ZASTOSOWANIA MATEMATYKI
APPLICATIONES MATHEMATICAE
VIII, 4 (1966)

W. KLONECKI (Poznań)

O ROZKŁADZIE ZNAKÓW WSKAŹNIKÓW PERKALA

1. Jedną z metod, która jest stosowana z dużym powodzeniem przy analizie zespołu cech obok metody Hotellinga [2], Spearmana [6] i Thurstone'a [7], jest metoda wskaźników przyrodniczych Perkala [3]. Pozwala ona na uchwycenie tzw. przyrodniczego podobieństwa populacyjnego w odróżnieniu od podobieństwa geometrycznego. Pożądane jest obliczenie rozkładu tych wskaźników przyrodniczych, gdyż wtedy możliwe jest testowanie hipotez dotyczących przyrodniczego podobieństwa populacyjnego w ujęciu Perkala [3]. W niniejszej pracy podaję metodę, która pozwala wyznaczyć rozkład znaków wskaźników przyrodniczych Perkala dla pewnej symetrycznej klasy rozkładów. Otrzymane rozkłady znaków tych wskaźników mogą być przydatne przy testowaniu hipotez dotyczących postaci rozkładów badanych zespołów cech. Tablice rozkładów zostaną opublikowane w oddzielnej publikacji.

Rozkład znaków wskaźników przyrodniczych został otrzymany przez uogólnienie metody Davida [1], stosowanej przez niego przy wyprowadzeniu wzoru Shepparda dla trójwymiarowej zmiennej losowej o rozkładzie normalnym. Klasa rozkładów, którą badał David, spełniała warunek symetryczności analogiczny do warunku symetryczności podanego w założeniu A. Natomiast klasa rozkładów będąca przedmiotem naszych rozważań spełnia ponadto drugi warunek symetryczności podany w założeniu B. Dzięki temu dodatkowemu warunkowi symetryczności otrzymujemy pewne relacje, wzory (18)-(22), pozwalające określić postać rozkładu znaków wskaźników przyrodniczych Perkala z dokładnością do pewnych parametrów dających się wyrazić przez całki, rozważane m. in. przez Rubena [4], których wartości dają się obliczyć przy użyciu elektronowych maszyn matematycznych. Wzory (32) i (33) mogą być także otrzymane z wzoru rekurencyjnego podanego przez Rubena [4], wyprowadzonego przez niego ze znanego wzoru różniczkowego Schläfliego [5].

Na zagadnienie obliczenia rozkładu znaków przyrodniczych zwrócił mi uwagę J. Perkal, za co składam mu serdeczne podziękowanie, jak także

za cenne uwagi w czasie pisania niniejszej pracy. Dziękuję także Z. Cylkowskiemu za zwrócenie mi uwagi na to, że wzór (33) wynika z wzoru (32), co uprościło jego wyprowadzenie.

2. Wskaźniki przyrodnicze określone są w następujący sposób: niech zmienne losowe $X_1, X_2, \ldots, X_n$ oznaczają badane cechy, $EX_1, EX_2, \ldots, EX_n$ ich wartości średnie, a $D^2X_1, D^2X_2, \ldots, D^2X_n$ wariancje. Wskaźnikami przyrodniczymi nazywają się składowe wektora

$$(1)\qquad (W_1, W_2, \ldots, W_n) = \left(\frac{X_1-EX_1}{DX_1}-m, \frac{X_2-EX_2}{DX_2}-m, \ldots, \frac{X_n-EX_n}{DX_n}-m\right),$$

gdzie

$$(2)\qquad m = \frac{1}{n}\sum_{i=1}^{n}\frac{X_i-EX_i}{DX_i}.$$

Ze wzorów (1) oraz (2) wynika, że wskaźniki przyrodnicze $(W_1, W_2, \ldots, W_n)$ spełniają związek

$$(3)\qquad W_1+W_2+\ldots+W_n = 0.$$

Niniejsza praca jest poświęcona badaniu rozkładów zmiennej losowej

$$(4)\qquad (\operatorname{sign} W_1, \operatorname{sign} W_2, \ldots, \operatorname{sign} W_n),$$

gdzie

$$\operatorname{sign} W_i = \begin{cases} +1 & \text{dla } W_i \geqslant 0, \\ -1 & \text{dla } W_i < 0 \end{cases} \qquad (i = 1, 2, \ldots, n).$$

Z określenia zmiennej losowej (4) wynika, że może ona przyjmować z prawdopodobieństwem dodatnim co najwyżej $2(2^{n-1}-1)$ wartości.

Rozkład zmiennej losowej (4) jest wyznaczony przez rozkład zmiennej losowej $(X_1, X_2, \ldots, X_n)$. Tak np. prawdopodobieństwo układu nierówności

$$(5)\qquad W_1 \geqslant 0, \quad W_2 < 0, \quad \ldots, \quad W_n \geqslant 0$$

wyraża się przez całkę

$$(6)\qquad \int dF(w_1, w_2, \ldots, w_n),$$

gdzie F jest dystrybuantą zmiennej losowej $(W_1, W_2, \ldots, W_n)$, a obszar całkowania jest określony przez warunki $w_1 \geqslant 0, w_2 < 0, \ldots, w_n \geqslant 0$ odpowiadające układowi nierówności (5).

Wzory na prawdopodobieństwo poszczególnych układów znaków zmiennej losowej (4) przyjmują prostszą postać w układzie $(w_1', w_2', \ldots, w_n')$, w którym jedna z osi jest prostopadła do płaszczyzny $w_1+w_2+\ldots+w_n = 0$.

