ZASTOSOWANIA MATEMATYKI
APPLICATIONES MATHEMATICAE
IX, 2 (1967)

Z. CYLKOWSKI (Wrocław) i W. KLONECKI (Poznań)

ROZKŁADY ZNAKÓW WSKAŹNIKÓW PERKALA

1. Wstęp. W niniejszej pracy podajemy rozkłady znaków wskaźników Perkala [2] dla zespołów liczących od dwóch do osiemnastu cech. Rozkłady te obliczono na podstawie wzorów podanych w pracy jednego z autorów [1]. Wartości parametrów występujących w tych wzorach obliczono na elektronicznej maszynie cyfrowej Elliott-803 w Katedrze Metod Numerycznych Uniwersytetu Wrocławskiego stosując rekurencyjny wzór Rubena [4]. Podajemy także przykład zastosowania podanych tablic przy weryfikacji hipotezy dotyczącej rozkładu zespołów cech. Używamy tutaj oznaczeń, jakie zostały wprowadzone w pracy [1].

Poczytujemy sobie za miły obowiązek złożyć na tym miejscu wyrazy podziękowania Profesorowi J. Łukaszewiczowi, który tę pracę przejrzał i którego cenne uwagi wykorzystaliśmy.

2. Wskaźniki Perkala i rozkład ich znaków. Jeśli zmienne losowe $X_1, X_2, \ldots, X_n$ oznaczają badane cechy, $EX_1, EX_2, \ldots, EX_n$ ich oczekiwane wartości, a $D^2X_1, D^2X_2, \ldots, D^2X_n$ wariancje, to składowe wektora

$$(W_1, W_2, \ldots, W_n) = \left(\frac{X_1-EX_1}{DX_1}-m, \frac{X_2-EX_2}{DX_2}-m, \ldots, \frac{X_n-EX_n}{DX_n}-m\right),$$

gdzie

$$m = \frac{1}{n}\sum_{i=1}^{n}\frac{X_i-EX_i}{DX_i},$$

noszą nazwę *wskaźników Perkala*. Przez *rozkład znaków wskaźników Perkala* należy rozumieć rozkład zmiennej losowej $(Z_1, Z_2, \ldots, Z_n)$, gdzie

$$Z_i = \begin{cases} 1, & \text{jeśli} \quad W_i \geqslant 0, \\ 0, & \text{jeśli} \quad W_i < 0 \end{cases} \qquad (i = 1, 2, \ldots, n),$$

a przez *rozkład ilości dodatnich wskaźników Perkala* rozkład sumy $S_n = Z_1+Z_2+\ldots+Z_n$.

Rozkład zmiennej losowej $(Z_1, Z_2, \ldots, Z_n)$ jest wyznaczony przez rozkład zmiennej $(X_1, X_2, \ldots, X_n)$. Ograniczymy się do obliczenia rozkładu znaków wskaźników Perkala dla pewnej szczególnej klasy rozkładów o tej własności, że dla wszystkich rozkładów tej klasy zmienna losowa $(Z_1, Z_2, \ldots, Z_n)$ ma ten sam rozkład co dla rozkładu jednostajnego w n-wymiarowej kuli jednostkowej ze środkiem w początku układu współrzędnych. Do tak określonej klasy rozkładów należą rozkłady normalne o zerowych wartościach oczekiwanych, jednostkowych odchyleniach standardowych i niezmiennicze względem obrotów dookoła osi o równaniu $x_1 = x_2 = \ldots = x_n$.

Obecnie podamy wartości oczekiwane i wariancje zmiennych losowych $Z_1, Z_2, \ldots, Z_n$ oraz zmiennej S_n pod warunkiem, że zmienna losowa $(X_1, X_2, \ldots, X_n)$ ma rozkład jednostajny w n-wymiarowej kuli jednostkowej ze środkiem w początku układu współrzędnych.

Zmienne losowe $Z_1, Z_2, \ldots, Z_n$ są zależne, gdyż $Z_1+Z_2+\ldots+Z_n = 0$. Każda ze zmiennych $Z_1, Z_2, \ldots, Z_n$ przyjmuje wartości 0 i 1, każdą z nich z prawdopodobieństwem 1/2. Mamy więc $EZ_i = 1/2$ oraz $D^2Z_i = 1/4$ dla $i = 1, 2, \ldots, n$. Kowariancja zmiennych Z_i i Z_j, gdzie $i \neq j$, jest równa

$$\operatorname{cov}(Z_i, Z_j) = \Pr\{Z_i = 1, Z_j = 1\} - 1/4.$$

Korzystając z tożsamości (patrz [1]) $Pr\{Z_i = 1, Z_j = 1\} = V_{0,2}^{(2)}(\theta)$, możemy napisać

$$\operatorname{cov}(Z_i, Z_j) = V_{0,2}^{(2)}(\theta) - 1/4, \qquad \text{gdzie} \qquad \theta = \arccos\left(-\frac{1}{n-1}\right),$$

lub

$$\operatorname{cov}(Z_i, Z_j) = \frac{1}{2\pi} \operatorname{arctg} \frac{1}{\sqrt{n(n-2)}},$$

co wynika z określenia funkcji $V_{0,2}^{(2)}(\theta)$.

