ZASTOSOWANIA MATEMATYKI
APPLICATIONES MATHEMATICAE
XI, 3 (1970)

E. PLESZCZYŃSKA (Warszawa)

WSPÓŁCZYNNIK ZALEŻNOŚCI MONOTONICZNEJ Z PRÓBKI (II)

1. Wstęp. W części (I) tej pracy [3] zdefiniowano klasę par monotonicznie zależnych ciągłych zmiennych losowych. W praktyce, przy badaniu współzależności zmiennych losowych możemy często przyjąć założenie, że para zmiennych losowych jest monotonicznie zależna. W szczególności zdarza się to w problemach dotyczących cech zastępczych, kiedy to kandydaturę pewnej cechy Z jako cechy zastępczej dla innej cechy W można postawić na ogół wtedy, gdy W i Z są monotonicznie zależne. Niejednokrotnie pomiar badanej cechy w problemach technicznych lub przyrodniczych można również traktować jako cechę zastępczą związaną monotonicznie z ową cechą badaną, gdyż różne metody pomiarowe polegają zwykle na wyznaczeniu wartości jakiejś monotonicznej funkcji mierzonej cechy, przy czym często nie udaje się wykluczyć losowych szumów.

W problemach praktycznych stawia się pytanie, jak silnie są zależne obie cechy i jak to badać statystycznie. W niniejszej pracy chcemy skomentować i uzupełnić dotąd stosowane metody statystycznego formułowania i rozwiązywania tego problemu przy założeniu monotonicznej zależności cech.

Proponowane metody powinny się okazać szczególnie przydatne wtedy, gdy dokładny pomiar wartości jednej z badanych cech jest bardzo trudny i kosztowny lub nawet zupełnie niemożliwy, tak że możemy mierzyć tylko wartości jakiejś nieznanej ściśle monotonicznej funkcji tej cechy. Istnieją wprawdzie metody, które pozwalają badać statystycznie zależność między dwiema cechami, gdy dokładne pomiary obu cech są trudne lub niemożliwe, ale nie są one dostatecznie efektywne w przypadku, gdy trudności dotyczą pomiaru tylko jednej z cech.

Praca stanowi kontynuację pracy [3], a więc definicje i wyniki uzyskane tam są tu używane bez podawania źródła; w szczególności stosujemy następujące oznaczenia: M, M^+, M^-, O_M, O_M^+, O_M^-, F, F^+, F^-, N, $H^{(2k+1)}$, A_n, $\Phi(a, b)$, $\Phi_W(a)$, $\Phi_Z(b)$, $\varphi(a, b)$, $\varphi_W(a)$, $\varphi_Z(b)$.

2. Modele statystyczne z parametrem charakteryzującym zależność. Współzależność pary zmiennych losowych (W, Z) przyjęto najczęściej wyrażać za pomocą współczynnika korelacji

$$\varrho_{W,Z} = \frac{E[(W-EW)(Z-EZ)]}{\sigma_W \cdot \sigma_Z}$$

(jeżeli momenty drugiego rzędu istnieją) i zazwyczaj sformułowanie problemu badania współzależności w języku statystycznym polega na weryfikacji hipotez o $\varrho_{W,Z}$ lub estymacji tego parametru. Jak wiadomo, w dwuwymiarowym rozkładzie normalnym $\varrho_{W,Z} = \varrho$. Jednakże $\varrho_{W,Z}$ nie jest parametrem określającym rodzaj zależności monotonicznej, gdyż $|\varrho_{W,Z}|$ jest równe 1 wtedy i tylko wtedy, gdy W jest liniową funkcją Z. Stąd np. w klasie N, jak dobrze wiadomo, nawet dla $|\varrho|$ dowolnie bliskich 1 współczynnik korelacji może być dowolnie bliski zera: dla każdego $\varepsilon > 0$ i dla zmiennej X o rozkładzie normalnym można znaleźć taką funkcję γ z klasy F, żeby $|\varrho_{X}, {}_{\gamma(X)}| < \varepsilon$ (np. przyjmując $\gamma(X) = X^{2k+1}$ dla dostatecznie dużych k); jeśli zaś (W, Z) ma rozkład normalny z parametrem ϱ, to łatwo obliczyć, że $\varrho_{W,\gamma(Z)} = \varrho \cdot \varrho_{Z,\gamma(Z)}$, a zatem $|\varrho_{W,\gamma(Z)}| < \varepsilon$. Wobec tego współczynnik korelacji nie jest odpowiednim parametrem w przypadku dowolnej zależności monotonicznej.

Spośród innych współczynników zależności, najbardziej znany jest stosunek korelacyjny

$$\eta_{W,Z} = \sqrt{\frac{\int\limits_B \big(E(W|Z=b)-EW\big)^2 d\Phi_Z(b)}{E(W-EW)^2}}$$

oraz współczynnik informacyjny

$$i_{W,Z} = \sqrt{1-e^{-2I(1:2)}},$$

gdzie dla rozkładów ciągłych

$$I(1:2) = \iint\limits_{A\times B} \varphi(a,b) \ln \frac{\varphi(a,b)}{\varphi_Z(b)\,\varphi_W(a)}\, da\, db;$$

oba te współczynniki są nieujemne (w dwuwymiarowym rozkładzie normalnym są one równe $|\varrho|$), więc również nie są odpowiednimi parametrami do badania zależności monotonicznej.

Wobec tego proponujemy wprowadzić pewien nowy współczynnik $\mu_{W,Z}$ charakteryzujący zależność monotoniczną, który jest parametrem określającym rodzaj zależności monotonicznej i który jest równy ϱ w pewnej rodzinie rozkładów z klasy N z parametrem ϱ zawierającej dwuwymiarowe rozkłady normalne.

Zdefiniujmy parametr $X_{0,5}$ rozkładu zmiennej losowej X jako środek przedziału złożonego z punktów u spełniających następujące warunki: $P(X \geqslant u) \geqslant \frac{1}{2}$ i $P(X \leqslant u) \geqslant \frac{1}{2}$. Niech

$$E(Z; W > W_{0,5}) = \iint\limits_{S(W_{0,5})} b d\Phi(a, b),$$

gdzie

$$S(W_{0,5}) = \{(a, b) : a \in A, b \in B, a > W_{0,5}\};$$

będziemy się posługiwać analogicznym zapisem także w przypadku, gdy warunek $W > W_{0,5}$ zastąpimy przez $W < W_{0,5}$, $Z > Z_{0,5}$, $Z < Z_{0,5}$. Natomiast przez $E(Z|W = W_{0,5})$ będziemy oznaczać warunkową wartość oczekiwaną Z, gdy $W = W_{0,5}$.

Możemy teraz zdefiniować:

$$\mu_{W,Z} = \frac{E(Z; W > W_{0,5}) - E(Z; W < W_{0,5}) + E(Z|W = W_{0,5})[P(W < W_{0,5}) - P(W > W_{0,5})]}{E(Z; Z > Z_{0,5}) - E(Z; Z < Z_{0,5}) + Z_{0,5}[P(Z < Z_{0,5}) - P(Z > Z_{0,5})]}.$$

Gdy $P(W = W_{0,5}) = 0$, ostatni wyraz licznika jest równy zero; w szczególności jest tak, gdy rozkład brzegowy W jest ciągły. Analogicznie jest dla mianownika przy $P(Z = Z_{0,5}) = 0$.