Taką transformacją jest np.

$$
\begin{aligned}
w_1' &= \frac{1}{\sqrt{n(n-1)}}[-(n-1)w_1+w_2+w_3+\ldots+w_{n-2}+w_{n-1}+w_n],\\
w_2' &= \frac{1}{\sqrt{(n-1)(n-2)}}[-(n-2)w_2+w_3+\ldots+w_{n-2}+w_{n-1}+w_n],\\
&\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\\
w_{n-2}' &= \frac{1}{\sqrt{3\cdot 2}}[-2w_{n-2}+w_{n-1}+w_n],\\
w_{n-1}' &= \frac{1}{\sqrt{2\cdot 1}}[-w_{n-1}+w_n],\\
w_n' &= \frac{1}{\sqrt{n}}[w_1+w_2+w_3+\ldots+w_{n-2}+w_{n-1}+w_n].
\end{aligned}
\tag{7}
$$

Dla transformacji (7) transformacja odwrotna ma postać

$$
\begin{aligned}
w_1 &= -\sqrt{\frac{n-1}{n}}w_1' + \sqrt{\frac{1}{n}}w_n',\\
w_2 &= \sqrt{\frac{1}{n(n-1)}}w_1' - \sqrt{\frac{n-2}{n-1}}w_2' + \sqrt{\frac{1}{n}}w_n',\\
w_3 &= \sqrt{\frac{1}{n(n-1)}}w_1' + \sqrt{\frac{1}{(n-1)(n-2)}}w_2' - \sqrt{\frac{n-3}{n-2}}w_3' + \sqrt{\frac{1}{n}}w_n',\\
&\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\\
w_{n-1} &= \sqrt{\frac{1}{n(n-1)}}w_1' + \sqrt{\frac{1}{(n-1)(n-2)}}w_2' +\\
&\quad + \sqrt{\frac{1}{(n-2)(n-3)}}w_3' + \ldots - \sqrt{\frac{1}{2}}w_{n-1}' + \sqrt{\frac{1}{n}}w_n',\\
w_n &= \sqrt{\frac{1}{n(n-1)}}w_1' + \sqrt{\frac{1}{(n-1)(n-2)}}w_2' +\\
&\quad + \sqrt{\frac{1}{(n-2)(n-3)}}w_3' + \ldots + \sqrt{\frac{1}{2}}w_{n-1}' + \sqrt{\frac{1}{n}}w_n'.
\end{aligned}
\tag{8}
$$

W układzie współrzędnych $(w_1', w_2', \ldots, w_n')$ warunek $w_1+w_2+\ldots+w_n = 0$ sprowadza się do postaci $w_n' = 0$. Możemy więc ograniczyć się do rozważania $(n-1)$-wymiarowej przestrzeni z układem współrzędnych $(w_1', w_2', \ldots, w_{n-1}')$. W tym układzie współrzędnych zmienna losowa $(W_1, W_2, \ldots, W_n)$ daje się przedstawić np. w postaci $(W_1', W_2', \ldots, W_{n-1}')$.

Prawdopodobieństwo układu nierówności (5) w układzie $(w_1', w_2', \ldots, w_{n-1}')$ wyraża się przez całkę

$$\int dF'(w_1', w_2', \ldots, w_{n-1}'), \tag{9}$$

gdzie $F'(w_1', w_2', \ldots, w_{n-1}')$ jest dystrybuantą zmiennej losowej $(W_1', W_2', \ldots, W_{n-1}')$, a całkowanie rozciąga się na obszar określony przez warunki

$$\begin{aligned}
&-\sqrt{\frac{n-1}{n}}\,w_1' \geqslant 0,\\
&\sqrt{\frac{1}{n(n-1)}}\,w_1' - \sqrt{\frac{n-2}{n-1}}\,w_2' < 0,\\
&\ldots\ldots\ldots\ldots\ldots\ldots\\
&\sqrt{\frac{1}{n(n-1)}}\,w_1' + \sqrt{\frac{1}{(n-1)(n-2)}}\,w_2' + \ldots +\\
&\qquad + \sqrt{\frac{1}{3\cdot 2}}\,w_{n-2}' + \sqrt{\frac{1}{2\cdot 1}}\,w_{n-1}' \geqslant 0.
\end{aligned} \tag{10}$$

Znaki tych nierówności odpowiadają kolejnym znakom, jakie mamy w układzie nierówności (5).

Zauważmy, że normalne wszystkich n płaszczyzn określonych przez równania

$$\begin{aligned}
&-\sqrt{\frac{n-1}{n}}\,w_1' = 0,\\
&\sqrt{\frac{1}{n(n-1)}}\,w_1' - \sqrt{\frac{n-2}{n-1}}\,w_2' = 0,\\
&\ldots\ldots\ldots\ldots\ldots\ldots\\
&\sqrt{\frac{1}{n(n-1)}}\,w_1' + \sqrt{\frac{1}{(n-1)(n-2)}}\,w_2' + \ldots +\\
&\qquad + \sqrt{\frac{1}{3\cdot 2}}\,w_{n-2}' + \sqrt{\frac{1}{2\cdot 1}}\,w_{n-1}' = 0.
\end{aligned} \tag{11}$$

tworzą ze sobą kąt, którego cosinus jest równy $-1/(n-1)$.

Stąd więc wynika, że znalezienie rozkładu zmiennej losowej (4) sprowadza się do obliczenia miar obszarów, na które płaszczyzny (11) rozcinają $(n-1)$-wymiarową przestrzeń.

Jeżeli dany jest rozkład zmiennej losowej $(X_1, X_2, \ldots, X_n)$, to możemy wyznaczyć rozkład zmiennej losowej $(W_1, W_2, \ldots, W_n)$, a zatem także miary prawdopodobieństwa obszarów, na które płaszczyzny (11) rozcinają przestrzeń. Jednak efektywne obliczenie wartości tych prawdopodobieństw może być bardzo kłopotliwe, w zależności od postaci rozkładu zmiennej losowej $(X_1, X_2, \ldots, X_n)$. W niniejszej pracy ograniczymy się do rozważania tylko pewnej klasy rozkładów zmiennej losowej $(X_1, X_2, \ldots, X_n)$ o takiej własności, że dla wszystkich rozkładów tej klasy zmienna losowa (4) ma ten sam rozkład, co dla rozkładu jednostajnego w kuli jednostkowej ze środkiem w początku układu.

Do tak określonej klasy rozkładów należą rozkłady, które są niezmiennicze względem grupy obrotów dookoła osi o równaniu $x_1 = x_2 = \ldots = x_n$. W szczególności więc do tej klasy należy łączny rozkład niezależnych zmiennych losowych o jednakowym rozkładzie.

