



URSZULA LUBOŃSKA, WOJCIECH NIEMIRO

Warszawa

O wnioskowaniu dotyczącym binarnej cechy ukrytej

(Praca wpłynęła do Redakcji 25.11.1988)

1. Wstęp. Nasze rozważania będą dotyczyły wnioskowania o cesze ukrytej w modelu statystycznym zdefiniowanym w ogólnych zarysach w pracy [2]. Schemat problemu jest następujący. Obserwujemy zmienne losowe $Z_1, \dots, Z_n$. Uważamy, że stanowią one subiektywne oceny pewnej ukrytej cechy U , niedostępnej obserwacji, podane przez n osób, które nazwiemy *ekspertami*. Model statystyczny opisujący taką sytuację jest sprecyzowany przez podanie rodziny $\mathcal{P}$ rozkładów prawdopodobieństwa wektora losowego

$$(U, Z) = (\underset{\substack{\uparrow \\ \text{cecha ukryta}}}{U}, \underbrace{Z_1, Z_2, \dots, Z_n}_{\text{opinie ekspertów}})$$

(na przestrzeni $\mathbb{R}^{n+1}$). Podobnie, jak autorzy pracy [2], będziemy rozważali nieparametryczne rodziny $\mathcal{P}$, składające się ze wszystkich rozkładów spełniających pewne warunki, dotyczące rodzaju zależności pomiędzy zmiennymi U oraz $Z_1, \dots, Z_n$. Sformułujemy te warunki w p. 2.

Dla $P \in \mathcal{P}$ oznaczymy przez $P_{|Z}$ rozkład brzegowy (n -wymiarowy) wektora opinii ekspertów $Z = (Z_1, \dots, Z_n)$. Autorzy pracy [2] skupili uwagę na badaniu własności rozkładu $P_{|Z}$ wynikających z przyjętych założeń o rodzinie $\mathcal{P}$. Nas interesować będzie zadanie rekonstrukcji rozkładu P na podstawie znajomości rozkładu $P_{|Z}$. Założymy, że dysponujemy na tyle dużą ilością obserwacji wektora losowego Z , że możemy przyjąć rozkład $Q = P_{|Z}$ tego wektora za znany. Będziemy się starali wyznaczyć rodzinę

$$(1) \quad \{P \in \mathcal{P} : P_{|Z} = Q\},$$

czyli podrodzinę łącznych rozkładów wektora (U, Z) o zadanym rozkładzie brzegowym wektora Z . Rozwiążemy to zadanie w p. 3 w przypadku zmiennej

ukrytej U przyjmującej dwie wartości.

Rozpatrzmy następnie zadanie wnioskowania o cesze ukrytej przy pełnej znajomości rozkładu cech obserwowalnych (por. [1]). Wykorzystując znajomość rodziny (1), chcemy wnioskować o U na podstawie obserwacji $Z_1, \dots, Z_n$ (opinii łącznie wszystkich ekspertów, części spośród nich lub pojedynczego eksperta Z_j). W ramach statystycznej teorii decyzji można interesujące nas zadanie wnioskowania sformułować jako poszukiwanie rozwiązania minimaxowego, określonego w sposób następujący. Niech Δ będzie zbiorem decyzji, $L(u, d)$ stratą związaną z podjęciem decyzji $d \in \Delta$ przy założeniu, że $U = u \in \mathbb{R}$. Minimaksową regułą wnioskowania nazwiemy taką funkcję $\delta: \mathbb{R}^k \rightarrow \Delta$ ($1 \leq k \leq n$; argumentami δ są pewne współrzędne $Z_{j_1}, \dots, Z_{j_k}$ wektora $Z = (Z_1, \dots, Z_n)$), że

$$(2) \quad \sup_{\{P \in \mathcal{P}: P|_Z = Q\}} \mathbf{E}_P L(U, \delta(Z_{j_1}, \dots, Z_{j_k})) \\ = \inf_{\delta'} \sup_{\{P \in \mathcal{P}: P|_Z = Q\}} \mathbf{E}_P L(U, \delta'(Z_{j_1}, \dots, Z_{j_k})),$$

gdzie Q jest znanym rozkładem brzegowym wektora Z .