Podamy teraz własności $\mu_{W,Z}$, przy czym dla skrócenia zapisu rozkład zmiennej losowej (W, Z) będziemy oznaczać przez $R(W, Z)$.

WŁASNOŚĆ 1. $|\mu_{W,Z}| \leqslant 1$, *gdzie*

$$\mu_{W,Z} = \begin{cases} 1 & \text{dla } R(W, Z) \in O_M^+, \\ -1 & \text{dla } R(W, Z) \in O_M^- \end{cases}$$

oraz $|\mu_{W,Z}| < 1$ *dla rozkładów ciągłych.*

Dowód. Weźmy pod uwagę takie funkcje $w(a, b)$ dla $a \in A$ i $b \in B$, że $0 \leqslant w(a, b) \leqslant 1$, $\iint\limits_{A\times B} w(a, b) d\Phi(a, b) = \frac{1}{2}$ i istnieje $I = \iint\limits_{A\times B} bw(a, b) d\Phi(a, b)$. Oczywiście kres górny I jest równy $E(Z; Z > Z_{0,5}) + Z_{0,5}\big(\frac{1}{2} - P(Z > Z_{0,5})\big)$. Zatem w szczególności dla

$$w(a, b) = \begin{cases} 1 & \text{dla } a > W_{0,5}, \\ \dfrac{\frac{1}{2} - P(W > W_{0,5})}{P(W = W_{0,5})} & \text{dla } a = W_{0,5}, \text{ jeśli } P(W = W_{0,5}) > 0, \\ 0 & \text{dla pozostałych} \end{cases}$$

mamy

$$E(Z; W > W_{0,5}) + E(Z|W = W_{0,5})\big(\tfrac{1}{2} - P(W > W_{0,5})\big) \leqslant$$
$$\leqslant E(Z; Z > Z_{0,5}) + Z_{0,5}\big(\tfrac{1}{2} - P(Z > Z_{0,5})\big).$$

Podobnie rozpatrując kres dolny I stwierdzamy, że

$$E(Z;\, W > W_{0,5}) + E(Z|W = W_{0,5})\big(\tfrac{1}{2} - P(W > W_{0,5})\big) \geqslant$$
$$\geqslant E(Z;\, Z < Z_{0,5}) + Z_{0,5}\big(\tfrac{1}{2} - P(Z < Z_{0,5})\big).$$

W ten sam sposób można też ograniczyć z obu stron

$$E(Z;\, W < W_{0,5}) + E(Z|W = W_{0,5})\big(\tfrac{1}{2} - P(W < W_{0,5})\big),$$

a więc z definicji $\mu_{W,Z}$ mamy $|\mu_{W,Z}| \leqslant 1$.

Załóżmy teraz, że rozpatrujemy takie $R(W, Z)$, iż $P(W = W_{0,5}) = 0 = P(Z = Z_{0,5})$, a więc w szczególności rozkłady ciągłe oraz $R(W, Z) \in O_M^+$ lub O_M^-. Wtedy, jeśli $\mu_{W,Z} = 1$, to z definicji $\mu_{W,Z}$ po redukcji otrzymujemy

$$\int\limits_{a > W_{0,5}} \int\limits_{b < Z_{0,5}} b\, d\Phi(a, b) = \int\limits_{a < W_{0,5}} \int\limits_{b > Z_{0,5}} b\, d\Phi(a, b),$$

a stąd wynika, że

$$P(W > W_{0,5}, Z < Z_{0,5}) = P(W < W_{0,5}, Z > Z_{0,5}) = 0. \tag{2.1}$$

Z kolei, jeśli (2.1) jest spełnione, to $\mu_{W,Z} = 1$, a więc jest to warunek konieczny i dostateczny. Podobnie, warunkiem koniecznym i dostatecznym na to, aby $\mu_{W,Z} = -1$, jest

$$P(W > W_{0,5}, Z > Z_{0,5}) = P(W < W_{0,5}, Z < Z_{0,5}) = 0.$$

Warunki te nie są spełnione dla rozkładów ciągłych, są natomiast spełnione odpowiednio dla rozkładów z klasy O_M^+ i O_M^-, a stąd wynika już druga część tezy.

WŁASNOŚĆ 2. *Jeśli W i Z są niezależne, to $\mu_{W,Z} = 0$.*

Dowód. Gdy W i Z są niezależne, licznik $\mu_{W,Z}$ jest równy $EZ \cdot [P(W > W_{0,5}) - P(W < W_{0,5}) + P(W < W_{0,5}) - P(W > W_{0,5})] = 0$.

WŁASNOŚĆ 3. *$0 < \mu_{W,Z} < 1$, gdy $R(W, Z) \in M^+$ oraz $-1 < \mu_{W,Z} < 0$, gdy $R(W, Z) \in M^-$.*

Dowód. Wiadomo, że w rozkładzie ciągłym

$$\int\limits_B P(W > W_{0,5}|b)\, d\Phi_Z(b) = P(W > W_{0,5}) = \tfrac{1}{2}$$

oraz że

$$\int\limits_B b\, d\Phi_Z(b) = EZ.$$

Jeśli $R(W, Z) \in M^+$, to $P(W > W_{0,5}|b)$ jest ściśle rosnącą funkcją b a więc

$$E(Z;\, W > W_{0,5}) = \int\limits_B bP(W > W_{0,5}|b)\, d\Phi_Z(b) > \tfrac{1}{2} EZ.$$

Podobnie stwierdzamy, że $E(Z;\, W < W_{0,5}) < \frac{1}{2} EZ$, a więc, na podstawie własności 1, $0 < \mu_{W,Z} < 1$.

Druga część dowodu przebiega analogicznie.

Z własności 1,2 i 3 wynika, że $\mu_{W,Z}$ jest parametrem określającym rodzaj zależności monotonicznej między zmiennymi losowymi W i Z.

WŁASNOŚĆ 4. *Jeśli* (W, Z) *ma rozkład ciągły, to*

$$\mu_{\gamma(W),aZ+b} = \begin{cases} (\operatorname{sgn} a)\mu_{W,Z}, & \textit{gdy } \gamma \textit{ jest typu } F^+, \\ (-\operatorname{sgn} a)\mu_{W,Z}, & \textit{gdy } \gamma \textit{ jest typu } F^-. \end{cases}$$

Dowód. Ponieważ w rozkładzie ciągłym mediana $\gamma(W)$ jest równa $\gamma(W_{0,5})$, więc

$$F\big(Z; \gamma(W) > \text{mediana } \gamma(W)\big) = \int_B bP\big(\gamma(W) > \text{mediana } \gamma(W) \,|\, Z = b\big) d\Phi_Z(b) =$$

$$= \begin{cases} E(Z;\ W > W_{0,5}) & \text{dla } \gamma \text{ typu } F^+, \\ E(Z;\ W < W_{0,5}) & \text{dla } \gamma \text{ typu } F^-. \end{cases}$$

Podobnie, $E\big(Z; \gamma(W) < \text{mediana } \gamma(W)\big)$ i stąd

$$\mu_{\gamma(W),Z} = \begin{cases} \mu_{W,Z} & \text{dla } \gamma \text{ typu } F^+, \\ -\mu_{W,Z} & \text{dla } \gamma \text{ typu } F^-. \end{cases} \tag{2.2}$$

Z kolei, podstawiając $aZ+b$ zamiast Z, otrzymujemy natychmiast, że $\mu_{W,aZ+b} = (\operatorname{sgn} a)\mu_{W,Z}$, co w połączeniu z (2.2) kończy dowód.