3. Obecnie sformułujemy założenia i wyprowadzimy kilka wzorów, z których będziemy korzystać przy dowodzie zasadniczego twierdzenia.

Niech $\mathscr{X}_m$, gdzie $m \geqslant 1$, będzie m-wymiarową przestrzenią euklidesową, a $H_1 = 0, H_2 = 0, \ldots, H_\gamma = 0$ $(1 \leqslant \gamma \leqslant m+1)$ układem hiperpłaszczyzn o równaniach $H_i = a_{i1}x_1 + a_{i2}x_2 + \ldots + a_{im}x_m = 0$ $(i = 1, 2, \ldots, \gamma)$. Płaszczyzny te rozcinają przestrzeń $\mathscr{X}_m$ na skończoną ilość obszarów. Każdy taki obszar da się przedstawić jednoznacznie w postaci

$$e_1 H_1 > 0, \quad e_2 H_2 > 0, \quad e_\gamma H_\gamma > 0, \tag{12}$$

gdzie $e_i = +1$ lub $e_i = -1$ $(i = 1, 2, \ldots, \gamma)$. Niech w przestrzeni $\mathscr{X}_m$ będzie określony pewien rozkład prawdopodobieństwa P.

ZAŁOŻENIE A. *Jakikolwiek obierzemy ciąg $(e_1, e_2, \ldots, e_\gamma)$ liczb ± 1, miary prawdopodobieństwa obszarów wyznaczonych przez warunki*

$$e_1 H_1 > 0, \quad e_2 H_2 > 0, \quad \ldots, \quad e_\gamma H_\gamma > 0 \tag{13}$$

oraz

$$-e_1 H_1 > 0, \quad -e_2 H_2 > 0, \quad \ldots, \quad -e_\gamma H_\gamma > 0 \tag{14}$$

są równe

ZAŁOŻENIE B. *Miara prawdopodobieństwa obszaru*

$$e_1 H_1 > 0, \quad e_2 H_2 > 0, \quad \ldots, \quad e_\gamma H_\gamma > 0 \tag{15}$$

zależy tylko od ilości wyrazów $+1$ oraz ilości wyrazów -1 w ciągu $(e_1, e_2, \ldots, e_\gamma)$, gdzie $e_i = \pm 1$ $(i = 1, 2, \ldots, \gamma)$.

ZAŁOŻENIE C. *Masa prawdopodobieństwa rozłożona na hiperpłaszczyznach $H_1 = 0, H_2 = 0, \ldots, H_\gamma = 0$ jest równa zero.*

Jasne jest, że jeżeli dla danej miary prawdopodobieństwa P założenia A, B i C są spełnione dla układu hiperpłaszczyzn $H_1 = 0, H_2 = 0, \ldots, H_\gamma = 0$, to są one spełnione dla danej miary P także dla każdego układu $H_{i_1} = 0, H_{i_2} = 0, \ldots, H_{i_{\gamma'}} = 0$ hiperpłaszczyzn spośród $H_1 = 0$, $H_2 = 0, \ldots, H_\gamma = 0$, gdzie $0 < \gamma' < \gamma$.

Wystarczy pokazać, że tak jest dla $\gamma' = \gamma-1$. Dowód wynika z relacji

(16) $$P\{e_1H_{i_1} > 0, \ldots, e_{\gamma-1}H_{i_{\gamma-1}} > 0, H_{i_\gamma} > 0\} + \\ + P\{e_1H_{i_1} > 0, \ldots, e_{\gamma-1}H_{i_{\gamma-1}} > 0, H_{i_\gamma} < 0\} = \\ = P\{e_1H_{i_1} > 0, \ldots, e_{\gamma-1}H_{i_{\gamma-1}} > 0\}.$$

Istotnie, założenia A, B i C spełniają funkcje po lewej stronie wzoru (16). Spełnia je więc i funkcja po prawej stronie.

Z założenia B wynika, że miarę prawdopodobieństwa obszaru określonego przez nierówności $e_1H_1 > 0$, $e_2H_2 > 0, \ldots, e_\gamma H_\gamma > 0$ możemy oznaczyć symbolem

(17) $$V_{\beta,\gamma}^{(m)}(H_1, H_2, \ldots, H_\gamma, P),$$

gdzie β oznacza ilość wyrazów $+1$, a $\gamma-\beta$ ilość wyrazów -1 w ciągu $(e_1, e_2, \ldots, e_\gamma)$. Dla skrótu będziemy pisać $V_{\beta,\gamma}$.

Uzupełnimy jeszcze określenie funkcji (17) dla $\beta = \gamma = 0$ i $m = 0$, przyjmując $V_{0,0}^{(m)} = 1$ $(m = 0, 1, \ldots)$.

TWIERDZENIE 1. *Jeżeli założenia* A, B i C *są spełnione, wówczas funkcja* $V_{\beta,\gamma} = V_{\beta,\gamma}^{(m)}(H_1, H_2, \ldots, H_\gamma, P)$ *ma następujące właściwości:*

(18) $$\sum_{i=0}^{\gamma} \binom{\gamma}{i} V_{i,\gamma} = 1,$$

(19) $$V_{\beta+1,\gamma} + V_{\beta,\gamma} = V_{\beta,\gamma-1},$$

(20) $$V_{\beta,\gamma} = V_{\gamma-\beta,\gamma},$$

(21) $$V_{\beta,2\beta+1} = V_{\beta+1,2\beta+1} = \tfrac{1}{2}V_{\beta,2\beta},$$

(22) $$V_{\beta,\gamma} = \sum_{i=0}^{\gamma'-\gamma} \binom{\gamma'-\gamma}{i} V_{\beta+i,\gamma'}, \qquad \gamma' \geqslant \gamma \geqslant 0.$$

Dowód. Z założenia C wynika, że

(23) $$\sum P\{e_1H_1 > 0, e_2H_2 > 0, \ldots, e_\gamma H_\gamma > 0\} = 1,$$

gdzie sumowanie jest rozciągnięte na wszystkie ciągi $(e_1, e_2, \ldots, e_\gamma)$ liczb ± 1. Z założenia B natomiast wynika, że miara prawdopodobieństwa obszaru $e_1H_1 > 0$, $e_2H_2 > 0, \ldots, e_\gamma H_\gamma > 0$ zależy tylko od ilości wyrazów $+1$ oraz ilości wyrazów -1 w ciągu $(e_1, e_2, \ldots, e_\gamma)$. A zatem wzór (23) można napisać w postaci (18).

