Marcel Josífko

Charakteristická funkce Kendallova koeficientu korelace pořadí

Časopis pro pěstování matematiky, Vol. 83 (1958), No. 1, 56--59

Persistent URL: http://dml.cz/dmlcz/108184

Časopis pro pěstování matematiky, roč. 83 (1958), Praha

CHARAKTERISTICKÁ FUNKCE KENDALLOVA KOEFICIENTU KORELACE POŘADÍ

MARCEL JOSÍFKO, Praha

DT: 519.272.1

(Došlo dne 21. prosince 1956)

V poznámce je odvozena charakteristická funkce náhodné veličiny S (čitatele ve výrazu (1)) pro Kendallův koeficient korelace pořadí v případě nezávislosti. Charakteristické funkce je použito k jednoduchému odvození známých vlastností Kendallova koeficientu korelace pořadí.

Buď $\{(x_\nu, y_\nu)\}_{\nu=1}^n$ náhodný výběr o rozsahu n z dvojrozměrné populace charakterisované náhodnou veličinou (X, Y) se spojitým rozložením. Přiřaďme každému prvku tohoto náhodného výběru dvojici přirozených čísel (i, q_i) tak, že i značí pořadí prvku ve výběru uspořádaném podle rostoucích hodnot x a q_i značí pořadí tohoto prvku ve výběru uspořádaném podle rostoucích hodnot y. Symbolem R pak označme minimální počet inversí, jimiž se permutace $(q_1, q_2, \ldots, q_n)$ převádí na přirozenou posloupnost $(1, 2, \ldots, n)$. Kendallův koeficient korelace pořadí se pro tento náhodný výběr definuje vzorcem

$$\tau = 1 - \frac{2R}{\binom{n}{2}} = \frac{S}{\binom{n}{2}}. \tag{1}$$

Za předpokladu, že náhodné veličiny X a Y jsou nezávislé, jsou všechny permutace $(q_1, q_2, \ldots, q_n)$ stejně pravděpodobné (pořadí podle y nezávisí na uspořádání výběru podle rostoucích hodnot x). Označíme-li symbolem $f(s; n)$ počet těch permutací $(q_1, q_2, \ldots, q_n)$, které vedou k téže hodnotě s náhodné veličiny S ze vzorce (1), můžeme psát:

$$P_n(S = s) = \frac{f(s; n)}{n!}. \tag{2}$$

Vidíme snadno, že je $f(s; n) > 0$ pouze pro hodnoty $s = \binom{n}{2} - 2k$, kde $k = 0, 1, \ldots, \binom{n}{2}$; pro ostatní s je $f(s; n) = 0$.

Jednoduchou úvahou se odvodí rekurentní vzorec pro výpočet hodnot $f(s; n)$ (Viz na př. [1] str. 403):

$$f(s; n+1) = \sum_{k=0}^{n} f(s+n-2k; n)\,,$$
$$f(0; 1) = 1\,. \tag{3}$$

Pomocí vzorce (3) dostaneme pro charakteristickou funkci náhodné veličiny S tyto vztahy

$$\varphi_n(t) = \frac{1}{n!} \cdot \sum_s e^{its} \,.\, f(s; n)\,,$$

$$\varphi_{n+1}(t) = \frac{1}{(n+1)!} \sum_s e^{its} \sum_{k=0}^{n} f(s+n-2k; n)\,,$$

$$\varphi_{n+1}(t) = \frac{1}{n+1} \cdot \varphi_n(t) \cdot \sum_{k=0}^{n} e^{it(2k-n)}\,.$$

Sečtením geometrické řady v posledním výrazu a elementární úpravou dospíváme k jednoduché rekurentní formuli pro charakteristickou funkci $\varphi_n(t)$

$$\varphi_{n+1}(t) = \varphi_n(t) \cdot \frac{1}{n+1} \cdot \frac{\sin\,[(n+1)\,t]}{\sin t}\,. \tag{4}$$

Odtud

$$\varphi_n(t) = \frac{1}{n!} \prod_{\nu=1}^{n} \frac{\sin\,(\nu t)}{\sin t}\,. \tag{5}$$

Ke stejnému výsledku dospějeme úpravou vytvořujícího polynomu pro γ-rozdělení (viz [2]).