WŁASNOŚĆ 5. *Jeśli* (W, Z) *ma rozkład normalny z parametrem* ϱ, *to*

$$\mu_{\gamma(W),aZ+b} = \begin{cases} (\operatorname{sgn} a)\varrho & \textit{dla } \gamma \textit{ typu } F^+, \\ (-\operatorname{sgn} a)\varrho & \textit{dla } \gamma \textit{ typu } F^-. \end{cases}$$

Dowód. Niech $W' = W - EW/\sigma_W$, $Z' = Z - EZ/\sigma_Z$. Na podstawie własności 4

$$\mu_{\gamma(W),aZ+b} = \begin{cases} (\operatorname{sgn} a)\mu_{W',Z'} & \text{dla } \gamma \text{ typu } F^+, \\ (-\operatorname{sgn} a)\mu_{W',Z'} & \text{dla } \gamma \text{ typu } F^-. \end{cases}$$

Pozostaje pokazać, że $\mu_{W',Z'} = \varrho$. Niech

$$g(u) = \frac{1}{\sqrt{2\pi}} e^{-u^2/2}, \quad G(a) = \int_{-\infty}^{a} g(u)\,du.$$

Zatem

$$\mu_{W',Z'} = \frac{-2\int_{-\infty}^{\infty} uG(-\varrho u/\sqrt{1-\varrho^2})g(u)\,du}{\int_0^{\infty} ug(u)\,du - \int_{-\infty}^{0} ug(u)\,du} = \sqrt{2\pi}\left[g(u)G\left(-\frac{\varrho u}{\sqrt{1-\varrho^2}}\right)\right]_{-\infty}^{\infty} -$$

$$- \sqrt{2\pi}\int_{-\infty}^{\infty} g(u)\frac{1}{\sqrt{2\pi}}\left(-\frac{\varrho}{\sqrt{1-\varrho^2}}\right) e^{-\varrho^2u^2/2(1-\varrho^2)}\,du =$$

$$= \varrho\int_{-\infty}^{\infty} \frac{1}{\sqrt{2\pi}\sqrt{1-\varrho^2}} e^{-u^2/2(1-\varrho^2)}\,du = \varrho.$$

Z własności 5 wynika więc, że $\mu_{W,Z}$ jest równy ϱ dla każdego rozkładu z klasy N, w którym brzegowy rozkład Z jest normalny.

WŁASNOŚĆ 6. *$\mu_{W,Z}$ jest parametrem zależności monotonicznej w klasie N.*

Dowód. Wystarczy pokazać, że w rodzinie rozkładów z klasy N o odpowiednio jednakowych rozkładach brzegowych, $\mu_{W,Z}$ jest ściśle rosnącą funkcją ϱ, przy czym $\mu_{W,Z}(\varrho) = -\mu_{W,Z}(-\varrho)$.

Mianownik $\mu_{W,Z}$ jest zawsze dodatni i jest taki sam dla wszystkich rozkładów o jednakowym rozkładzie brzegowym zmiennej Z. Licznik $\mu_{W,Z}$ można przedstawić w postaci

$$2\psi(\varrho)-EZ, \qquad \text{gdzie } \Psi(\varrho) = \int_B bP(W>W_{0,5}|Z=b)\,d\Phi_Z(b).$$

Wobec tego wystarczy pokazać, że $\Psi(\varrho)$ jest ściśle rosnąca i że $2\Psi(-\varrho)-EZ = EZ-2\Psi(\varrho)$, czyli $\Psi(-\varrho)+\Psi(\varrho) = EZ$.

Niech (X, Y) będzie taką zmienną losową o rozkładzie normalnym z parametrem ϱ, że $EX = EY = 0$, $\sigma_X = \sigma_Y = 1$, $Z = \gamma_1(X)$, $W = \gamma_2(Y)$, przy czym γ_1 i γ_2 są obie typu F^+. Zatem

$$\Psi(\varrho) = \int_{-\infty}^{\infty}\gamma_1(u)\left(1-G\left(-\frac{\varrho u}{\sqrt{1-\varrho^2}}\right)\right)g(u)\,du. \tag{2.3}$$

Ponieważ $G(-u) = 1-G(u)$, więc

$$\Psi(\varrho_2)-\Psi(\varrho_1) = \int_{-\infty}^{\infty}\gamma_1(u)\left[G\left(\frac{-\varrho_1 u}{\sqrt{1-\varrho_1^2}}\right)-G\left(\frac{-\varrho_2 u}{\sqrt{1-\varrho_2^2}}\right)\right]g(u)\,du = \tag{2.4}$$

$$= \int_0^{\infty}\big(\gamma_1(u)-\gamma_1(-u)\big)\left[G\left(\frac{-\varrho_1 u}{\sqrt{1-\varrho_1^2}}\right)-G\left(\frac{-\varrho_2 u}{\sqrt{1-\varrho_2^2}}\right)\right]g(u)\,du.$$

Dla $u>0$ funkcja $G(-\varrho u/\sqrt{1-\varrho^2})$ jest ściśle malejącą funcją ϱ, gdyż

$$\frac{dG(-\varrho u/\sqrt{1-\varrho^2})}{d\varrho} = -\frac{ug(-\varrho u/\sqrt{1-\varrho^2})}{(1-\varrho^2)^{3/2}} < 0.$$

Zatem dla $\varrho_1<\varrho_2$ wyrażenie w nawiasie kwadratowym we wzorze (2.4) jest dodatnie, a że γ_1 jest z założenia ściśle rosnąca, więc $\Psi(\varrho_2)-\Psi(\varrho_1)>0$. Z kolei z (2.3) i z własności $G(u)$ wynika od razu, że $\Psi(\varrho)+\Psi(-\varrho) = EZ$.

Gdy γ_1 i γ_2 są obie typu F^-, dowód przebiega analogicznie.

3. Współczynnik $\hat{\mu}^{(n)}_{W,Z}$. W [3] wprowadzono współczynnik $H^{(n)}_{W,Z}$; obecnie zdefiniujemy współczynnik $\hat{\mu}^{(n)}_{W,Z}$ — statystykę o podobnej strukturze, będącą „próbkowym odpowiednikiem" parametru $\mu_{W,Z}$. W dalszej części pracy przedstawione są rozważania i wyniki badań nad użytecznością obu statystyk w praktyce.