Prawdopodobieństwo, że np. $(Z_1, Z_2, \ldots, Z_n) = (0, 1, 1, 0, \ldots, 0, 1)$ zależy tylko od ilości zer i jedynek w ciągu $(0, 1, 1, 0, \ldots, 0, 1)$, a nie zależy od tego, na jakich miejscach są zera, a na jakich jedynki. Zgodnie z oznaczeniami wprowadzonymi w [1] napiszemy więc

$$\Pr\{S_n = \beta\} = \binom{n}{\beta} V_{\beta,n}^{(n-1)}.$$

Wartość oczekiwana zmiennej losowej S_n jest równa $n/2$, a wariancja wyraża się następującym wzorem:

$$D^2S_n = \frac{n}{4} - 2\binom{n}{2} \frac{1}{2\pi} \operatorname{arctg} \frac{1}{\sqrt{n(n-2)}}.$$

Stąd wynika, że

$$\lim_{n\to\infty} \frac{D^2 S_n}{n} = 1/4 - 1/2\pi = 0{,}090845.$$

W trzeciej kolumnie tablicy 1 podano wartości prawdopodobieństwa $V_{\beta,n}^{(n-1)}$ dla rozkładu znaków wskaźników Perkala (tzn. dla rozkładu zmiennej losowej $(Z_1, Z_2, \ldots, Z_n)$) dla $n = 3, 4, \ldots, 18$ oraz dla $1 \leqslant \beta \leqslant n/2$. Wartości tych prawdopodobieństw dla $n/2 < \beta \leqslant n-1$ otrzymujemy ze wzoru $V_{\beta,n}^{(n-1)} = V_{n-\beta,n}^{(n-1)}$.

W czwartej kolumnie tablicy 1 podano wartości prawdopodobieństwa $\binom{n}{\beta} V_{\beta,n}^{(n-1)}$ dla $n = 3, 4, \ldots, 18$ i dla $1 \leqslant \beta \leqslant n/2$. Pozostałe wartości prawdopodobieństw dla $n/2 < \beta \leqslant n-1$ otrzymujemy ze wzoru $\Pr\{S_n = \beta\} = \Pr\{S_n = n-\beta\}$.

Korzystając z tablicy 1 można obliczyć wyrazy ciągu $\{a_n\} = \{D^2 S_n/n\}$ dla $n = 3, 4, \ldots, 18$. W szczególności mamy $a_{15} = 0{,}090709$, $a_{16} = 0{,}090726$ $a_{17} = 0{,}090741$ i $a_{18} = 0{,}09075$.

Nie potrafiliśmy udowodnić, że zmienna losowa S_n ma rozkład asymptotycznie normalny. Okazuje się jednak, że już dla $n = 15$ prawdopodobieństwa wartości zmiennej S_n różnią się nieznacznie od odpowiednich wartości uzyskanych z przybliżenia normalnego, jak to widać z tablicy 2.

3. Metoda numerycznych obliczeń. Formułę nadającą się do obliczeń na elektronicznej maszynie cyfrowej dla wspomianych na wstępie wartości parametrów $V_{0,\gamma}^{(\gamma)}$ wyprowadziliśmy ze wzoru Rubena [4].

Niech $P_\gamma(\varrho) = V_{0,\gamma}^{(\gamma)}(\arccos \varrho)$. Mamy

$$P_0(\varrho) \equiv 1, \qquad P_1(\varrho) \equiv 1/2$$

oraz

$$\frac{dP_\gamma(\varrho)}{d\varrho} = \frac{\gamma(\gamma-1)}{4\pi} \frac{1}{\sqrt{1-\varrho^2}} P_{\gamma-2}\left(\frac{\varrho}{1+2\varrho}\right),$$

dla $\gamma \geqslant 2$.