Oczywiście, założenie o znajomości rozkładu brzegowego cech obserwowalnych jest dużym uproszczeniem. W praktyce wyznaczenie podrodziny (1) lub minimaxowego rozwiązania (2) musi być zawsze poprzedzone estymacją rozkładu Q w rodzinie $\{P|_Z: P \in \mathcal{P}\}$. Takie zagadnienia estymacji nie będą rozważane w tej pracy.

2. Ogólne założenia o modelu. W paragrafie tym przytoczymy pewne założenia, dotyczące łącznego rozkładu wektora (U, Z) , sformułowane w pracy [2]. W dalszym ciągu pracy rozważać będziemy rodziny składające się z rozkładów spełniających wszystkie lub niektóre z poniższych założeń.

Będziemy stale zakładać, że opinie ekspertów na temat tego samego obiektu są niezależne:

ZAŁOŻENIE (I). *Zmienne losowe $Z_1, Z_2, \dots, Z_n$ są warunkowo niezależne pod warunkiem $U = u$, dla dowolnego $u \in \mathbb{R}$.*

Niekiedy uzasadnione jest przyjęcie następującego założenia o jednorodności ekspertów:

ZAŁOŻENIE (II). *Dla wszystkich $j = 1, \dots, n$ dwuwymiarowe rozkłady par (U, Z_j) są takie same.*

Rozważać będziemy ponadto pewne założenia dotyczące struktury zależności pomiędzy zmiennymi losowymi U i Z_j . Nieformalny sens każdego z dwu poniższych założeń jest taki, że większym wartościom cechy ukrytej U odpowiadają wyższe oceny Z_j ekspertów.

ZAŁOŻENIE (III). *Zmienne losowe U i Z_j są dodatnio regresyjnie zależne (zależne w sensie $RD+$), dla $j = 1, \dots, n$.*

Z definicji (por. [3]), zależność typu $RD+$ pomiędzy U i Z_j oznacza, że dla $u' \leq u$ zachodzi nierówność $P(Z_j < z | U = u') \geq P(Z_j < z | U = u)$ dla każdego $z \in \mathbf{R}$.

W części dalszych rozważań korzystać będziemy z następującego, mocniejszego założenia:

ZAŁOŻENIE (III'). *Zmienne losowe U i Z_j są dodatnio zależne w sensie $LRD+$, dla $j = 1, \dots, n$.*

Z definicji (por. [3]), zależność typu $LRD+$ pomiędzy U i Z_j oznacza, że dla dowolnych u_i, z_i ($i = 1, 2, 3$) takich, że $u_1 < u_2 < u_3$, $z_1 < z_2 < z_3$ mamy

$$\begin{aligned} P(u_1 < U \leq u_2, z_1 < Z_j \leq z_2) P(u_2 < U \leq u_3, z_2 < Z_j \leq z_3) \\ \geq P(u_1 < U \leq u_2, z_2 < Z_j \leq z_3) P(u_2 < U \leq u_3, z_1 < Z_j \leq z_2). \end{aligned}$$

Jeżeli dla każdego $u \in \mathbf{R}$ istnieje gęstość $g(u, \cdot)$ warunkowego rozkładu ($Z_j | U = u$) względem pewnej wspólnej miary, to zależność $LRD+$ jest równoważna temu, że iloraz $g(u', z)/g(u, z)$ maleje ze wzrostem z dla dowolnych $u' \leq u$ (op. cit.).

Wszystkie wymienione założenia są inwariantne ze względu na rosnące przekształcenia zmiennych Z_j . Jeżeli wektor $(U, Z_1, \dots, Z_n)$ spełnia którekolwiek z założeń (I, II, III lub III'), $\phi_1, \dots, \phi_n$ są funkcjami rosnącymi, to wektor $(U, \phi_1(Z_1), \dots, \phi_n(Z_n))$ również spełnia to założenie. Jeżeli jednowymiarowe rozkłady brzegowe zmiennych Z_j są ciągłe, to możemy bez straty ogólności przyjąć, że są to rozkłady jednostajne: dla $0 \leq z \leq 1$, $j = 1, \dots, n$

$$(3) \quad P(Z_j < z) = z.$$

Istotnie, jeżeli dystrybuenta F_j zmiennej Z_j jest ciągłą, to kładąc $\phi_j = F_j$ dostajemy przekształconą zmienną $\phi_j(Z_j)$ o rozkładzie jednostajnym.