Niech $\hat{W}_{0,5}^{(n)}$ i $\hat{Z}_{0,5}^{(n)}$ oznaczają medianę z próbki A_n odpowiednio dla zmiennych losowych W i Z (a więc jeśli $Z_{s_1} \leqslant \ldots \leqslant Z_{s_n}$, to

$$\hat{Z}_{0,5}^{(n)} = \begin{cases} Z_{s_{k+1}} & \text{jeśli } n = 2k+1, \\ \dfrac{Z_{s_k} + Z_{s_{k+1}}}{2} & \text{jeśli } n = 2k). \end{cases}$$

Niech dla dowolnego zdania a

$$\delta(a) = \begin{cases} 1, & \text{gdy } a \text{ jest prawdziwe}, \\ 0, & \text{gdy } a \text{ jest fałszywe}. \end{cases}$$

Definiujemy:

$$\mu_{W,Z}^{(n)} = \frac{\sum_{j=1}^{n} Z_j \operatorname{sgn}(W_j - \hat{W}_{0,5}^{(n)}) - \sum_{j=1}^{n} Z_j \delta(W_j = \hat{W}_{0,5}^{(n)}) \dfrac{\sum_{j=1}^{n} \operatorname{sgn}(W_j - \hat{W}_{0,5}^{(n)})}{\sum_{j=1}^{n} \delta(W_j = \hat{W}_{0,5}^{(n)})}}{\sum_{j=1}^{n} Z_j \operatorname{sgn}(Z_j - \hat{Z}_{0,5}^{(n)}) - \hat{Z}_{0,5}^{(n)} \sum_{j=1}^{n} \operatorname{sgn}(Z_j - \hat{Z}_{0,5}^{(n)})}.$$

Gdy tylko jeden element próbki ma wartość cechy W równą $W_{0,5}^{(n)}$, a więc w szczególności wtedy, gdy rozkład W jest ciągły, drugi składnik w liczniku $\hat{\mu}_{W,Z}^{(n)}$ jest równy zero; w analogicznym przypadku dla zmiennej Z znika drugi wyraz w mianowniku.

Dowód, że $|\hat{\mu}_{W,Z}^{(n)}| \leqslant 1$ wynika z własności 1 współczynnika $\mu_{W,Z}$, gdyż $\hat{\mu}_{W,Z}^{(n)}$ jest parametrem $\mu_{W,Z}$ w próbce A_n traktowanej jako cała populacja. Również natychmiastowe są dowody, że

1° $\hat{\mu}_{W,Z} = 1$ (lub -1) dla $R(W,Z) \in O_M^+$ (lub O_M^-);

2° dla niezależnych W i Z rozkład $\hat{\mu}_{W,Z}^{(n)}$ jest symetryczny, a więc w szczególności $E\hat{\mu}_{W,Z}^{(n)} = 0$;

3° jeśli (W,Z) ma rozkład ciągły, to

$$\hat{\mu}_{\gamma(W),aZ+b}^{(n)} = \begin{cases} (\operatorname{sgn} a)\hat{\mu}_{W,Z}^{(n)}, & \text{gdy } \gamma \text{ jest typu } F^+, \\ (-\operatorname{sgn} a)\hat{\mu}_{W,Z}^{(n)}, & \text{gdy } \gamma \text{ jest typu } F^-. \end{cases}$$

Asymptotyczne własności $\hat{\mu}$ opisuje następujące twierdzenie:

TWIERDZENIE 1. *Jeśli (W,Z) ma rozkład ciągły, przy czym $EZ^4 < \infty$ i istnieje $\mu_{W,Z}$, to* $\operatorname{plim}_{n\to\infty} \hat{\mu}_{W,Z}^{(n)} = \mu_{W,Z}$.

Dowód. Niech (W, Z) ma rozkład ciągły i niech Z' oznacza zmienną losową o gęstości

$$\frac{1}{P(W > W_{0,5})} \int_{a > W_{0,5}} \varphi(a, b)\, da.$$

Stąd $EZ' = 2E(Z;\, W > W_{0,5})$. Niech

$$S_n = \sum_{j=1}^{n} \delta(W_j > W_{0,5});$$

jest to zatem zmienna losowa o rozkładzie dwumianowym ze średnią $n/2$ i wariancją $n/4$.

Pokażemy, że ciąg $\{L_n\}$, gdzie

$$L_n = \begin{cases} \dfrac{\sum\limits_{j=1}^{n} Z_j\, \delta(W_j > \hat{W}_{0,5}^{(n)})}{S_n} & \text{dla } s_n > 0, \\ 0 & \text{dla } s_n = 0, \end{cases}$$

jest stochastycznie zbieżny do EZ'.

Jeżeli $S_n > 0$, to L_n można zapisać w postaci $\sum\limits_{j=1}^{n} Z'_j / S_n$, gdzie $\{Z'_1, \ldots, Z'_{S_n}\}$ jest ciągiem tych zmiennych losowych $Z_1, \ldots, Z_n$, dla których $W_j > W_{0,5}$. Jest to ciąg niezależnych zmiennych losowych o jednakowym rozkładzie, takim jak rozkład Z'.

Oznaczmy przez $A_n(\varepsilon)$ zdarzenie polegające na tym, że $|L_n - EZ'| > \varepsilon$. Dla $1 \leqslant k_0 \leqslant n$ możemy napisać

$$\begin{aligned} (3.1) \qquad P\big(A_n(\varepsilon)\big) &= P\big(A_n(\varepsilon), S_n < k_0\big) + \sum_{k=k_0}^{n} P\big(A_n(\varepsilon), S_n = k\big) \leqslant \\ &\leqslant P(S_n < k_0) + \sum_{k=k_0}^{n} P\big(A_n(\varepsilon) \,|\, S_n = k\big) \cdot P(S_n = k). \end{aligned}$$

Ponieważ dla $k \geqslant 1$

$$P\big(A_n(\varepsilon) \,|\, S_n = k\big) = P\left(\left| \frac{\sum\limits_{j=1}^{k} (Z'_j - EZ')}{k} \right| > \varepsilon \right),$$

więc dla dowolnych $\eta > 0$ i $\varepsilon > 0$ istnieje takie k_0, że dla k i n spełniających nierówność $k_0 \leqslant k \leqslant n$ mamy $P\big(A_n(\varepsilon) \,|\, S_n = k\big) < \eta/2$. Z drugiej strony, dla każdego $\eta > 0$ i k_0 istnieje takie $n_0 > k_0$, że $P(S_n < k_0) < \eta/2$ dla dowolnego $n > n_0$. Zatem, na podstawie (3.1), do η i ε można dobrać takie k_0 i n_0, że dla $n > n_0$ mamy $\operatorname{plim}\limits_{n \to \infty} L_n = EZ'$.

Niech

$$T_n(a) = \frac{1}{n}\sum_{j=1}^{n} Z_j\,\delta(W_j > a).$$

Zatem $T_n(W_{0,5}) = L_n(S_n/n)$, a ponieważ $\operatorname*{plim}_{n\to\infty} S_n/n = \frac{1}{2}$, więc na podstawie twierdzenia o stochastycznej zbieżności iloczynu dwóch ciągów otrzymujemy $\operatorname*{plim}_{n\to\infty} T_n(W_{0,5}) = \frac{1}{2}EZ' = E(Z;\, W > W_{0,5})$. Pokażemy teraz, że również $\operatorname*{plim}_{n\to\infty} T_n(\hat{W}^{(n)}_{0,5}) = E(Z;\, W > W_{0,5})$, czyli że $U_n = T_n(\hat{W}^{(n)}_{0,5}) - T_n(W_{0,5})$ dąży stochastycznie do zera.

