^2 + k^2\mu_2\mu_{k-1}^2}{n}}\right)$$

Estymatory nieobciążone momentów centralnych

$$\tilde{M}_2^n = \frac{n}{n-1} \tilde{M}_2$$

$$\tilde{M}_3^n = \frac{n^2}{(n-1)(n-2)} \tilde{M}_3$$

$$\tilde{M}_4^n = \frac{n(n^2 - 2n + 3)}{(n-1)(n-2)(n-3)} \tilde{M}_4 - \frac{3n(2n-3)}{(n-1)(n-2)(n-3)} \tilde{M}_2$$

Estymator nieobciążony współczynnika asymetrii

$$\hat{a} = \frac{n^2}{(n-1)(n-2)} \frac{\frac{1}{n} \sum_{i=1}^n (x_i - \bar{x})^3}{\hat{s}^3}$$

gdzie

$$\hat{s}^2 = \frac{1}{n-1} \sum_{i=1}^n (x_i - \bar{x})^2$$

jest wariancją nieobciążoną (z próby)

Estymator nieobciążony kurtozy

$$\hat{k} = \frac{n^2(n+1)}{(n-1)(n-2)(n-3)} \frac{\frac{1}{n} \sum_{i=1}^n (x_i - \bar{x})^4}{\hat{s}^4} - 3 \frac{(n-1)^2}{(n-2)(n-3)} + 3$$

↓

gdzie

$$\hat{s}^2 = \frac{1}{n-1} \sum_{i=1}^n (x_i - \bar{x})^2$$

jest **wariancją nieobciążoną (z próby)**

Test współczynnika asymetrii.

Niech
$$M_k = \frac{1}{n} \sum_{i=1}^n (x_i - \bar{x})^k$$

empiryczny (z próby) moment centralny rzędu k .

Stawiamy hipotezy:

$$H_0(a = 0), \quad H_1(a \neq 0)$$

Test skorygowany

$$U_n = n \frac{\frac{M_3^2}{M_2^3}}{9 + M_6 M_2^{-3} - 6 M_4 M_2^{-2}}$$

Test uproszczony (rozkład zbliżony do normalnego):

$$U_n = n \frac{\frac{M_3^2}{M_2^3}}{6}$$

Zbiór krytyczny $K = \langle k, \infty \rangle$.

W obu przypadkach wartości krytyczne k odczytujemy z tablicy rozkładu chi kwadrat z jednym stopniem swobody.

Test kurtozy.

Niech
$$M_k = \frac{1}{n} \sum_{i=1}^n (x_i - \bar{x})^k$$

empiryczny (z próby) moment centralny rzędu k .

Stawiamy hipotezy:

$$H_0(k = 3), \quad H_1(k \neq 3)$$

Stosujemy statystykę:

$$U_n = n \frac{\left(\frac{M_4}{M_2^2} - 3 \right)^2}{24}$$

Zbiór krytyczny $K = (k, \infty)$.

Wartość krytyczną k odczytujemy z tablicy rozkładu chi kwadrat z jednym stopniem swobody.

Test Fishera (dodatkowy test normalności rozkładu)

Niech $P_k = \sum x_i^k$.

Określamy kumulanty

$$K_1 = \frac{1}{n} P_1$$

$$K_2 = \frac{n P_2 - P_1^2}{n(n-1)}$$

$$K_3 = \frac{n^2 P_3 - 3n P_2 P_1 + 2 P_1^3}{n(n-1)(n-2)}$$

$$K_4 = \frac{(n^3 + n^2) P_4 - 4(n^2 + n) P_3 P_1 - 3(n^2 - n) P_2^2 + 12n P_2 P_1^2 - 6 P_1^4}{n(n-1)(n-2)(n-3)}$$

Test współczynnika asymetrii.

Stawiamy hipotezy:

$$H_0(a = 0), \quad H_1(a \neq 0)$$

Obliczamy wartość statystyki:

$$U_a = \sqrt{\frac{n}{6}} \frac{K_3}{K_2^{3/2}}$$

Statystyka ma asymptotyczny rozkład $N(0;1)$

$$K = (-\infty; -k) \cup (k; \infty)$$

gdzie k odczytujemy z tablicy $N(0, 1)$ dla poziomu istotności $1 - \alpha/2$.

Decyzje:

Jeśli $U_n \in K$ to H_0 odrzucamy ,

Jeśli $U_n \notin K$ to nie ma podstaw do odrzucenia H_0 .

Test kurtozy.

Stawiamy hipotezy:

$$H_0(k = 3), \quad H_1(k \neq 3)$$

Obliczamy wartość statystyki:

$$U_k = \left(\frac{n}{24} \right)^{1/2} \frac{K_4}{K_2^2}$$

Statystyka ma asymptotyczny rozkład $N(0;1)$

$$K = (-\infty; -k] \cup [k; \infty)$$

gdzie k odczytujemy z tablicy $N(0, 1)$ dla poziomu istotności $1 - \alpha/2$.

Decyzje:

Jeśli $U_n \in K$ to H_0 odrzucamy ,

Jeśli $U_n \notin K$ to nie ma podstaw do odrzucenia H_0 .