3. Przypadek binarnej cechy ukrytej. Niech cecha ukryta U będzie zmienną losową binarną, przyjmującą wartości 0 lub 1. Dla podkreślenia tego faktu, w tym paragrafie oznaczmy tę zmienną odmiennym symbolem I . O zmiennych losowych Z_j założymy, że mają rozkłady ciągłe. Rozpatrujemy zatem układ zmiennych losowych

$$(I, Z_1, \dots, Z_n).$$

Oznaczmy

$$\begin{aligned} (1) \quad \pi &= P(I = 0) = 1 - P(I = 1), \\ (2) \quad P(Z_j < z | I = 0) &= F_j(z) \\ P(Z_j < z | I = 1) &= G_j(z) \quad (j = 1, \dots, n), \end{aligned}$$

$$(3) \quad H(z_1, \dots, z_n) = P(Z_1 < z_1, \dots, Z_n < z_n).$$

Przyjmijmy, że $0 < \pi < 1$ i układ zmiennych losowych $(I, Z_1, \dots, Z_n)$ spełnia założenie (I), które możemy zapisać w następującej postaci:

$$(4) \quad \begin{aligned} P(Z_1 < z_1, \dots, Z_n < z_n | I = 0) &= F_1(z_1) \dots F_n(z_n), \\ P(Z_1 < z_1, \dots, Z_n < z_n | I = 1) &= G_1(z_1) \dots G_n(z_n). \end{aligned}$$

Założymy, że rozkłady brzegowe zmiennych Z_j są ciągłe. Zgodnie uwagą z końca poprzedniego paragrafu przyjmujemy, że rozkłady te są jednostajne, czyli

$$(5) \quad H(z, 1, \dots, 1) = \dots = H(1, \dots, 1, z) = z \quad (0 \leq z \leq 1).$$

Na mocy (5) mamy $z = \pi F_j(z) + (1 - \pi)G_j(z)$, a więc

$$(6) \quad G_j(z) = \frac{z - \pi F_j(z)}{1 - \pi}.$$

W dalszym ciągu rozważać będziemy przypadek tylko dwóch ekspertów, czyli przyjmujemy $n = 2$.

STWIERDZENIE 1. *Jeżeli układ zmiennych losowych (I, Z_1, Z_2) spełnia założenie (I), to zachodzi następujący związek:*

$$(7) \quad H(z_1, z_2) - z_1 z_2 = \frac{\pi}{1 - \pi} (F_1(z_1) - z_1)(F_2(z_2) - z_2).$$

Dowód. Wykorzystując (4) i (6) mamy

$$\begin{aligned} H(z_1, z_2) &= \pi F_1(z_1)F_2(z_2) + (1 - \pi)G_1(z_1)G_2(z_2) \\ &= \pi F_1(z_1)F_2(z_2) + \frac{1}{1 - \pi} [z_1 - \pi F_1(z_1)][z_2 - \pi F_2(z_2)] \\ &= \frac{\pi}{1 - \pi} [F_1(z_1)F_2(z_2) - z_1 F_2(z_2) - z_2 F_1(z_1) + z_1 z_2]. \quad \blacksquare \end{aligned}$$

Przeformułujemy teraz założenia (III) i (III'). Założenie (III) sprowadza się do nierówności

$$(8) \quad G_j(z) \leq F_j(z)$$

dla $0 \leq z \leq 1$, $j = 1, 2$.

STWIERDZENIE 2. *Niech układ zmiennych losowych (I, Z_1, Z_2) spełnia założenie (I). Zachodzenie każdej z następujących dwóch nierówności:*

$$1) \ z \leq F_j(z), \quad 2) \ G_j(z) \leq z,$$

dla $0 \leq z \leq 1$, $j = 1, 2$, jest równoważne warunkowi (III). Nierówność

$$H(z_1, z_2) \geq z_1 z_2$$

zachodzi dla $z_1 z_2 \in [0, 1]$ wtedy i tylko wtedy, kiedy warunek (III) jest spełniony albo dla wektora (I, Z_1, Z_2) , albo dla wektora $(1 - I, Z_1, Z_2)$.