Mamy

$$\begin{aligned}
EU_n^2 &= \frac{1}{n^2}E\Big\{\sum_{j=1}^{n} Z_j[\delta(W_j > \hat{W}^{(n)}_{0,5}) - \delta(W_j < W_{0,5})]\Big\}^2 \leqslant \\
&\leqslant \frac{1}{n^2}\sum_{j,k=1}^{n} E\{|Z_jZ_k|\cdot|\delta(W_j > \hat{W}^{(n)}_{0,5}) - \delta(W_j > W_{0,5})|\} = \\
&= \frac{1}{n^2}\sum_{j,k=1}^{n} E\{|Z_jZ_k|\,\delta[(\hat{W}^{(n)}_{0,5} \leqslant W_j < W_{0,5}) \vee (W_{0,5} < W_j \leqslant \hat{W}^{(n)}_{0,5})]\} = \\
&= \frac{1}{n^2}\sum_{j,k=1}^{n} E\{|Z_jZ_k|\,\delta(|W_j - W_{0,5}| < |W_{0,5} - \hat{W}^{(n)}_{0,5}|)\}.
\end{aligned}$$

Niech $M = EZ^4$, $D_j = |W_j - W_{0,5}|$ i $F_n = |\hat{W}^{(n)}_{0,5} - W_{0,5}|$. Na podstawie nierówności Schwarza mamy

$$EU_n^2 \leqslant \frac{\sqrt{M}}{n}\sum_{j=1}^{n}\sqrt{P(D_j < F_n)}.$$

Gdy (W, Z) ma rozkład ciągły, ciąg $\{\hat{W}^{(n)}_{0,5}\}$ jest stochastycznie zbieżny do $W_{0,5}$, a zatem dla każdego $\eta > 0$ i $\varepsilon > 0$ istnieje takie n_0, że $P(F_n > \varepsilon) < \eta^2/2M$ dla $n > n_0$. Z ciągłości (W, Z) wynika również, że dla każdego $\eta > 0$ istnieje takie $\varepsilon > 0$, że $P(D_j < \varepsilon) < \eta^2/2M$ dla każdego j. Zatem dla każdego $\eta > 0$ istnieją takie $\varepsilon > 0$ i n_0, że dla $n > n_0$ i dla każdego j

$$\begin{aligned}
P(D_j < F_n) &= P(D_j < F_n,\, D_j < \varepsilon) + P(D_j < F_n,\, D_j > \varepsilon,\, F_n < \varepsilon) + \\
&+ P(D_j < F_n,\, D_j > \varepsilon,\, F_n > \varepsilon) \leqslant P(D_j < \varepsilon) + P(F_n > \varepsilon) < \frac{\eta^2}{M},
\end{aligned}$$

czyli

$$EU_n^2 \leqslant \frac{\sqrt{M}}{n} \sum_{j=1}^{n} \sqrt{\frac{\eta^2}{M}} = \eta \qquad \text{dla } n > n_0.$$

Stąd wynika, że $\operatorname*{plim}_{n\to\infty} U_n = 0$.

Ponieważ dla rozkładów ciągłych

$$\hat{\mu}_{W,Z} = \frac{\frac{1}{n}\sum_{j=1}^{n} Z_j\,\delta(W_j > \hat{W}_{0,5}^{(n)}) - \frac{1}{n}\sum_{j=1}^{n} Z_j\,\delta(W_j < \hat{W}_{0,5}^{(n)})}{\frac{1}{n}\sum_{j=1}^{n} Z_j\,\delta(Z_j > \hat{Z}_{0,5}^{(n)}) - \frac{1}{n}\sum_{j=1}^{n} Z_j\,\delta(Z_j < \hat{Z}_{0,5}^{(n)})}, \tag{3.2}$$

więc z dotychczasowych rozważań wynika, że pierwszy składnik licznika $\hat{\mu}$ dąży stochastycznie do $E(Z; W > W_{0,5})$, czyli do pierwszego składnika $\mu_{W,Z}$. Podobne dowody można przeprowadzić dla drugiego wyrazu licznika i obu wyrazów mianownika w (3.2). Stosując następnie twierdzenia o stochastycznej zbieżności ciągu sum i ilorazów otrzymujemy tezę twierdzenia.

4. Zakres zastosowań $\hat{\mu}^{(n)}$ i $H^{(n)}$. Spośród statystyk stosowanych w badaniach zależności dwu zmiennych najczęściej używanym jest współczynnik korelacji liniowej z próbki $\hat{\varrho}_{W,Z}^{(n)}$. Ponieważ ciąg $\{\hat{\varrho}_{W,Z}^{(n)}\}$ jest stochastycznie zbieżny do $\varrho_{W,Z}$, więc dla dużych próbek stosują się do $\hat{\varrho}_{W,Z}^{(n)}$ uwagi krytyczne podane w punkcie 2 w związku z $\varrho_{W,Z}$.

Przy stosowaniu $\hat{\varrho}_{W,Z}^{(n)}$ warunkiem koniecznym jest to, żeby obie badane cechy były mierzone na skali co najmniej przedziałowej. Wobec tego $\hat{\varrho}_{W,Z}^{(n)}$ ma węższy zakres zastosowań od $\hat{\mu}_{W,Z}^{(n)}$ i $H_{W,Z}^{(n)}$, które wymagają pomiaru jednej tylko cechy na skali co najmniej przedziałowej, a drugiej cechy na skali co najmniej porządkowej; stąd w niektórych badaniach $\hat{\mu}$ i H są dopuszczalne, podczas gdy z góry trzeba wykluczyć możność stosowania $\hat{\varrho}$. Z drugiej strony, statystyk tych nie można używać, jeśli obie cechy są mierzone na skali co najwyżej porządkowej, a więc $\hat{\mu}$ i H mają z kolei węższy zakres zastosowań niż różne współczynniki korelacji rangowej.

W praktyce bywa również tak, że pewna cecha może wprawdzie być mierzona na skali przedziałowej, ale koszt pomiaru (określany nakładem pieniędzy, czasu lub trudu eksperymentatora) jest bardzo wysoki. Cechę taką nazywa się często „trudną" w odróżnieniu od cech „łatwych", których pomiary są tanie. Występuje często potrzeba znalezienia zastępczych cech łatwych dla interesujących nas cech trudnych. Badanie zależności między taką parą cech jest badaniem wstępnym np. wtedy, gdy opieramy kontrolę jakości lub badanie niezawodności na cesze zastępczej Z zamiast na „ważnej" cesze W ([1], [2]). Statystyki $\hat{\mu}$ i H są przystosowane do takich właśnie sytuacji, w których dokładny pomiar jednej z badanych cech jest bardzo kosztowny (np. badania niszczące, trudne oznaczenia chemiczne itd.), natomiast bez trudu można ustalić, która z dwu wartości tej cechy jest „większa". W takich przypadkach oczekiwany koszt wyznaczenia wartości $\hat{\varrho}_{W,Z}^{(n_1)}$ (lub innej statystyki wymagającej dokładnych pomiarów obu cech) może być znaczny w porównaniu z kosztem wyznaczenia wartości $\hat{\mu}_{W,Z}^{(n_2)}$ lub $H_{W,Z}^{(n_2)}$ nawet wtedy, gdy $n_1 \leqslant n_2$. Oczywiście, dokładne oceny wymagają bliższego sprecyzowania założeń.