Test normalności.

Stawiamy hipotezę:

H_0 (X - ma rozkład normalny)

Obliczamy wartość statystyki:

$$U = (U_a)^2 + (U_k)^2$$

Statystyka ma rozkład Y_2

Zbiór krytyczny $K = [k, \infty)$.

Wartość krytyczną k odczytujemy z tablicy rozkładu chi kwadrat z dwoma stopniami swobody.

Przykład.

Wiedząc, że:

n	P1	P2	P3	P4
100	-4,04849	116,8441	-2,91853	338,0933

Oblicz

K1	K2	K3	K4

oraz

Ua	Uk	U

dla $\alpha = 0,05$ sprawdź powyższe hipotezy.

Testy oparte o momenty z próby cd.

Test D'Agostino oparty o współczynnik asymetrii.

Test D'Agostino (1970) za statystykę testową wykorzystuje przekształcony współczynnik asymetrii z próby. Niech

$$a = \frac{\frac{1}{n} \sum_{i=1}^n (x_i - \bar{x})^3}{s^3}$$

gdzie

$$s^2 = \frac{1}{n} \sum_{i=1}^n (x_i - \bar{x})^2$$

W celu „normalizacji” rozkładu statystyki testowej stosujemy przekształcenie

Niech

$$A = a \sqrt{\frac{(n+1)(n+3)}{6(n-2)}}$$

$$B = \frac{3(n^2 + 27n - 70)(n+1)(n+3)}{(n-2)(n+5)(n+7)(n+9)}$$

$$W^2 = \sqrt{2(B-1)} - 1$$

$$E = \frac{1}{\sqrt{\ln W}}$$

$$F = \frac{A}{\sqrt{\frac{2}{W^2 - 1}}}$$

$$U_a = E \ln \left(F + \sqrt{F^2 + 1} \right)$$

Stawiamy hipotezy:

Stawiamy hipotezy:

$$H_0(a = 0), \quad H_1(a \neq 0)$$

(można też stosować jednostronne hipotezy alternatywne)

Statystyka ma rozkład $N(0;1)$

$$K = (-\infty; -k) \cup (k; \infty)$$

gdzie k odczytujemy z tablicy $N(0, 1)$ dla poziomu istotności $1 - \alpha/2$.

Decyzje:

Jeśli $U_n \in K$ to H_0 odrzucamy ,

Jeśli $U_n \notin K$ to nie ma podstaw do odrzucenia H_0 .

Test D'Agostino oparty o kurtozę

Test za statystykę testową wykorzystuje przekształconą kurtozę z próby. Niech

$$k = \frac{\frac{1}{n} \sum_{i=1}^n (x_i - \bar{x})^4}{s^4}$$

gdzie $s^2 = \frac{1}{n} \sum_{i=1}^n (x_i - \bar{x})^2$

W celu „normalizacji” rozkładu statystyki testowej stosujemy przekształcenie
Niech

$$A = \frac{k - \frac{3(n-1)}{n+1}}{\sqrt{\frac{24n(n-2)(n-3)}{(n+1)^2(n+3)(n+5)}}}$$

$$B = \frac{6(n^2 - 5n + 2)}{(n+7)(n+9)} \sqrt{\frac{6(n+3)(n+5)}{n(n-2)(n-3)}}$$

$$C = 6 + \frac{8}{B} \left(\sqrt{1 + \frac{4}{B^2}} + \frac{2}{B} \right)$$

$$U_k = \frac{\left(1 - \frac{2}{9C} - \sqrt[3]{\frac{1 - \frac{2}{C}}{1 + A\sqrt{\frac{2}{C-4}}}} \right)}{\sqrt{\frac{2}{9C}}}$$

Stawiamy hipotezy:

$$H_0(k = 3), \quad H_1(k \neq 3)$$

(można też stosować jednostronne hipotezy alternatywne)

Statystyka ma asymptotyczny rozkład $N(0;1)$

$$K = (-\infty; -k] \cup [k; \infty)$$

gdzie k odczytujemy z tablicy $N(0, 1)$ dla poziomu istotności $1 - \alpha/2$.

Decyzje:

Jeśli $U_n \in K$ to H_0 odrzucamy ,

Jeśli $U_n \notin K$ to nie ma podstaw do odrzucenia H_0 .

Test typu omnibus D'Agostino oparty o kurtozę i współczynnik asymetrii.

Łącząc dwa powyższe testy otrzymuje się test wrażliwy na odstępstwa od normalności zarówno w postaci niezerowej asymetrii jak i kurtozy istotnie różnej od 3 (D'Agostino 1973).

Stawiamy hipotezę:

H_0 (X - ma rozkład normalny)

Obliczamy wartość statystyki:

$$U = (U_a)^2 + (U_k)^2$$

Statystyka ma rozkład Y_2

Zbiór krytyczny $K = \leq k, \infty$.

Wartość krytyczną k odczytujemy z tablicy rozkładu chi kwadrat z dwoma stopniami swobody.

L.Kowalski 4.04.2016