

Rodzina przekształceń Boxa-Coxa (dla $x > 0$)

$$T_p(x) = \begin{cases} \frac{x^p - 1}{p} & p \neq 0 \\ \ln x & p = 0 \end{cases}$$

$$T'_p(x) = x^{p-1}$$

Zadanie. Znaleźć p takie, że wektor $T_p(x)$ będzie symetryczny.

Korzystając z współczynnika symetrii γ zadanie to można sformułować w przybliżony sposób

Zadanie. Znaleźć p takie, że $\gamma(\alpha_k, T_p(x)) \simeq 0$ dla wszystkich $\alpha_k < 0.5$

Niech $x_{(n+1-k)} > x_{(k)}$

Dla n nieparzystych

$$\gamma(\alpha_k, T_p(x)) = \begin{cases} \frac{x_{(n+1-k)}^p + x_{(k)}^p - 2 * x_{(r)}^p}{x_{(n+1-k)}^p - x_{(k)}^p} & p \neq 0 \\ \frac{\ln(x_{(n+1-k)}) + \ln(x_{(k)}) - 2 * \ln(x_{(r)})}{\ln(x_{(n+1-k)}) - \ln(x_{(k)})} & p = 0 \end{cases}, r = \frac{n+1}{2}$$

Dla n parzystych, takich, że $x_{(\frac{n}{2})} < x_{(\frac{n}{2}+1)}$ wektor uzupełniamy o dodatkowy wyraz, równy $\tilde{x} = \frac{x_{(\frac{n}{2})} + x_{(\frac{n}{2}+1)}}{2}$.

Niech

$$u_- \stackrel{df}{=} \frac{\tilde{x}}{x_{(k)}} > 1, u_+ \stackrel{df}{=} \frac{x_{(n+1-k)}}{\tilde{x}} > 1,$$

Wtedy

$$f(p) \stackrel{df}{=} \gamma(\alpha_k, T_p(x)) = \begin{cases} \frac{u_+^p + u_-^p - 2}{u_+^p - u_-^p} & p \neq 0 \\ \frac{\ln u_+ - \ln u_-}{\ln u_+ + \ln u_-} & p = 0 \end{cases}$$

oraz $f(p)$ jest ciągłą funkcją p .

Zadanie sprowadza się do rozwiązania równania $f(p) = 0$.

Zadanie to ma rozwiązanie, gdyż

$$\lim_{p \rightarrow \infty} f(p) = \lim_{p \rightarrow \infty} \frac{u_+^p - 2}{u_+^p} = 1,$$

$$\lim_{p \rightarrow -\infty} f(p) = \lim_{p \rightarrow -\infty} \frac{u_-^p - 2}{-u_-^p} = -1$$

Prosta, przechodząca przez punkty $(-1, f(-1))$ i $(1, f(1))$ ma równanie

$$\frac{f(1) - f(-1)}{2}(p - 1) + f(1)$$

Z metody siecznych przybliżone rozwiązanie równania

$$f(p) = 0$$

jest rozwiązaniem równania

$$\frac{f(1)-f(-1)}{2}(p-1)+f(1)=0$$

skąd

$$\begin{aligned} p &\simeq \frac{f(-1)+f(1)}{f(-1)-f(1)} = \\ &= \frac{\frac{u_+^{-1}+u_-^{-1}-2}{u_+^{-1}-u_-^{-1}} + \frac{u_++u_-^{-1}-2}{u_++u_-^{-1}}}{\frac{u_+^{-1}+u_-^{-1}-2}{u_+^{-1}-u_-^{-1}} - \frac{u_++u_-^{-1}-2}{u_++u_-^{-1}}} = \frac{\frac{1+u_+u_-^{-1}-2u_+}{1-u_+u_-} + \frac{u_++u_-^{-1}-2u_-}{u_++u_-^{-1}}}{\frac{1+u_+u_-^{-1}-2u_+}{1-u_+u_-} - \frac{u_++u_-^{-1}-2u_-}{u_++u_-^{-1}}} \\ &= \frac{(1+u_+u_-^{-1}-2u_+)-(u_++u_-^{-1}-2u_-)}{(1+u_+u_-^{-1}-2u_+)+(u_++u_-^{-1}-2u_-)} \\ &= \frac{2(u_-^{-1}-u_+)}{2(1+u_+u_-^{-1}-u_+-u_-)} \\ &= \frac{1}{u_+-1} - \frac{1}{u_-^{-1}-1} \end{aligned}$$

Przykład 1.

1	2	3	4	5	6	7
1	4	9	16	25	36	49

	1	2	3
u_-	$\frac{16}{1} = 16$	$\frac{16}{4} = 4$	$\frac{16}{9} = 1.78$
u_+	$\frac{49}{16} = 3.06$	$\frac{36}{16} = 2.25$	$\frac{25}{16} = 1.56$
p	$\frac{1}{3.06-1} - \frac{1}{16-1} = 0.42$	$\frac{1}{2.25-1} - \frac{1}{4-1} = 0.47$	$\frac{1}{1.56-1} - \frac{1}{1.78-1} = 0.50$

Przykład 2. Planety

1	2	3	4	5	6	7	8	9
0.4	0.7	1.0	1.5	5.2	9.5	19.2	30.0	39.4

	1	2	3	4	5
u_-	13.0	7.43	5.20	3.47	
u_+	7.58	5.77	3.69	1.83	
p	0.07	0.05	0.13	0.80	

Mediana wartości p , równa 0.10 jest najlepszym oszacowaniem p . Wybierając najbliższą wartość całkowitą otrzymać można skalę, symetryzującą dane o planetach. Jest to skala logarytmiczna. Obliczona w niej średnia Czebyszewa wynosi $m = 1.38, \bar{x} = 1.45, \tilde{x} = 1.65$ ■

Jaka wartość typowa?

$$\begin{array}{ccc} x & \xrightarrow{T_p} & y \\ \downarrow & & \downarrow \\ \tilde{x} & \xleftarrow{T_p^{-1}} & \tilde{y} \end{array}$$