Jako przykład rozważymy sytuację, gdy cechą W jest trwałość. Jest to cecha trudna do badania ze względu na czas trwania doświadczeń. Przy stosowaniu $\hat{\varrho}_{W,Z}^{(n_1)}$ w każdym doświadczeniu musi upłynąć czas $\max(W_1, \ldots, W_{n_1})$, a zatem oczekiwany czas trwania doświadczenia dla $n_1 > 1$ i przy danym rozkładzie (W, Z) jest $E(\max(W_1, \ldots, W_{n_1})) > > EW$. Natomiast przy stosowaniu $H_{W,Z}^{(n_2)}$ badanie można przerwać z chwilą, gdy zakończy życie element medialny w próbce, to jest ten element, który ma wartość cechy Z równą medianie $Z_{s_{k+1}}$ wartości $Z_1, \ldots, Z_{2k+1}$, gdzie $n_2 = 2k+1$; aby bowiem obliczyć wartość H, trzeba wiedzieć tylko, które elementy w próbce mają trwałość większą, a które mniejszą od medialnego. Jeśli przez $W_{s_{k+1}}$ oznaczymy trwałość elementu medialnego, to przy posługiwaniu się $H^{(n_2)}$ oczekiwany czas trwania doświadczenia jest równy $EW_{s_{k+1}}$. Podamy obecnie warunek dostateczny na to, żeby $EW_{s_{k+1}} < EW$ dla $k > 0$.

Oznaczmy przez (X, Y) zmienną losową o rozkładzie normalnym, w którym $EX = EY = 0$, $\sigma_X = \sigma_Y = 1$. Dla danego rozkładu (W, Z) z klasy N niech γ_1 i γ_2 będą takimi funkcjami, że $Z = \gamma_1(X)$, $W = \gamma_2(Y)$. Będziemy mówić, że rozkład (W, Z) spełnia *warunek A*, jeśli należy do klasy N i jeśli $E(W|\gamma_1(x)) + E(W|\gamma_1(-x))$ jest funkcją rosnącą dla $x > 0$, a więc jeśli $E(W|z)$ jest dostatecznie szybko rosnącą lub dostatecznie szybko malejącą funkcją z. Zauważmy, że warunek ten spełnia każdy rozkład z klasy N o $|\varrho| > 0$, w którym zmienna losowa W ma rozkład logarytmo-normalny, a więc rozkład, którym często daje się opisać rozkład trwałości. Wtedy bowiem łatwo policzyć, że

$$E(W|\gamma_1(x)) = (EW)\exp\left(\varrho Cx - \varrho^2 \frac{C^2}{2}\right), \quad \text{gdzie } C = \sqrt{\ln\left(\frac{\sigma_W^2}{(EW)^2} + 1\right)},$$

a więc

$$E(W|\gamma_1(x)) + E(W|\gamma_1(-x)) = (EW)\exp\left(-\varrho^2 \frac{C^2}{2}\right)(e^{\varrho Cx} + e^{-\varrho Cx});$$

funkcja ta dla $|\varrho| > 0$ ma dodatnią pochodną dla $x > 0$, gdyż $EW > 0$.

Przypuszczamy, że warunek ten zachodzi także dla innych rozkładów z klasy N o rozkładzie brzegowym cechy W prawostronnie skośnym.

Twierdzenie 2. *Jeśli $R(W, Z)$ spełnia warunek A, to $EW_{s_{k+1}} < EW$ dla $k > 0$.*

Dowód. Oznaczmy przez $\varphi_Z^{(k+1)}$ gęstość rozkładu mediany $\hat{Z}_{0,5}^{(2k+1)}$:

$$\varphi_Z^{(k+1)}(b) = \frac{(2k+1)!}{(k!)^2}\{[\Phi_Z(b)][1-\Phi_Z(b)]\}^k\varphi_Z(b).$$

Mamy

$$EW_{s_{k+1}} = \int_B E(W|b)\varphi_Z^{(k+1)}(b)\,db;$$

podstawiając $b = \gamma_1(x)$ i wprowadzając poprzednio zdefiniowane funkcje g i G, mamy

$$EW_{s_{k+1}} = \frac{(2k+1)!}{(k!)^2}\int_{-\infty}^{\infty} E\big(W|\gamma_1(x)\big)\big[G(x)\big(1-G(x)\big)\big]^k g(x)\,dx =$$

$$= \frac{(2k+1)!}{(k!)^2}\int_0^{\infty}\big[E\big(W|\gamma_1(x)\big)+E\big(W|\gamma_1(-x)\big)\big]\big[G(x)\big(1-G(x)\big)\big]^k g(x)\,dx.$$

Podobnie

$$EW = \int_0^{\infty}\big[E\big(W|\gamma_1(x)\big)+E\big(W|\gamma_1(-x)\big)\big]g(x)\,dx.$$

Wiadomo, że

$$\frac{(2k+1)!}{(k!)^2}\int_0^{\infty}\big[G(x)\big(1-G(x)\big)\big]^k g(x)\,dx = \int_0^{\infty} g(x)\,dx = \frac{1}{2}$$

oraz że dla każdego $a > 0$ i $k > 0$

$$\frac{(2k+1)!}{(k!)^2}\int_0^{a}\big[G(x)\big(1-G(x)\big)\big]^k g(x)\,dx > \int_0^{a} g(x)\,dx.$$

Jeśli zatem $E\big(W|\gamma_1(x)\big)+E\big(W|\gamma_1(-x)\big)$ jest rosnącą funkcją x dla $x > 0$, to $EW_{s_{k+1}} < EW$, a więc twierdzenie jest udowodnione.

Zatem w tym przykładzie oczekiwany czas trwania doświadczenia jest zawsze mniejszy przy stosowaniu $H^{(n_2)}$ niż $\hat{\varrho}^{(n_1)}$, bez względu na n_1

i n_2. Jeśli przyjmiemy, że koszt badania jest wprost proporcjonalny do czasu trwania doświadczenia, to oczekiwane koszty badania przy H są zawsze niższe od oczekiwanych kosztów przy $\hat{\varrho}$.

Żeby jednak zalecić stosowanie H zamiast $\hat{\varrho}$, należałoby sformułować wymagania dotyczące oczekiwanych konsekwencji statystycznych posługiwania się obiema statystykami i stwierdzić, czy przy danym $R(W, Z)$ istnieją, i jakie są, najmniejsze liczności próbek n_1 i n_2, przy których te wymagania są spełnione. Na przykład, jeśli naszym celem jest testowanie dwu prostych alternatywnych hipotez $\varrho = \varrho_i$ $(i = 1, 2)$, to wymagania mają zazwyczaj postać warunków ustalających maksymalne prawdopodobieństwa błędu I i II rodzaju, α i β. Wiadomo, że można zawsze znaleźć taką dostatecznie dużą liczność próbki n_1, że przy $\hat{\varrho}^{(n_1)}_{W,Z}$ warunki te będą spełnione. W przypadku $H^{(n_2)}_{W,Z}$ tak nie jest, gdyż można udowodnić, że $H^{(n_2)}_{W,Z}$ ma asymptotyczny rozkład o dodatniej wariancji, a zatem istnienie liczności próbki n_2 gwarantującej spełnienie warunków zależy od rozkładu (W, Z) przy $\varrho = \varrho_1$ i ϱ_2 oraz od α i β. Jeśli n_2 istnieje, statystykę $H^{(n_2)}_{W,Z}$ uważamy za lepszą od $\hat{\varrho}^{(n_1)}_{W,Z}$.