Dowód. Nierówność (8) możemy, na mocy (6), przepisać w postaci $z - \pi F_j(z) \leq (1 - \pi)F_j(z)$, a więc jest ona równoważna nierówności $z \leq F_j(z)$. Równoważność nierówności (8) i $G_j(z) \leq z$ pokazuje się podobnie. Wreszcie, dla dowodu ostatniej tezy naszego stwierdzenia wystarczy zauważyć, że prawa strona we wzorze (7) jest stale nieujemna wtedy i tylko wtedy, kiedy obie funkcje $F_1(z) - z$ i $F_2(z) - z$ są nieujemne lub obie są niedodatnie.

Ostatnia teza Stwierdzenia 2 mówi o dodatniej kwadrantowej zależności (zależności typu $QD+$, por. [3]) wektora (Z_1, Z_2) . Zauważmy, że założenie (I) i równość (5) gwarantują, że dystrybuanty F_j i G_j spełniają warunek Lipschitza. Istotnie, oznaczając ilorazy różnicowe $(F_j(b) - F_j(a))/(b - a)$ i $(G_j(b) - G_j(a))/(b - a)$ odpowiednio przez Δ_F i Δ_G mamy, na mocy (6), $0 \leq \Delta_G = (1 - \pi\Delta_F)/(1 - \pi)$, skąd $\Delta_F \leq 1/\pi$ oraz $0 \leq \pi\Delta_F = 1 - (1 - \pi)\Delta_G$, skąd $\Delta_G \leq 1/(1 - \pi)$. Dystrybuanty F_j i G_j mają zatem gęstości względem miary Lebesgue'a, które oznaczmy przez f_j i g_j ; dwuwymiarową gęstość odpowiadającą dystrybuancie H oznaczmy przez h .

Założenie (III') sprowadza się zatem, w rozpatrywanym obecnie przypadku, do tego, że $g_j(z)/f_j(z)$ jest niemalejącą funkcją argumentu z .

STWIERDZENIE 3. *Jeśli układ zmiennych losowych (I, Z_1, Z_2) spełnia założenie (I), to założenie (III') można wyrazić za pomocą jednego następujących równoważnych warunków (dla $j = 1, 2$):*

- 1) F_j jest funkcją wklęsłą na odcinku $[0, 1]$,
- 2) G_j jest funkcją wypukłą na odcinku $[0, 1]$.

Nierówność

$$h(a_1, a_2)h(b_1, b_2) \geq h(a_1, b_2)h(b_1, a_2)$$

zachodzi dla $0 \leq a_j < b_j \leq 1$ ($j = 1, 2$) wtedy i tylko wtedy, kiedy warunek (III') jest spełniony albo dla wektora (I, Z_1, Z_2) , albo dla wektora $(1 - I, Z_1, Z_2)$.

Dowód. Ponieważ z (6) wynika, że $g_j(z) = (1 - \pi f_j(z))/(1 - \pi)$, więc $g_j(z)/f_j(z) = (1 - \pi)^{-1}(1/f_j(z) - \pi)$ rośnie wtedy i tylko wtedy, kiedy f_j maleje. Równoważność 2) i (III') pokazuje się podobnie. Wreszcie zauważmy, że

$$h(z_1, z_2) = \frac{\pi}{1 - \pi} [f_1(z_1) - 1][f_2(z_2) - 1] + 1.$$

Wynika stąd tożsamość $h(a_1, a_2)h(b_1, b_2) - h(a_1, b_2)h(b_1, a_2) = \pi(1 - \pi)^{-1} \times (f_1(b_1) - f_1(a_1))(f_2(b_2) - f_2(a_2))$. Wyrażenie po prawej stronie jest dla $0 \leq a_j < b_j \leq 1$ nieujemne wtedy i tylko wtedy, kiedy obie gęstości f_j są niemalejące lub obie są nierosnące. Kończy to dowód ostatniej części stwierdzenia. ■

Ostatnia teza Stwierdzenia 3 mówi o dodatniej zależności typu $LRD+$ wektora (Z_1, Z_2) .

Stwierdzenia 1, 2 i 3 pozwolą nam na scharakteryzowanie rodziny tych dwuwymiarowych rozkładów, które są łącznymi rozkładami opinii dwóch ekspertów w rozpatrywanym modelu.