Aby udowodnić wzór (19), zauważmy, że

(24) $$V_{\beta+1,\gamma} = P\{H_1 > 0, \ldots, H_\beta > 0, H_{\beta+1} > 0, H_{\beta+2} < 0, \ldots, H_\gamma < 0\},$$

(25) $$V_{\beta,\gamma} = P\{H_1 > 0, \ldots, H_\beta > 0, H_{\beta+1} < 0, H_{\beta+2} < 0, \ldots, H_\gamma < 0\}.$$

Dodając stronami (24) i (25), otrzymujemy

$$(26)\qquad V_{\beta+1,\gamma}+V_{\beta,\gamma} = P\{H_1>0,\ldots,H_\beta>0,H_{\beta+2}>0,\ldots,H_\gamma<0\}.$$

Ciąg $(e_1, e_2, \ldots, e_\gamma) = (+1, \ldots, +1, -1, \ldots, -1)$, określający obszar $H_1>0, \ldots, H_\beta>0$, $H_{\beta+2}<0, \ldots, H_\gamma<0$, ma β elementów równych $+1$ oraz $\gamma-\beta-1$ elementów równych -1. Zatem prawą stronę wzoru (26) można oznaczyć przez $V_{\beta,\gamma-1}$. W ten sposób otrzymujemy wzór (19).

Wzór (20) jest oczywisty wobec założenia A.

Podstawmy teraz $\gamma = 2\beta+1$ we wzorze (19) oraz (20),

$$(27)\qquad V_{\beta+1,2\beta+1}+V_{\beta,2\beta+1} = V_{\beta,2\beta},$$

$$(28)\qquad V_{\beta,2\beta+1} = V_{\beta+1,2\beta+1}.$$

Z zestawienia wzorów (27) i (28) wynika natychmiast wzór (21).

Następny wzór (22) można otrzymać ze wzoru (19). Wykonujemy najpierw następujące proste przekształcenia algebraiczne:

$$(29)\qquad \sum_{i=0}^{\gamma'-\gamma}\binom{\gamma'-\gamma}{i}V_{\beta+i,\gamma'} = V_{\beta,\gamma'}+\sum_{i=1}^{\gamma'-\gamma-1}\binom{\gamma'-\gamma}{i}V_{\beta+i,\gamma'}+V_{\beta+\gamma'-\gamma,\gamma'} =$$

$$=\binom{\gamma'-\gamma-1}{0}V_{\beta,\gamma'}+\sum_{i=1}^{\gamma'-\gamma-1}\left[\binom{\gamma'-\gamma-1}{i}+\binom{\gamma'-\gamma-1}{i-1}\right]V_{\beta+i,\gamma'}+\binom{\gamma'-\gamma-1}{\gamma'-\gamma-1}V_{\beta+\gamma'-\gamma,\gamma'}=$$

$$= \sum_{i=0}^{\gamma'-\gamma-1}\binom{\gamma'-\gamma-1}{i}V_{\beta+i,\gamma'}+\sum_{i=0}^{\gamma'-\gamma-1}\binom{\gamma'-\gamma-1}{i}V_{\beta+i+1,\gamma'} =$$

$$= \sum_{i=0}^{\gamma'-\gamma-1}\binom{\gamma'-\gamma-1}{i}[V_{\beta+i,\gamma'}+V_{\beta+1+i,\gamma'}].$$

Korzystając ze wzoru (19), otrzymujemy

$$(30)\qquad \sum_{i=0}^{\gamma'-\gamma}\binom{\gamma'-\gamma}{i}V_{\beta+i,\gamma'} = \sum_{i=0}^{\gamma'-\gamma-1}\binom{\gamma'-\gamma-1}{i}V_{\beta+i,\gamma'-1}.$$

Analogicznie możemy pokazać, że dla $\delta = 1, 2, \ldots, \gamma'-\gamma$

$$(31)\qquad \sum_{i=0}^{\gamma'-\gamma}\binom{\gamma'-\gamma}{i}V_{\beta+i,\gamma'} = \sum_{i=0}^{\gamma'-\gamma-\delta}\binom{\gamma'-\gamma-\delta}{i}V_{\beta+i,\gamma'-\delta}.$$

Jeżeli za δ podstawimy teraz $\gamma'-\gamma$, otrzymujemy wzór (22).

Korzystając z wzorów (18)-(22) udowodnimy twierdzenie o zależnościach między funkcjami $V_{\beta,\gamma}$. Twierdzenie to mówi, że każda funkcja $V_{\beta,\gamma}$ daje się wyrazić jako kombinacja liniowa funkcji $V_{0,2}, V_{0,4}, V_{0,6}, \ldots$

Drugi wzór, tj. (33), w tym twierdzeniu jest szczególnym przypadkiem wzoru, jaki podał David [1].