Uvedená charakteristická funkce umožňuje snadné odvození momentů náhodné veličiny S a důkaz konvergence jejího rozložení k rozložení normálnímu. K tomu účelu vztah (5) upravíme použitím známého vyjádření sinu ve tvaru nekonečného součinu

$$\sin z = z \,.\, \prod_{j=1}^{\infty} \left(1 - \frac{z^2}{\pi^2 j^2}\right). \tag{6}$$

V našem případě je

$$\varphi_n(t) = \prod_{\nu=1}^{n} \prod_{j=1}^{\infty} \left(1 - \frac{\nu^2 t^2}{\pi^2 j^2}\right) \left(1 - \frac{t^2}{\pi^2 j^2}\right)^{-1}.$$

V okolí nuly můžeme psát

$$\log \varphi_n(t) = - \sum_{\nu=1}^{n} \sum_{j=1}^{\infty} \sum_{k=1}^{\infty} \frac{1}{k} \left(\frac{t^2}{j^2 \pi^2}\right)^k \cdot (\nu^{2k} - 1)\,.$$

Je hned vidět, že pro $|t| \leqq n^{-1}$ platí

$$\log \varphi_n(t) = - \sum_{k=1}^{\infty} \frac{t^{2k}}{k} \cdot \sum_{\nu=1}^{n} (\nu^{2k} - 1) \cdot \frac{1}{\pi^{2k}} \sum_{j=1}^{\infty} \frac{1}{j^{2k}}\,.$$

Uvážíme-li, že $\sum_{j=1}^{\infty} \frac{1}{j^2} = \frac{\pi^2}{6}$, dostáváme pro charakteristickou funkci $\varphi_n(t)$ tento výraz

$$\varphi_n(t) = \exp\left\{-\frac{t^2}{2} \cdot \frac{1}{3} \sum_{\nu=1}^{n} (\nu^2 - 1) - \sum_{k=2}^{\infty} t^{2k} \frac{\sum_{\nu=1}^{n} (\nu^{2k} - 1)}{k \cdot \pi^{2k}} \cdot \sum_{j=1}^{\infty} \frac{1}{j^{2k}}\right\},$$
$$|t| \leqq n^{-1}. \tag{7}$$

Odtud přímo určíme rozptyl náhodné veličiny S

$$\sigma_S^2 = \frac{1}{3} \sum_{\nu=1}^{n} (\nu^2 - 1) = \frac{n(n-1)(2n+5)}{18}. \tag{8}$$

Standardisovaná náhodná veličina

$$U = \frac{S}{\sigma_S} \tag{9}$$

má, jak je ihned patrné z (7) a (8), tyto kumulanty:

$$\varkappa_2 = 1, \quad \varkappa_{2k} = O(n^{-k+1}) \quad \text{pro } k = 2, 3, \ldots$$

Liché kumulanty jsou vesměs rovny nule.

Náhodná veličina U má tedy charakteristickou funkci

$$\psi_n(t) = \exp\left\{-\frac{t^2}{2} + O(n^{-1})\right\}.$$

Platí tedy $\lim \psi_n(t) = \psi(t) = e^{-\frac{t^2}{2}}$ pro každé pevné t, takže náhodná veličina U má asymptoticky normální rozložení $N(0, 1)$.

Kendallův koeficient korelace pořadí τ má tedy rovněž, za předpokladu nezávislosti náhodných veličin X a Y, asymptoticky normální rozložení s průměrem 0 a rozptylem

$$\sigma_\tau^2 = \frac{2n+5}{9 \cdot \binom{n}{2}};$$

$2k$-tý kumulant náhodné veličiny τ je

$$\varkappa_{2k}(\tau) = \frac{(2k)!}{k \cdot \pi^{2k}} \sum_{j=1}^{\infty} \frac{1}{j^{2k}} \cdot \binom{n}{2}^{-2k} \cdot \sum_{\nu=1}^{n} (\nu^{2k} - 1).$$

LITERATURA

[1] *M. G. Kendall:* The advanced theory of statistics I. (1948).

[2] *J. Hájek:* Některá pořadová rozdělení a jejich použití. Čas. pro pěst. mat. *80* (1955), 17—31.

Резюме

СОБСТВЕННАЯ ФУНКЦИЯ КОЭФФИЦИЕНТА КЕНДАЛЛА КОРРЕЛЯЦИИ РАНГОВ

МАРЦЕЛЬ ИОСИФКО (Marcel Josífko), Прага

(Поступило в редакцию 21/XII 1956 г.)

В статье выводится собственная функция случайной величины S (числителя в выражении (1)) для коэффициента Кендалла корреляции рангов в случае независимости. Собственная функция используется для простого вывода известных уже свойств коэффициента Кендалла корреляции рангов.

Summary

THE CHARACTERISTIC FUNCTION OF THE KENDALL'S RANK CORRELATION COEFFICIENT

MARCEL JOSÍFKO, Praha

(Received December 21, 1956)

The characteristic function for the random variable S (the numerator of Kendall's coefficient of rank correlation (1)) in the case of independence is given. The characteristic function is applied to the simple derivation of cumulants and asymptotic distribution of Kendall's τ.