Niech H będzie dwuwymiarową dystrybuantą rozkładu dodatnio kwadrantowo zależnego o brzegowych rozkładach jednostajnych na odcinku $[0, 1]$. Warunkiem koniecznym na to, aby H była dystrybuantą brzegową wektora (Z_1, Z_2) dla pewnego układu (I, Z_1, Z_2) , spełniającego założenie (I) jest, na mocy wzoru (7), możliwość przedstawienia H w postaci

$$(9) \quad H(z_1, z_2) = r_1(z_1)r_2(z_2) + z_1z_2$$

dla pewnych funkcji $r_1, r_2 : [0, 1] \rightarrow \mathbb{R}$. Okazuje się, że jest to również warunek dostateczny. Mówi o tym pierwsza część następującego Twierdzenia 1. Sformułowanie tego twierdzenia poprzedzimy kilkoma spostrzeżeniami. Jeżeli dystrybuanta H jest postaci (9) to, z wyjątkiem trywialnego przypadku $H(z_1, z_2) \equiv z_1z_2$, funkcje r_1 i r_2 są wyznaczone jednoznacznie z dokładnością do stałych mnożących. Ponieważ $H(z_1, z_2) \geq z_1z_2$, więc można założyć, że $r_1, r_2 \geq 0$ i, oczywiście, $r_j(0) = r_j(1) = 0$. Ponadto, dla wielkości s_j i k_j określonych (dla $j = 1, 2$) wzorami

$$(10) \quad s_j = \inf_{0 \leq a < b \leq 1} \frac{r_j(b) - r_j(a)}{b - a}, \quad k_j = \sup_{0 \leq a < b \leq 1} \frac{r_j(b) - r_j(a)}{b - a}$$

mamy $-\infty < s_j < 0 < k_j < \infty$ oraz zachodzą nierówności $s_1k_2 \geq -1$ i $s_2k_1 \geq -1$. W samej rzeczy, ponieważ H jest dystrybuantą, więc dla $0 \leq a_j < b_j \leq 1$ mamy $0 \leq H(b_1, b_2) - H(b_1, a_2) - H(a_1, b_2) + H(a_1, a_2)$. Stąd $(r_1(b_1) - r_1(a_1))(r_2(b_2) - r_2(a_2)) \geq -(b_1 - a_1)(b_2 - a_2)$, a zatem $s_1k_2 \geq -1$ i $k_1s_2 \geq -1$.

Twierdzenie 1. Niech $H(\cdot, \cdot)$ będzie dwuwymiarową dystrybuantą taką, że $H(z, 1) = H(1, z) = z$ ($0 \leq z \leq 1$) i $H(z_1, z_2) \geq z_1z_2$ dla $z_1, z_2 \in [0, 1]$.

1) Na to, żeby H była dystrybuantą brzegową wektora (Z_1, Z_2) dla pewnego wektora losowego (I, Z_1, Z_2) , spełniającego założenia (I) i (III) potrzeba i wystarcza aby H można było przedstawić w postaci (9).

2) Założmy, że dystrybuanta H spełnia warunek (9). Wszystkie łączne rozkłady wektora (I, Z_1, Z_2) o brzegowej dystrybuancie H , spełniające (I, III), wyznaczone są przez parametr π (wzór (1)) i funkcje F_j (wzór (2)) następującej postaci:

$$(11) \quad \pi = \frac{1}{1 + \alpha_1\alpha_2}$$

$$(12) \quad F_j(z) = \alpha_j r_j(z) + z$$

dla dowolnych liczb α_1 i α_2 spełniających nierówności

$$k_2 \leq \alpha_1 \leq -1/s_1, \quad k_1 \leq \alpha_2 \leq -1/s_2,$$

gdzie s_j i k_j są określone wzorami (10).

3) Warunkiem koniecznym i dostatecznym na to, żeby opisanie w powyższy sposób rozkładu wektora (I, Z_1, Z_2) spełniały ponadto warunek (III') jest wklęsłość funkcji r_1 i r_2 .

Dowód. Konieczność warunku sformułowanego w części 1) twierdzenia została już pokazana.