TWIERDZENIE 2. *Jeżeli założenia* A, B *i* C *są spełnione, to dla funkcji* $V_{\beta,\gamma} = V_{\beta,\gamma}^{(m)}(H_1, H_2, \ldots, H_\gamma, P)$ *mamy następujące dwie relacje:*

$$V_{\beta,\gamma} = \sum_{i=0}^{\beta} (-1)^{\beta+i} \binom{\beta}{i} V_{0,\gamma-i}, \tag{32}$$

$$V_{0,2\beta+1} = \tfrac{1}{2} \sum_{i=0}^{2\beta} (-1)^{i} \binom{2\beta+1}{i} V_{0,i}. \tag{33}$$

Dowód. Pierwszy z tych wzorów udowodnimy w oparciu o wzór (19) za pomocą indukcji matematycznej. Wzór (32) jest prawdziwy dla $\beta = 0$ i dowolnego $\gamma \geqslant 0$. Pokażemy, że ze słuszności wzoru (32) dla γ i $\beta = 0, 1, \ldots, \gamma$ wynika słuszność wzoru (32) dla $\gamma+1$ i $\beta = 1, 2, \ldots, \gamma+1$. Istotnie, w myśl wzoru (19)

$$V_{\beta+1,\gamma+1} + V_{\beta,\gamma+1} = V_{\beta,\gamma}, \tag{34}$$

a stąd

$$V_{\beta+1,\gamma+1} = V_{\beta,\gamma} - V_{\beta,\gamma+1}. \tag{35}$$

Po wykonaniu elementarnych przekształceń

$$\begin{aligned} V_{\beta+1,\gamma+1} &= \sum_{i=0}^{\beta} (-1)^{\beta+i} \binom{\beta}{i} V_{0,\gamma-i} - \sum_{i=0}^{\beta} (-1)^{\beta+i} \binom{\beta}{i} V_{0,\gamma+1-i} = \\ &= \sum_{i=0}^{\beta+1} (-1)^{\beta+1+i} \binom{\beta+1}{i} V_{0,\gamma+1-i}, \qquad \text{c.b.d.o.} \end{aligned} \tag{36}$$

Następny wzór (33) wyprowadzimy ze wzoru (32). Napiszemy wzór (32) w postaci

$$\begin{aligned} V_{2\beta+1,2\beta+1} &= \sum_{i=0}^{2\beta+1} (-1)^{1+i} \binom{2\beta+1}{i} V_{0,2\beta+1-i} = \\ &= -V_{0,2\beta+1} + \sum_{i=0}^{2\beta} (-1)^{i} \binom{2\beta+1}{i+1} V_{0,2\beta-i}. \end{aligned} \tag{37}$$

Biorąc teraz pod uwagę, że $V_{2\beta+1,2\beta+1} = V_{0,2\beta+1}$, otrzymujemy przekształcając wzór (37)

$$\begin{aligned} V_{0,2\beta+1} &= \tfrac{1}{2} \sum_{i=0}^{2\beta} (-1)^{i} \binom{2\beta+1}{i+1} V_{0,2\beta-i} = \\ &= \tfrac{1}{2} \sum_{i=0}^{2\beta} (-1)^{2\beta-i} \binom{2\beta+1}{2\beta+1-i} V_{0,i} = \\ &= \tfrac{1}{2} \sum_{i=0}^{2\beta} (-1)^{i} \binom{2\beta+1}{i} V_{0,i}. \end{aligned} \tag{38}$$

Wzór (33) jest zatem udowodniony.

4. Z tego, co powiedzieliśmy w rozdziale 2, wynika, że dla rozwiązania postawionego zagadnienia, tzn. znalezienia rozkładu zmiennej losowej (4) dla klasy rozkładów zdefiniowanej w rozdziale 2, można założyć, iż zmienna losowa $(W_1', W_2', \ldots, W_{n-1}')$ ma w układzie $(w_1', w_2', \ldots, w_{n-1}')$ rozkład jednostajny w kuli jednostkowej ze środkiem w początku układu współrzędnych. Przy tych założeniach znalezienie rozkładu zmiennej losowej (4) sprowadza się do obliczenia miar prawdopodobieństwa obszarów, na które płaszczyzny (11) rozcinają kulę jednostkową ze środkiem w początku układu. Przy obliczaniu miar tych obszarów możemy, jak zaraz pokażemy, wykorzystać wyniki poprzedniego rozdziału, gdyż spełnione są warunki A, B i C.

Niech w m-wymiarowej przestrzeni euklidesowej $\mathscr{X}_m$ określony będzie rozkład jednostajny w kuli jednostkowej ze środkiem w początku układu oraz niech danych będzie $\gamma (0 < \gamma \leqslant m+1)$ hiperpłaszczyzn przechodzących przez środek kuli. Załóżmy, że normalne wszystkich tych hiperpłaszczyzn tworzą ze sobą ten sam kąt. W szczególności, jeżeli $\gamma = m+1$, to $\theta = \arccos(-1/m)$.

Udowodnimy, że warunki A, B i C są spełnione. W dowodzie będziemy korzystali z następującego oczywistego lematu.

LEMAT. *Niech dane będą w przestrzeni* $\mathscr{X}_m$ *dwa układy hiperpłaszczyzn* $H_1 = 0, H_2 = 0, \ldots, H_\gamma = 0$ *oraz* $H_1^* = 0, H_2^* = 0, \ldots, H_\gamma^* = 0$. *Załóżmy, że hiperpłaszczyzny te przechodzą przez środek pewnej kuli. Niech* $\|\sphericalangle(H_i, H_j)\|$ *oraz* $\|\sphericalangle(H_i^*, H_j^*)\|$ $(i, j = 1, 2, \ldots, \gamma)$ *oznaczają odpowiednio dla pierwszego i dla drugiego układu hiperpłaszczyzn macierze kątów, jakie tworzą ze sobą ich normalne. Jeżeli macierze* $\|\sphericalangle(H_i, H_j)\|$ *oraz* $\|\sphericalangle(H_i^*, H_j^*)\|$ *są równe (porządek kolumn nie odgrywa roli), to objętość brył, jakie stożek* $H_i > 0, H_2 > 0, \ldots, H_\gamma > 0$ *oraz stożek* $H_1^* > 0, H_2^* > 0, \ldots, H_\gamma^* > 0$ *wycina z kuli, są równe.*

Z podanego lematu widać, że założenie A jest spełnione. Istotnie, niech dane będą hiperpłaszczyzny $H_1 = 0, H_2 = 0, \ldots, H_\gamma = 0$. W myśl tego lematu miary prawdopodobieństwa obszarów $e_1 H_1 > 0, e_2 H_2 > 0, \ldots, e_\gamma H_\gamma > 0$ oraz $-e_1 H_1 > 0, -e_2 H_2 > 0, \ldots, -e_\gamma H_\gamma > 0$ są równe, gdyż jak łatwo zauważyć, zachodzi równość $\|\sphericalangle(e_i H_i, e_j H_j)\| = \|\sphericalangle(-e_i H_i, -e_j H_j)\|$. Nie korzystamy tutaj z założenia, że normalne hiperpłaszczyzn tworzą ten sam kąt.