Funkcję $P_\gamma(\varrho)$ można rozwinąć w przedziale

$$\left(-\frac{1}{2\left[\frac{\gamma}{2}\right]+1}, \frac{1}{2\left[\frac{\gamma}{2}\right]+1}\right)$$

w szereg Maclaurina:

$$P_\gamma(\varrho) = \sum_{n=0}^{\infty} \frac{1}{n!} \left.\frac{d^n P_\gamma(\varrho)}{d\varrho^n}\right|_{\varrho=0} \varrho^n.$$

Pokażemy teraz, że pochodne $d^n P_\gamma(\varrho)/d\varrho^n$, $n = 0, 1, \ldots$, wyrażające się wzorem

$$\frac{d^n P_\gamma(\varrho)}{d\varrho^n}\bigg|_{\varrho=0} =$$

$$= \begin{cases} \left(\dfrac{1}{2}\right)^\gamma & \text{jeśli } n = 0, \\ \dfrac{\gamma(\gamma-1)}{4\pi} \displaystyle\sum_{k=0}^{n-1} \binom{n-1}{k} \frac{d^{n-k-1}\left(\dfrac{1}{\sqrt{1-\varrho^2}}\right)}{d\varrho^{n-k-1}}\Bigg|_{\varrho=0} \frac{d^k P_{\gamma-2}\left(\dfrac{\varrho}{1+2\varrho}\right)}{d\varrho^k}\Bigg|_{\varrho=0} & \text{jeśli } n > 0, \end{cases}$$

dadzą się wyrazić rekurencyjnie przez pochodne

$$\frac{d^k P_{\gamma-2}(\varrho)}{d\varrho^k}\bigg|_{\varrho=0}, \quad k = 1, 2, \ldots$$

W tym celu znajdziemy najpierw zależności między pochodnymi

$$\frac{d^k P_{\gamma-2}\left(\dfrac{\varrho}{1+2\varrho}\right)}{d\varrho^k}\Bigg|_{\varrho=0}$$

i pochodnymi

$$\frac{d^k P_{\gamma-2}(\varrho)}{d\varrho^k}\bigg|_{\varrho=0},$$

gdzie $k = 1, 2, \ldots$

Po n-krotnym zróżniczkowaniu względem ϱ tożsamości

$$(1+2\varrho)^2 \frac{d}{d\varrho}\left(\frac{d^k P_{\gamma-2}(u)}{du^k}\right) = \frac{d^{k+1} P_{\gamma-2}(u)}{du^{k+1}}, \quad k = 0, 1, \ldots,$$

gdzie $u = \varrho/(1+2\varrho)$, mamy

$$(1+2\varrho)^2 \frac{d^{n+1}}{d\varrho^{n+1}}\left(\frac{d^k P_{\gamma-2}(u)}{du^k}\right) + 4n(1+2\varrho)\frac{d^n}{d\varrho^n}\left(\frac{d^k P_{\gamma-2}(u)}{du^k}\right) +$$
$$+4n(n-1)\frac{d^{n-1}}{d\varrho^{n-1}}\left(\frac{d^k P_{\gamma-2}(u)}{du^k}\right) = \frac{d^n}{d\varrho^n}\left(\frac{d^{k+1} P_{\gamma-2}(u)}{du^{k+1}}\right).$$

Po podstawieniu $\varrho = 0$ i wprowadzeniu symbolu

$$p_{n,k}^{(\gamma-2)} = \frac{d^n}{d\varrho^n}\left(\frac{d^k P_{\gamma-2}(u)}{du^k}\right)_{\varrho=0}, \quad n, k = 0, 1, \ldots,$$

otrzymane wyrażenie daje napisać się w następującej postaci:

$$p_{n+1,k}^{(\gamma-2)} = p_{n,k+1}^{(\gamma-2)} - 4n\left[p_{n,k}^{(\gamma-2)} + (n-1)p_{n-1,k}^{(\gamma-2)}\right].$$