Założmy teraz że zachodzi (9). Łatwo zauważyć, biorąc pod uwagę wzór (7), że w celu znalezienia wszystkich rozkładów wektora (I, Z_1, Z_2) spełniających (I, III) o brzegowej dystrybucie H wystarczy rozpatrzeć π oraz funkcje F_j określone wzorami (11) i (12). Równość (7) zachodzi wtedy automatycznie. Funkcje G_j określimy wzorem (6). Pozostaje tylko zbadać, dla jakiego wyboru współczynników α_1 i α_2 funkcje F_j oraz G_j są dystrybucjami rozkładów prawdopodobieństwa. Ponieważ $F_j(0) = G_j(0) = 0$ i $F_j(1) = G_j(1) = 1$, więc wystarczy sprawdzić, czy funkcje F_j i G_j są rosnące. Zauważmy, że ilorazy różnicowe $(F_j(b) - F_j(a))/(b - a) = \alpha_j(r_j(b) - r_j(a))/(b - a) + 1$ są nieujemne wtedy i tylko wtedy, kiedy $\alpha_j s_j + 1 \geq 0$, czyli $\alpha_j \leq -1/s_j$. Z kolei ilorazy różnicowe $(G_j(b) - G_j(a))/(b - a) = (1 - \pi)^{-1} \{1 - \pi[\alpha_j(r_j(b) - r_j(a))/(b - a) + 1]\}$ są nieujemne wtedy i tylko wtedy, kiedy $\alpha_j k_j \leq (1 - \pi)/\pi$. Wobec równości $(1 - \pi)/\pi = \alpha_1 \alpha_2$ nierówność $\alpha_1 k_1 \leq (1 - \pi)/\pi$ jest równoważna $\alpha_2 \geq k_1$, a nierówność $\alpha_2 k_2 \leq (1 - \pi)/\pi$ równoważna $\alpha_1 \geq k_2$. Ponieważ pokazaliśmy wcześniej, że przedziały $[k_2, -1/s_1]$ i $[k_1, -1/s_2]$ są niepuste, kończy to dowód тез 1) i 2) naszego twierdzenia.

Dla uzasadnienia tezy 3) wystarczy powołać się na Stwierdzenie 3. Zauważmy jeszcze, że dla wklęsłej funkcji r_j mamy po prostu $k_j = \frac{dr_j(z)}{dz}|_{z=0}$ i $s_j = \frac{dr_j(z)}{dz}|_{z=1}$. ■

U w a g a. Z części 2) Twierdzenia 1 wynika w szczególności nieestymowalność π na podstawie obserwacji (Z_1, Z_2) w rozważanym modelu. Znając rozkład opinii ekspertów możemy jedynie stwierdzić, że

$$1 - \frac{1}{1 + s_1 s_2} \leq \pi \leq \frac{1}{1 + k_1 k_2}. \quad \blacksquare$$

Przeanalizujemy teraz sytuację, kiedy dołączamy do rozpatrywanych dotychczas założeń warunek (II), mówiący o jednorodności opinii ekspertów. Zaczniemy od następującego spostrzeżenia. Jeżeli wektor (I, Z_1, Z_2) spełnia założenia (I, II), to prawdziwa jest tożsamość (dla $z_1, z_2 \in [0, 1]$):

$$(13) \quad [H(z_1, z_2) - z_1 z_2]^2 = [H(z_1, z_1) - z_1^2][H(z_2, z_2) - z_2^2].$$

W samej rzeczy, przyjmując we wzorze (7) $F = F_1 = F_2$ i $z = z_1 = z_2$ otrzymujemy $(F(z) - z)^2 = \frac{1-\pi}{\pi}(H(z, z) - z^2)$. Korzystając ponownie ze wzoru (7), dostajemy (13). Załóżmy, że funkcja H spełnia tożsamość (13) i jest dystrybuantą rozkładu dodatnio kwadrantowo zależnego, o jednostajnych rozkładach brzegowych. Oznaczmy

$$(14) \quad r(z) = \sqrt{H(z, z) - z^2},$$

$$(15) \quad s = \inf_{0 \leq a < b \leq 1} \frac{r(b) - r(a)}{b - a}, \quad k = \sup_{0 \leq a < b \leq 1} \frac{r(b) - r(a)}{b - a}.$$

Mamy teraz $H(z_1, z_2) = r(z_1)r(z_2) + z_1z_2$ i na mocy uwag poprzedzających Twierdzenie 1 widać, że $-\infty < s \leq 0 \leq k < \infty$ i $ks \geq -1$.