Przejdźmy teraz do założenia B. Załóżmy teraz, że normalne hiperpłaszczyzn $H_1 = 0, H_2 = 0, \ldots, H_\gamma = 0$ tworzą ze sobą ten sam kąt. Obierzmy dwa dowolne ciągi $(e_1, e_2, \ldots, e_\gamma)$ oraz $(e_1^*, e_2^*, \ldots, e_\gamma^*)$ liczb ± 1 z taką samą ilością wyrazów $+1$. Wykażemy, że miary prawdopodobieństwa obszarów $e_1 H_1 > 0, e_2 H_2 > 0, \ldots, e_\gamma H_\gamma > 0$ oraz $e_1^* H_1 > 0, e_2^* H_2 > 0, \ldots, e_\gamma^* H_\gamma > 0$ są równe. Aby to udowodnić, wystarczy w myśl podanego lematu pokazać, że zachodzi równość

$$(39) \qquad \| \sphericalangle (e_i H_i, e_j H_j) \| = \| \sphericalangle (e_i^* H_i, e_j^* H_j) \|.$$

Dowód wzoru (39) wynika stąd, że: 1° Normalne hiperpłaszczyzn $H_1 = 0, H_2 = 0, \ldots, H_\gamma = 0$ tworzą ze sobą ten sam kąt. 2° Ilość wyrazów $+1$ w ciągach $(e_1, e_2, \ldots, e_\gamma)$ oraz $(e_1^*, e_2^*, \ldots, e_\gamma^*)$ jest taka sama.

Spełnione jest także założenie C. Istotnie, wynika to natychmiast z postaci rozkładu prawdopodobieństwa określonego w przestrzeni $\mathcal{X}_m$.

Jasne jest, że miara prawdopodobieństwa obszaru

$$(40) \qquad e_1 H_1 > 0, \quad e_2 H_2 > 0, \quad \ldots, \quad e_\gamma H_\gamma > 0$$

utworzonego przez hiperpłaszczyzny $H_1 = 0, H_2 = 0, \ldots, H_\gamma = 0$, których normalne tworzą ze sobą ten sam kąt θ, można oznaczyć przez

$$(41) \qquad V_{\beta,\gamma}^{(m)}(\theta),$$

gdzie β, γ i m są określone jak w (17). Innymi słowy, kąt θ oraz ilość wyrazów $+1$ w ciągu $(e_1, e_2, \ldots, e_\gamma)$ liczb ± 1 wyznaczają obszar o jednoznacznie określonej mierze prawdopodobieństwa.

Łatwo pokazać, że funkcja (41) ma następujące własności: jeżeli w γ-wymiarowej przestrzeni istnieje γ hiperpłaszczyzn, których normalne tworzą ze sobą ten sam kąt θ, to

$$(42) \qquad V_{\beta,\gamma}^{(\gamma+\delta)}(\theta) = V_{\beta,\gamma}^{(\gamma)}(\theta) \qquad (\delta = 1, 2, \ldots).$$

W szczególności

$$(43) \quad V_{\beta,\gamma}^{(\gamma+\delta)}\left[\arccos\left(-\frac{1}{\gamma-1}\right)\right] = V_{\beta,\gamma}^{(\gamma-1)}\left[\arccos\left(-\frac{1}{\gamma-1}\right)\right] \quad (\gamma = 0, 1, \ldots),$$

oraz

$$(44) \quad V_{0,\gamma}^{(\gamma-1)}\left[\arccos\left(-\frac{1}{\gamma-1}\right)\right] = V_{\gamma,\gamma}^{(\gamma-1)}\left[\arccos\left(-\frac{1}{\gamma-1}\right)\right] = 0 \quad (\gamma = 2, 3, \ldots).$$

Wzorów (42) i (43) dowodzić nie będziemy. Dowody tych wzorów sprowadzają się do znalezienia odpowiednich całek dla wyrażeń występujących w tych wzorach.

Wzór (44) wynika stąd, że ciągi $(e_1, e_2, \ldots, e_{\gamma+1}) = (+1, +1, \ldots, +1)$ oraz $(e_1, e_2, \ldots, e_{\gamma+1}) = (-1, -1, \ldots, -1)$ wyznaczają zbiory puste w γ-wymiarowych przestrzeniach.

Zestawiając udowodnione twierdzenia 1 i 2 z wzorami (42)-(44), otrzymujemy twierdzenie, które daje rozwiązanie postawionego zadania.

TWIERDZENIE 3. *Dla klasy rozkładów określonej w rozdziale 2 prawdopodobieństwo układu nierówności*

$$(45) \qquad e_1 W_1 > 0, \quad e_2 W_2 > 0, \quad \ldots, \quad e_n W_n > 0$$

jest tylko funkcją β (ilości wyrazów $+1$) w ciągu $(e_1, e_2, \ldots, e_n)$ liczb ± 1 i równa się

$$V_{\beta,n}^{(n-1)}\left[\arccos\left(-\frac{1}{n-1}\right)\right]. \tag{46}$$

Jeżeli w m-wymiarowej przestrzeni euklidesowej mamy układ γ hiperpłaszczyzn, których normalne tworzą ze sobą ten sam kąt, to dla funkcji $V_{\beta,\gamma}^{(m)}(\theta)$ zachodzą równości

$$V_{0,2\beta+1}^{(m)}(\theta) = \tfrac{1}{2}\sum_{i=0}^{2\beta}(-1)^i\binom{2\beta+1}{i}V_{0,i}^{(i)}(\theta) \tag{47}$$

oraz (18)-(22), (32), (42)-(44).