TABLICA 1

n	β	$V_{\beta,n}^{(n-1)}$	$\binom{n}{\beta} V_{\beta,n}^{(n-1)}$
3	1	0.16666 6667	0.50000 000
4	1	0.04386 9914	0.17547 966
	2	0.10817 3448	0.64904 069
5	1	0.00978 4688	0.04892 344
	2	0.04510 7656	0.45107 656
6	1	0.00192 2044	0.01153 226
	2	0.01487 7622	0.22316 432
	3	0.02653 0341	0.53060 683
7	1	0.00034 0695	0.00238 487
	2	0.00419 1353	0.08801 84
	3	0.01170 2764	0.40959 67
8	1	0.00005 5409	0.00044 327
	2	0.00104 9378	0.02938 258
	3	0.00427 9910	0.23967 494
	4	0.00658 5692	0.46099 842
9	1	0.00000 8369	0.00007 532
	2	0.00023 9163	0.00860 988
	3	0.00136 5246	0.11468 06
	4	0.00298 9160	0.37663 42
10	1	0.00000 11847	0.00001 1847
	2	0.00005 04182	0.00226 882
	3	0.00039 14575	0.04697 490
	4	0.00116 22390	0.24407 019
	5	0.00164 02717	0.41334 848
11	1	0.00000 01583	0.00000 1741
	2	0.00000 99435	0.00054 689
	3	0.00010 28976	0.01697 81
	4	0.00040 1085	0.13235 81
	5	0.00075 7825	0.35011 52
12	1	0.00000 00200 8	0.00000 02410
	2	0.00000 18501 6	0.00012 2110
	3	0.00002 51391 5	0.00553 0613
	4	0.00012 57097 3	0.06222 631
	5	0.00030 69564 8	0.24310 953
	6	0.00040 91151	0.37802 238
13	1	0.00000 00024 3	0.00000 00316
	2	0.00000 03269 0	0.00002 5499
	3	0.00000 57665	0.00164 921
	4	0.00003 63610	0.02599 81
	5	0.00011 18888	0.14400 09
	6	0.00019 13323	0.32832 63

TABLICA 1 (c. d.)

n	β	$V_{\beta,n}^{(n-1)}$	$\binom{n}{\beta} V_{\beta,n}^{(n-1)}$
14	1	0.00000 00002 81	0.00000 00039 4
	2	0.00000 00551 31	0.00000 50169
	3	0.00000 12514 9	0.00045 5541
	4	0.00000 98197 8	0.00982 960
	5	0.00003 73717 9	0.07481 833
	6	0.00007 98068 1	0.23965 98
	7	0.00010 21163 6	0.35046 33
15	1	0.00000 00000 32	0.00000 00005
	2	0.00000 00089 1	0.00000 0936
	3	0.00000 02585 4	0.00011 7634
	4	0.00000 24980 7	0.00340 987
	5	0.00001 15890 9	0.03480 20
	6	0.00003 02906 2	0.15160 46
	7	0.00004 81841 4	0.31006 50
16	1	0.00000 00000 03	0.00000 00000 5
	2	0.00000 00013 85	0.00000 01662
	3	0.00000 00510 85	0.00002 8608
	4	0.00000 06027 54	0.00109 701
	5	0.00000 33698 13	0.01471 935
	6	0.00001 06145 1	0.08500 100
	7	0.00002 05475 5	0.23506 39
	8	0.00002 54996 0	0.32817 99
17	1	0.00000 00000 003	0.00000 00000 0
	2	0.00000 00002 08	0.00000 00283
	3	0.00000 00096 93	0.00000 6592
	4	0.00000 01387 08	0.00033 0125
	5	0.00000 09258 90	0.00572 941
	6	0.00000 34718 58	0.04296 771
	7	0.00000 80455 59	0.15647 00
	8	0.00001 21141 96	0.29449 61
18	1	0.00000 00000 000	0.00000 00000 00
	2	0.00000 00000 301	0.00000 00046 1
	3	0.00000 00017 72	0.00000 14461
	4	0.00000 00305 80	0.00009 3575
	5	0.00000 02418 67	0.00207 232
	6	0.00000 10689 96	0.01984 484
	7	0.00000 29273 05	0.09315 856
	8	0.00000 52559 89	0.22999 16
	9	0.00000 63693 00	0.30967 54

Stąd widać, że możemy obliczyć $p_{k,0}^{(\gamma-2)}$, jeśli znamy już wartości $p_{0,k}^{(\gamma-2)}$ dla $k = 0, 1, \ldots$ Innymi słowy, mamy wzór rekurencyjny na obliczenie współczynników szeregu Maclaurina.

Przedstawiona metoda pozwala więc znaleźć przybliżone wartości funkcji $V_{0,\gamma}^{(\gamma)}$, $\gamma = 2, 4, \ldots$, występujących we wzorach na prawdopodobieństwo rozkładu znaków wskaźników Perkala.