Sformułujemy teraz odpowiednik Twierdzenia 1 dla modelu jednorodnych ekspertów.

Twierdzenie 2. Niech $H(\cdot, \cdot)$ będzie dwuwymiarową dystrybuantą taką, że $H(z, 1) = H(1, z) = z$ ($0 \leq z \leq 1$) i $H(z_1, z_2) \geq z_1z_2$ dla $z_1, z_2 \in [0, 1]$.

1) Na to, żeby H była dystrybuantą brzegową wektora (Z_1, Z_2) dla pewnego wektora losowego (I, Z_1, Z_2) , spełniającego założenia (I, II i III) potrzeba i wystarcza aby spełniona była tożsamość (13).

2) Załóżmy, że dystrybuanta H spełnia tożsamość (13). Wszystkie łączne rozkłady wektora (I, Z_1, Z_2) o brzegowej dystrybuancie H , spełniające (I, II, III) wyznaczone są przez parametr π i dystrybuantę $F = F_1 = F_2$ postaci:

$$(16) \quad \pi = \frac{1}{1 + \alpha^2},$$

$$(17) \quad F(z) = \alpha r(z) + z,$$

gdzie r jest funkcją określoną wzorem (14), zaś α jest dowolną liczbą spełniającą nierówność $k \leq \alpha \leq -1/s$ dla k i s danych wzorami (15).

3) Warunkiem koniecznym i dostatecznym na to, żeby opisane w powyższy sposób rozkłady wektora (I, Z_1, Z_2) spełniały ponadto warunek (III') jest wklęsłość funkcji r .

D o w ó d. Wynika natychmiast z Twierdzenia 1. ■

4. Minimaksowe reguły decyzyjne. Rozważmy teraz klasyczne zadanie dyskryminacji statystycznej w badanym przez nas modelu z binarną cechą ukrytą. Jeśli I jest zmienną losową o wartościach 0 lub 1, zaś Z – wektorem losowym w $\mathbb{R}^k$, to zadanie wnioskowania o I na podstawie Z formułuje się w analizie dyskryminacyjnej najczęściej w następujący sposób. Rozważa się zbiór decyzji $\Delta = \{0, 1\}$ i funkcję strat $L : \{0, 1\} \times \{0, 1\} \rightarrow \mathbb{R}$

określoną wzorem

$$(1) \quad L(i, d) = \begin{cases} c & \text{jeśli } i = 0 \text{ i } d = 1, \\ 1 - c & \text{jeśli } i = 1 \text{ i } d = 0, \\ 0 & \text{w pozostałych przypadkach} \end{cases}$$

(dla pewnego współczynnika $c \in (0, 1)$). Poszukuje się funkcji decyzyjnej $\delta : \mathbb{R}^k \rightarrow \{0, 1\}$ minimalizującej ryzyko

$$R_P(\delta) = E_P L(I, \delta(Z)),$$

czyli wartość oczekiwaną straty względem łącznego rozkładu P wektora (I, Z) . Rozwiązanie tak postawionego zadania dla danego rozkładu P jest dobrze znane (np. [4], p. 7.4).

Rozpatrzmy teraz, podobnie jak w poprzednim paragrafie, wektor losowy

$$(I, Z_1, Z_2)$$

gdzie I jest binarną cechą ukrytą. Zakładamy, że łączny rozkład P tego wektora należy do rodziny $\mathcal{P}$ wszystkich rozkładów na $\{0, 1\} \times \mathbb{R}^2$ spełniających założenia (I) i (III') oraz warunek unormowania (2.3) z paragrafu 2. Zajmiemy się wyznaczeniem minimaxowego rozwiązania problemu wnioskowania o I na podstawie obserwacji zmiennej Z_1 (opinii pojedynczego eksperta). Mówiąc dokładniej, zakładamy znajomość rozkładu brzegowego $Q = P|_Z$ wektora (Z_1, Z_2) i poszukujemy takiej funkcji decyzyjnej $\delta : [0, 1] \rightarrow \{0, 1\}$, która minimalizuje maksimum ryzyka

$$(2) \quad \sup_{\{P \in \mathcal{P} : P|_Z = Q\}} R_P(\delta) = \sup_{\{P \in \mathcal{P} : P|_Z = Q\}} E_P L(I, \delta(Z_1))$$