Wybór jednej z dwu rozpatrywanych statystyk był więc w tym przykładzie uzależniony od porównania oczekiwanych kosztów badania każdą z rozpatrywanych statystyk z licznościami próbki n_1 i n_2, przy których spełnione są warunki nałożone na moc testu. Sytuacja była tu szczególnie prosta, gdyż wynik porównania owych oczekiwanych kosztów nie zależał nie tylko od n_1 i n_2, ale również od ϱ_1 i ϱ_2 i w ogóle od rozkładu (W, Z). Gdyby konstruować kryterium wyboru najlepszej z kilku rozpatrywanych statystyk, uogólniając kryterium podane w tym przykładzie, trzeba by uwzględnić fakt, że na ogół oczekiwane koszty badania dla danej statystyki i danej liczności próbki nie są jednoznacznie ustalone, zależą bowiem od sposobu i warunków przeprowadzenia badań.

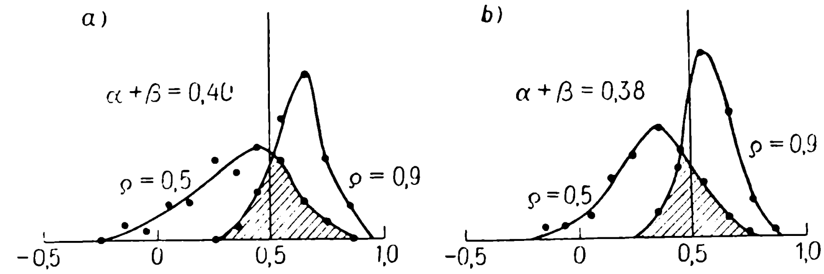
Jednakże rozważanie kryterium wyboru statystyki nie jest celem tej pracy, a omówiony wyżej przykład miał tylko zilustrować różnice występujące w oczekiwanych kosztach badania związanych z $\hat{\varrho}$ i H w problemie z jedną cechą trudną.

Można by skonstruować wiele podobnych przykładów dotyczących H lub $\hat{\mu}$. Warto tu dodać, że dla $\hat{\mu}^{(n)}_{W,Z}$ na mocy twierdzenia 1 można by zawsze dobrać taką liczność próbki, przy której byłyby spełnione warunki nałożone na moc testu przy weryfikacji hipotez $\varrho = \varrho_i$ $(i = 1, 2)$ dla $R(W, Z) \in N$.

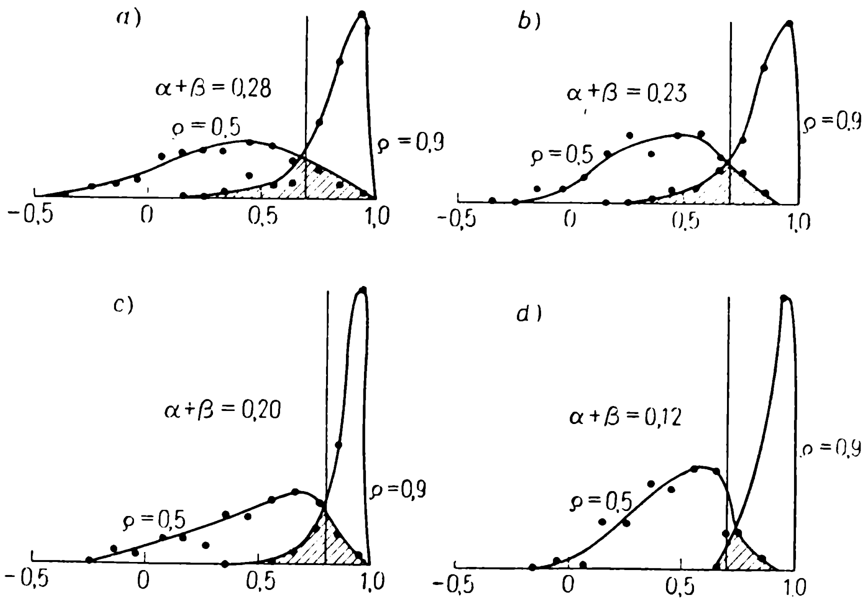
Przy dokonywaniu wyboru statystyki może również okazać się istotna szczególnie prosta w obliczeniach postać zarówno $\hat{\mu}$, jak i H.

5. Ilustracja numeryczna. Wnioskowanie statystyczne na podstawie $\hat{\mu}^{(n)}_{W,Z}$ lub $H^{(n)}_{W,Z}$ jest możliwe tylko wtedy, gdy znane są rozkłady tych statystyk przy danym $R(W, Z)$ i danej liczności próbki. Teoretyczne wyznaczenie tych rozkładów, gdy (W, Z) ma rozkład normalny, byłoby wprawdzie możliwe, jednakże metodami Monte Carlo przy użyciu maszyny cyfrowej można bez trudu oszacować rozkład każdej ze statystyk dla dowolnego n i przy odpowiednich założeniach o $R(W, Z)$.

Ponieważ $\hat{\mu}^{(n)}_{W,Z}$ i $H^{(n)}_{W,Z}$ pozostają niezmienione przy ściśle rosnących transformacjach zmiennej losowej W, więc jeśli założymy, że $R(W, Z) \in N$, to do wyznaczenia rozkładu każdej ze statystyk potrzebna jest jedynie



Rys. 1. Empiryczne rozkłady $\hat{\varrho}^{(n)}_{W,Z}$, gdy $(\sqrt[7]{W}, Z)$ ma rozkład normalny z parametrami $\{0, m_Z, 1, \delta_Z, \varrho\}$.
a) Empiryczne rozkłady $\hat{\varrho}^{(15)}$. b) Empiryczne rozkłady $\hat{\varrho}^{(25)}$.



Rys. 2. Empiryczne rozkłady $H^{(n)}_{W,Z}$ i $\hat{\mu}^{(n)}_{W,Z}$, gdy (W, Z) ma rozkład z klasy N z parametrem ϱ, a brzegowy rozkład Z jest normalny.
a) Empiryczne rozkłady $H^{(15)}$. b) Empiryczne rozkłady $H^{(25)}$. c) Empiryczne rozkłady $\hat{\mu}^{(15)}$. d) Empiryczne rozkłady $\hat{\mu}^{(25)}$.

znajomość rozkładu brzegowego zmiennej Z oraz parametru ϱ. Ma to znaczenie zwłaszcza w badaniach z „trudną” cechą W i „łatwą” cechą Z, gdyż wtedy jest niekiedy możliwe zebranie dodatkowych obserwacji zmiennej Z i oszacowanie jej rozkładu.

Na rys. 2 podane są empiryczne rozkłady $H^{(n)}_{W,Z}$ i $\hat{\mu}^{(n)}_{W,Z}$ dla $n = 15$ i 25, gdy (W, Z) ma rozkład z klasy N z parametrem $\varrho = 0{,}5$ i 0,9, a brzegowy rozkład Z jest normalny. Na rys. 1 umieszczone są analogiczne rozkłady dla $\hat{\varrho}^{(n)}_{W,Z}$, gdy $(\sqrt[7]{W}, Z)$ ma rozkład normalny $N(0, m_Z, 1, \sigma_Z, \varrho)$. Na to bowiem, aby wyznaczyć rozkład $\hat{\varrho}$ dla rozkładów z N, gdy W nie jest zmienną normalną, rozkład brzegowy W musi być całkowicie wyspecyfikowany. Założenia o $R(W, Z)$ przy $\hat{\varrho}$ są oczywiście szczególnym przypadkiem tych założeń przyjętych przy $\hat{\mu}$ i H.