5. Z twierdzenia 3 wynika, że rozkład zmiennej losowej (4) daje się wyrazić liniowo za pomocą wyrażeń $V_{0,2m}^{(2m)}(\theta)$, gdzie $m = 1, 2, \ldots$

Tak np. dla $n = 2, 3, 4, 5$ otrzymujemy następujące formuły na obliczenie rozkładu wskaźników przyrodniczych $(W_1, W_2, \ldots, W_n)$.

1° Dla $n = 2$

$$V_{1,2}^{(1)}(\theta) = \tfrac{1}{2} - V_{0,2}^{(1)}(\theta), \tag{48}$$

gdzie $\theta = \arccos(-1)$.

2° Dla $n = 3$

$$\begin{aligned} V_{0,3}^{(2)}(\theta) &= -\tfrac{1}{4} + \tfrac{3}{2}V_{0,2}^{(2)}(\theta) \\ V_{1,3}^{(2)}(\theta) &= \tfrac{1}{4} - \tfrac{1}{2}V_{0,2}^{(2)}(\theta), \end{aligned} \tag{49}$$

gdzie $\theta = \arccos(-\tfrac{1}{2})$.

3° Dla $n = 4$

$$\begin{aligned} V_{1,4}^{(3)}(\theta) &= -\tfrac{1}{4} + \tfrac{3}{2}V_{0,2}^{(2)}(\theta) - V_{0,4}^{(4)}(\theta) \\ V_{2,4}^{(3)}(\theta) &= \tfrac{1}{2} - 2V_{0,2}^{(2)}(\theta) + V_{0,4}^{(4)}(\theta), \end{aligned} \tag{50}$$

gdzie $\theta = \arccos(-\tfrac{1}{3})$.

4° Dla $n = 5$

$$\begin{aligned} V_{0,5}^{(4)}(\theta) &= \tfrac{1}{2} - \tfrac{5}{2}V_{0,2}^{(2)}(\theta) + \tfrac{5}{2}V_{0,4}^{(4)}(\theta), \\ V_{1,5}^{(4)}(\theta) &= -\tfrac{1}{2} + \tfrac{5}{2}V_{0,2}^{(2)}(\theta) - \tfrac{3}{2}V_{0,4}^{(4)}(\theta), \\ V_{2,5}^{(4)}(\theta) &= \tfrac{1}{4} - V_{0,2}^{(2)}(\theta) + \tfrac{1}{2}V_{0,4}^{(4)}(\theta), \end{aligned} \tag{51}$$

gdzie $\theta = \arccos(-\tfrac{1}{4})$.

Zauważmy teraz, że ze wzoru (44) wynika, że $V_{0,2}^{(1)}[\arccos(-1)] = 0$. Stąd i ze wzoru (48) dla dwuwymiarowej przestrzeni, $n = 2$, otrzymujemy

$$V_{1,2}^{(1)}[\arccos(-1)] = \tfrac{1}{2}. \tag{52}$$

Dla przestrzeni trójwymiarowej, $n = 3$, ze wzoru (49) i z tego, że $V_{0,3}^{(2)}[\arccos(-\frac{1}{2})] = 0$, otrzymujemy $V_{0,2}^{(2)}[\arccos(-\frac{1}{2})] = \frac{1}{6}$. Stąd $V_{1,3}^{(2)}[\arccos(-\frac{1}{2})] = \frac{1}{4}-\frac{1}{2}V_{0,2}^{(2)}[\arccos(-\frac{1}{2})] = \frac{1}{6}$. Mamy więc

$$V_{1,3}^{(2)}[\arccos(-\tfrac{1}{2})] = \tfrac{1}{6}. \tag{53}$$

Dla czterowymiarowej przestrzeni, $n = 4$, analogicznie jak poprzednio $V_{0,4}^{(3)}[\arccos(-\frac{1}{3})] = 0$ w myśl wzoru (44).

A zatem

$$\begin{aligned} V_{1,4}^{(3)}(\theta) &= -\tfrac{1}{4}+\tfrac{3}{2}V_{0,2}^{(2)}(\theta), \\ V_{2,4}^{(3)}(\theta) &= \tfrac{1}{2}-2V_{0,2}^{(2)}(\theta), \end{aligned} \tag{54}$$

gdzie $\theta = \arccos(-\frac{1}{3})$.

Wartość funkcji $V_{0,2}^{(2)}[\arccos(-\frac{1}{3})]$ można odczytać z tablic funkcji trygonometrycznych. Wynosi ona 0,19592.

Po podstawieniu otrzymujemy

$$\begin{aligned} V_{1,4}^{(3)}[\arccos(-\tfrac{1}{3})] &= 0{,}04388, \\ V_{2,4}^{(3)}[\arccos(-\tfrac{1}{3})] &= 0{,}10816. \end{aligned} \tag{55}$$

Dla przestrzeni pięciowymiarowej, $n = 5$, otrzymujemy w analogiczny sposób jak poprzednio następujące formuły:

$$\begin{aligned} V_{1,5}^{(4)}(\theta) &= -\tfrac{1}{5}+V_{0,2}^{(2)}(\theta), \\ V_{2,5}^{(4)}(\theta) &= \tfrac{3}{20}-\tfrac{1}{2}V_{0,2}^{(2)}(\theta), \end{aligned} \tag{56}$$

gdzie $\theta = \arccos(-\frac{1}{4})$.

Ponieważ $V_{0,2}^{(2)}[\arccos(-\frac{1}{4})] = 0{,}20979$, otrzymujemy po podstawieniu

$$\begin{aligned} V_{1,5}^{(4)}[\arccos(-\tfrac{1}{4})] &= 0{,}00979, \\ V_{2,5}^{(4)}[\arccos(-\tfrac{1}{4})] &= 0{,}04511. \end{aligned} \tag{57}$$

Aby otrzymać pozostałe wartości funkcji $V_{\beta,n}^{(n-1)}\left[\arccos\left(-\frac{1}{n-1}\right)\right]$, należy skorzystać ze wzoru (20). Otrzymane wartości podano w tablicy 1.