4. Zastosowania. Rozkłady przedstawione w tablicy 1 można wykorzystać przy weryfikacji hipotezy orzekającej, że rozkład wartości cech jest niezmienniczy względem obrotów dookoła osi o równaniu $x_1 = x_2 = \ldots = x_n$. W szczególności, jeśli założymy dodatkowo, że rozważany układ cech ma rozkład normalny, wówczas warunek niezmienniczości rozkładu względem prostej $x_1 = x_2 = \ldots = x_n$ równoważny jest warunkowi, że macierz kowariancji $\{\sigma_{ij}\}$ ma elementy

$$\sigma_{ij} = \begin{cases} \sigma_1 & \text{dla} \quad i = j, \\ \sigma_2 & \text{dla} \quad i \neq j. \end{cases}$$

Nawiązując do wskaźnika sumarycznego m (patrz [2]), określającego sumaryczną wielkość indywiduum ze względu na zespół badanych cech, możemy przy pomocy znaków wskaźników Perkala weryfikować hipotezę, że wskaźnik m zawiera całą informację, tzn. że po wytrąceniu wskaźnika m pozostałe reszty są ze sobą nieskorelowane. Proponowany tutaj test polega na porównaniu empirycznego rozkładu znaków wskaźników Perkala z ich rozkładem teoretycznym. Zgodność rozkładów sprawdzamy na przykład przy pomocy testu χ^2.

W sposób analogiczny można weryfikować hipotezy dotyczące rozkładów podzespołów cech lub reszt.

Proponowaną metodę weryfikacji takich hipotez zilustrujemy jednym przykładem zaczerpniętym z pracy Perkala [3].

TABLICA 2

n	β	$\binom{n}{\beta} V_{\beta,n}^{(n-1)}$	Przybliżenie normalne	Błąd procentowy
	1	0.000000	0.000000	—
	2	0.000001	0.000009	+800
	3	0.000118	0.000294	+150
15	4	0.003410	0.004755	+ 40
	5	0.034802	0.038153	+ 10
	6	0.151605	0.152431	+ 1
	7	0.310065	0.304358	− 2

Tablica tej pracy zawiera wskaźniki Perkala obliczone dla siedmiu cech antropologicznych i pięćdziesięciu osobników. W tablicy 3 podajemy zaobserwowany rozkład ilości dodatnich wskaźników (opuściliśmy te przypadki, gdzie nie podano znaków wszystkich wskaźników). Prócz tego tablica 3 zawiera w czwartej kolumnie rozkład teoretyczny przepisany z tablicy 1 dla $n = 7$. Nie ma tutaj podstaw do odrzucenia hipotezy, że rozkład badanego zespołu cech antropologicznych jest niezmienniczy względem obrotów dookoła osi $x_1 = x_2 = \ldots = x_n$ lub, innymi słowy, że po wyeliminowaniu wskaźnika symetrycznego m pozostałe reszty są niezależne. Obliczone χ^2 wynosi 4,69 i jest mniejsze od wartości krytycznej 7,81 dla 3 stopni swobody przy poziomie istotności $\alpha = 0{,}05$.

TABLICA 3

Ilość +	Frekwencje empiryczne	Frakcje empiryczne	Frakcje teoretyczne
1	0	0.000	0.002
2	7	0.156	0.088
3	17	0.378	0.410
4	20	0.444	0.410
5	1	0.022	0.088
6	0	0.000	0.002

Prace cytowane

[1] W. Klonecki, *O rozkładzie znaków wskaźników Perkala*, Zastosow. Matem. 8 (1966), str. 295-307.

[2] J. Perkal, *On the analysis of a set of characteristics*, tamże 5 (1960), str. 35-45.

[3] — *O wskaźnikach antropologicznych*, Przegląd Antropol. 19 (1953), str. 209-221.

[4] H. Ruben, *On the moments of order statistics in samples from normal populations*, Biometrika 41 (1954), str. 200-227.

Praca wpłynęła 5. 4. 1966

З. ЦЫЛЬКОВСКИ (Вроцлав)
В. КЛЬОНЕЦКИ (Познань)

РАСПРЕДЕЛЕНИЯ ЗНАКОВ ИНДЕКСОВ ПЭРКАЛЯ

РЕЗЮМЕ

В этой работе даны распределения индексов Пэркаля для $n = 3, 4, \ldots, 18$. Они вычислены методом, указанным в работе [1].

Найденными индексами можно воспользоваться при проверке гипотезы утверждающей, что распределение признаков инвариантно относительно вращения вокруг оси, уравнение которой $x_1 = x_2 = \ldots = x_n$. Предлагаемый метод проверки иллюстрируется примером из работы Пэркаля [3].

Z. CYLKOWSKI (Wrocław)
W. KLONECKI (Poznań)

THE DISTRIBUTIONS OF THE SIGNS OF PERKAL'S INDICES

SUMMARY

In this paper the distributions of the signs of Perkal's natural indices (see [2]) are presented for $n = 3, 4, \ldots, 18$. They have been computed by the method described in [1].

These distributions can be applied to test the statistical hypothesis asserting that the distribution under test remains invariant under rotation around the line $x_1 = x_2 = \ldots = x_n$.

The proposed method of verification is illustrated by an example taken from Perkal's paper [3].