Mówiąc nieściśle, przy wnioskowaniu statystycznym jest pożądane, aby rozkłady rozpatrywanej statystyki dla $\varrho = 0{,}5$ i $\varrho = 0{,}9$ były możliwie różne. Jako wskaźnik owych różnic obliczono dla każdej ze statystyk kres dolny sumy oszacowań prawdopodobieństw błędu I i II rodzaju, $\alpha+\beta$, w teście weryfikującym hipotezę $\varrho = 0{,}9$ wobec hipotezy $\varrho = 0{,}5$. Otrzymane wyniki przedstawia tablica 1.

TABLICA 1. Wartości $\alpha+\beta$.

n	$\hat{\mu}^{(n)}_{W,Z}$	$H^{(n)}_{W,Z}$	$\hat{\varrho}^{(n)}_{W,Z}$
15	0,20	0,28	0,40
25	0,12	0,23	0,38

A więc, gdyby naszym zadaniem było testowanie hipotezy $\varrho = 0{,}9$ wobec $\varrho = 0{,}5$ i gdyby rozkład $(\sqrt[7]{W}, Z)$ był normalny $N(0, m_Z, 1, \sigma_Z, \varrho)$, to dla $n = 25$ można by zaproponować test, w którym α i β byłyby w przybliżeniu równe 0,06 przy statystyce $\hat{\mu}^{(25)}_{W,Z}$, 0,13 przy $H^{(25)}_{W,Z}$ i aż 0,17 przy $\hat{\varrho}^{(25)}_{W,Z}$. Gdyby natomiast rozkład brzegowy W był normalny, to testy oparte na $\hat{\mu}$ i H nie uległyby zmianie, podczas gdy stosując $\hat{\varrho}$ można by zbudować test z α i β równymi około 0,05 dla $n = 15$ i około 0,015 dla $n = 25$ (przy obliczeniach korzystano z Fisherowskiego przekształcenia $U = \log(1+\hat{\varrho})/(1-\hat{\varrho})$. Z tego widać, jak wielką rolę we wnioskowaniu statystycznym na podstawie $\hat{\varrho}$ odgrywa postać brzegowego rozkładu W.

Przypuśćmy, że $R(W, Z)$ jest normalny i że testujemy hipotezę $\varrho = 0{,}9$ wobec $\varrho = 0{,}5$, przy czym żądamy, żeby $\alpha = \beta = 0{,}10$. Wymagania te są spełnione np. w teście opartym na $\hat{\mu}^{(15)}_{W,Z}$ oraz w teście opartym na $\hat{\varrho}^{(11)}_{W,Z}$. Obie te statystyki traktujemy więc jako równoważne i dalszy wybór między nimi uzależniamy od oczekiwanych kosztów badania. Zakładamy, że koszt badania zależy tylko od oceny wartości cechy W. Zatem w teście związanym z $\hat{\varrho}^{(11)}$ jest on równy $11a$, gdzie $a > 0$ jest kosztem dokładnego wyznaczenia jednej wartości cechy W; natomiast przy stosowaniu $\hat{\mu}^{(15)}$ koszt jednego porównania pary wartości (W_i, W_j) wynosi b $(0 < b < a)$. Oczekiwane koszty badania przy $\hat{\mu}^{(15)}$ są z założenia proporcjonalne do oczekiwanej liczby porównań i zależą od przyjętego algorytmu podziału próbki na elementy z wartościami cechy W większymi i mniejszymi od mediany, przy czym można zaproponować algorytmy zależne od ϱ (jeśli np. najpierw uporządkować próbkę ze względu na wartości skorelowanej cechy Z); w każdym razie jednak koszty te nie przekraczają oczywiście $\frac{1}{2}n(n-1)b = 105b$, a na ogół są znacznie mniejsze. A zatem nawet w korzystnym dla $\hat{\varrho}$ przypadku, gdy (W, Z) jest normalna, jesteśmy w podanym przykładzie skłonni wybrać $\hat{\mu}^{(15)}$ zamiast $\hat{\varrho}^{(11)}$, jeśli $11a > 105b$. W praktyce nieraz bywa tak, że $a \gg b$, np. wtedy, gdy dokładne badanie w przeciwieństwie do porównywania jest niszczące.

Jednakże wyniki tabelki 1 wskazują na to, że celowe jest stosowanie statystyk $\hat{\mu}$ i H szczególnie wtedy, gdy rozkład „trudnej" cechy W wyraźnie odbiega od normalnego.

6. Uwaga. Podjęto próby wprowadzenia statystyk analogicznych $\hat{\mu}$ i H do wnioskowania statystycznego o monotonicznej zależności dwu procesów stochastycznych z dyskretnym czasem; problem badania „serii zastępczych" wydaje się niemniej ważny w praktyce od problemu „cech zastępczych".

Prace cytowane

[1] J. Oderfeld, *Statystyczne badanie na cechy skorelowane*, Zastosow. Matem. 4 (1959), str. 255-264.

[2] E. Pleszczyńska, *Odsiewanie w badaniu statystycznym na cechy skorelowane*, Zastosow. Matem. 5 (1960), str. 47-60.

[3] — *Współczynnik zależności monotonicznej z próbki (I)*, Zastosow. Matem. 9 (1968), str. 203-217.

INSTYTUT MATEMATYCZNY POLSKIEJ AKADEMII NAUK
WARSZAWA

Praca wpłynęła 24. 12. 1968

Э. ПЛЕЩИНЬСКА (Варшава)

ВЫБОРОЧНЫЙ КОЭФФИЦИЕНТ МОНОТОННОЙ ЗАВИСИМОСТИ (II)

РЕЗЮМЕ

В § 4 работы обсуждены пределы применений двух статистик употребляемых в исследовании монотонной зависимости между двумя признаками: статистики $H^{(n)}$, введенной в [3], и статистики $\hat{\mu}^{(n)}$, определенной в этой работе. Подчеркнута пригодность этих статистик в том случае, когда точное измерение одного из признаков дорого и очень затруднительно или вообще невозможно, и когда

можно измерять только значения некоторой неизвестной, строго монотонной функции этого признака; это случается на практике прежде всего в исследовании так называемых *заменительных признаков*. В § 5 приведены численные примеры иллюстрирующие статистические выводы, основанные на этих статистиках.

Теорема 1 из § 3 касается стохастической сходимости последовательности $\{\hat{\mu}^{(n)}\}$ к некоторому параметру μ, который, как показано в § 2, хорошо характеризует — в смысле определений из [3] — монотонную зависимость случайных переменных.

E. PLESZCZYŃSKA (Warszawa)

A SAMPLE COEFFICIENT OF MONOTONIC DEPENDENCE (II)

SUMMARY

§ 4 of the paper discusses the domain of application of two statistics used in the investigation of the monotonic dependence between two characteristics, the statistics $H^{(n)}$, introduced in [3], and the statistics $\hat{\mu}^{(n)}$, defined in this paper. Underlined is the usefulness of H and $\hat{\mu}$ in the case where the measurement of one of the characteristics is costly and difficult, or even impossible, so that one can measure the values of some unknown, strictly monotonic function of that characteristics only; first of all, this takes place in practice in an investigation of the so called ***substitutional characteristics***. In § 5 numerical examples illustrating statistical inference based on both these statistics are given.

Theorem 1, proved in § 3, asserts the stochastical convergence of the sequence $\{\hat{\mu}^{(n)}\}$ to some parameter μ which, as has been shown in § 2, well characterizes — in the sense of the definitions introduced in [3] — the monotonic dependence of random variables.