TABLICA 1

Wartości funkcji $V_{\beta,n}^{(n-1)}\left[\arccos\left(-\frac{1}{n-1}\right)\right]$

n \ β	0	1	2	3	4	5
2	0	0,5	0			
3	0	0,16667	0,16667	0		
4	0	0,04388	0,10816	0,04388	0	
5	0	0,00979	0,04511	0,04511	0,00979	0

Rozkład ogólnej liczby dodatnich i ujemnych wskaźników przyrodniczych Perkala podany jest w tablicy 2.

TABLICA 2

Wartości funkcji $\binom{n}{\beta} V_{\beta,n}^{(n-1)}\left[\arccos\left(-\frac{1}{n-1}\right)\right]$

n \ β	0	1	2	3	4	5	Razem
2	0	1	0				1
3	0	0,5	0,5	0			1
4	0	0,17552	0,64896	0,17552	0		1
5	0	0,04895	0,45110	0,45110	0,04895	0	1

Rozkład wskaźników przyrodniczych Perkala dla $n \geqslant 6$ nie daje się wyznaczyć, gdyż wartości funkcji $V_{0,\gamma}^{(\gamma)}(\theta)$ dla $\gamma = 4, 6, 8, \ldots$ nie zostały dotąd stablicowane dla kątów, jakie tutaj występują. Prawdopodobieństwa te jednak dadzą się obliczyć za pomocą wzoru rekurencyjnego opublikowanego przez Rubena [4].

Jak już na wstępie podawaliśmy, rozkłady wskaźników przyrodniczych Perkala dla $n \geqslant 6$ zostaną opublikowane w oddzielnej publikacji.

Prace cytowane

[1] F. N. David, *A note on the evaluation of the multivariate normal integral*, Biometrika 40 (1953), str. 485-459.

[2] H. Hotelling, *Analysis of a complex of statistical variables into principal components*, Journal of Educational Psychology 24 (1933), str. 417-441 i 498-520.

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

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

[5] L. Schläfli, *On the multiple integral* $\overset{n}{\int} dxdy \ldots dz$ *whose limits are* $p_1 = a_1x+b_1y+ \ldots +h_1z > 0$, $p_2 > 0, \ldots, p_n > 0$, *and* $x^2+y^2+ \ldots +z^2 < 1$, Quart. J. Math. 2 (1858), str. 264-301, 3 (1860), str. 54-68 i 97-108.

[6] Ch. Spearman, *The abilities of man*, New York 1927.

[7] L. L. Thurstone, *Multiple-factor analysis*, Chicago 1947.

Praca wpłynęła 15. 6. 1965

В. КЛЬОНЕЦКИ (Познань)

О РАСПРЕДЕЛЕНИИ ЗНАКОВ ИНДЕКСОВ ПЭРКАЛЯ

РЕЗЮМЕ

В этой работе автор показывает, как найти распределение знаков естественных индексов Пэркаля [3] для классов распределений инвариантных относительно группы вращений вокруг оси $x_1 = x_2 = \ldots = x_n$.

Пусть $\mathfrak{X}_m$ обозначает m-мерное евклидово пространство, в котором определено равномерное распределение вероятности в единичном шаре, которого центр лежит в начале системы координат. В пространстве $\mathfrak{X}_m$ дано γ гиперплоскостей $(1 \leqslant \gamma \leqslant m+1)$ вида $H_i = a_{i1}x_1 + a_{i2}x_2 + \ldots + a_{im}x_m = 0$ $(i = 1, 2, \ldots, \gamma)$. Предположим, что нормали этих гиперплоскостей образуют между собой равные углы θ. При этих предположениях имеют место соотношения (32) и (33), причём $V_{\beta,\gamma}^{(m)}(\theta)$ является мерой вероятности конуса $e_1H_1 > 0, e_2H_2 > 0, \ldots, e_\gamma H_\gamma > 0$. β обозначает число выражений $+1$ в последовательности $(e_1, e_2, \ldots, e_\gamma)$, состоящей из чисел ± 1. Эти результаты были получены путём обобщения метода Давида [1]. Их можно было также получить другим путём, именно из рекуррентной формулы Рубена [4] для гиперсферических симплексов.

Из теоремы 3 следует, что распределение знаков индексов Пэркаля может быть представлено как линейная комбинация функций $V_{0,2}^{(2)}(\theta)$, $V_{0,4}^{(4)}(\theta), \ldots$

В етой работе дано распределение индексов Пэркаля в случае дву-, трёх-, четырёх- и пятимерной случайной величины.

W. KLONECKI (Poznań)

ON THE DISTRIBUTION OF SIGNS OF PERKAL INDICES

SUMMARY

The author shows how to obtain the distribution of the signs of Perkal's natural indices [3] for a class of distributions which remain invariant under the group of rotation around the line $x_1 = x_2 = \ldots = x_n$.

Let $\mathfrak{X}_m$ denote the m-dimensional euclidean space with uniform probability measure in the unit sphere centered at the origin. Let $H_1 = 0, H_2 = 0, \ldots, H_\gamma = 0$ denote hyperplanes, $H_i = a_{i1}x_1 + a_{i2}x_2 + \ldots + a_{im}x_m = 0$ $(i = 1, 2, \ldots, \gamma)$. The normals of any two of the hyperplanes are assumed to be equal, say to θ. Under these assumptions formulae (32) and (33) hold, where $V_{\beta,\gamma}^{(m)}(\theta)$ is the probability measure of the cone $e_1H_1 > 0, e_2H_2 > 0, \ldots, e_\gamma H_\gamma > 0$ and β is the number of $+1$'s and $\gamma - \beta$ the number of -1's in the sequance $(e_1, e_2, \ldots, e_\gamma)$, where $e_i = \pm 1$. Formulae (32) and (33) are obtained by generalizing a method originated by David [1]. However both formulae (32) and (33) could be obtained from Ruben's [4] recurrence formula for hyperspherical simplices.

It follows from theorem 3 that the distributions of the signs of Perkal's natural indices can be expressed as linear functions of the probabilities $V_{0,2}^{(2)}(\theta)$, $V_{0,4}^{(4)}(\theta), \ldots$

The distributions of the signs of natural indices are presented for two, three, four, and five-dimensional random variables.