dla funkcji strat L danej wzorem (1). Stosując oznaczenia z paragrafu 3 możemy ryzyko R_P w powyższym wzorze wyrazić w postaci

$$(3) \quad R_P(\delta) = \int_0^1 \{c\pi f_1(z_1)\delta(z_1) + (1-c)(1-\pi)g_1(z_1)[1-\delta(z_1)]\} dz_1$$

gdzie f_1 i g_1 są warunkowymi gęstościami rozkładu zmiennej Z_1 dla, odpowiednio, $I = 0$ i $I = 1$, zaś $\pi = P(I = 0)$. W rozważanym przypadku rodzina $\{P \in \mathcal{P} : P|_Z = Q\}$ jest dwuparametrową rodziną rozkładów opisaną w Twierdzeniu 1 w poprzednim paragrafie. Gęstość f_1 jest mianowicie postaci $f_1(z) = 1 + \alpha_1 r'_1(z)$, mamy $\pi = 1/(1 + \alpha_1 \alpha_2)$ i $g_1(z) = (1 - \pi f_1(z))/(1 - \pi) = z - (1/\alpha_2)r'_1(z)$. Parametry (α_1, α_2) przebiegają przedział $k_2 \leq \alpha_1 \leq -1/s_1$, $k_1 \leq \alpha_2 \leq -1/s_2$, zaś r'_1 jest znaną funkcją (po pochodną funkcji r_1 występującej w (3.9)).

Oznaczmy dla uproszczenia

$$\alpha = \alpha_1, \quad \beta = 1/\alpha_2, \quad \alpha^* = -1/s_1, \quad \alpha_* = k_2, \quad \beta^* = 1/k_1, \quad \beta_* = -s_2, \\ r(z) = r_1(z), \quad \rho(z) = r'_1(z),$$

$$R(\alpha, \beta, \delta) = R_P(\delta).$$

Wzór (3) przyjmuje teraz postać

$$R(\alpha, \beta, \delta) = \int_0^1 \left\{ c \frac{\beta}{\alpha + \beta} [1 + \alpha \rho(z)] \delta(z) + (1 - c) \frac{\alpha}{\alpha + \beta} [1 - \beta \rho(z)] [1 - \delta(z)] \right\} dz$$

i po prostych przekształceniach, korzystając z tego, że $\int_0^1 \rho(z) dz = 0$, otrzymujemy

$$R(\alpha, \beta, \delta) = \frac{1}{\alpha + \beta} \left\{ (1 - c)\alpha + [c\beta - (1 - c)\alpha] \int_0^1 \delta(z) dz + \alpha\beta \int_0^1 \rho(z) \delta(z) dz \right\}.$$

Z tego wzoru widać, że dla dowolnej funkcji decyzyjnej δ możemy znaleźć funkcję decyzyjną

$$(4) \quad \delta_\zeta(z) = \begin{cases} 0 & \text{dla } z \leq \zeta, \\ 1 & \text{dla } z > \zeta, \end{cases}$$

jednostajnie (dla dowolnych α i β) lepszą od funkcji δ . Istotnie, wystarczy wziąć $\zeta = 1 - \int_0^1 \delta(z) dz$ i zauważyć, że ρ jest funkcją nierosnącą. Wynika stąd, że w poszukiwaniu minimaksowej funkcji decyzyjnej możemy ograniczyć się do rozpatrywania funkcji postaci (4) (progowych reguł klasyfikacji). Ostatecznie zatem nasze zadanie przybiera postać: znaleźć punkt $\zeta \in [0, 1]$ w którym osiągnąony jest kres dolny

$$(5) \quad \inf_{0 \leq \zeta \leq 1} \sup_{\alpha_* \leq \alpha \leq \alpha^*} \sup_{\beta_* \leq \beta \leq \beta^*} R(\alpha, \beta, \delta_\zeta)$$

gdzie, jak łatwo zauważyć,

$$(6) \quad R(\alpha, \beta, \delta_\zeta) = \frac{1}{\alpha + \beta} \{ c\beta + \zeta[(1 - c)\alpha - c\beta] - \alpha\beta r(\zeta) \}.$$

Obliczmy pochodne cząstkowe funkcji $R(\alpha, \beta, \delta_\zeta)$:

$$(7) \quad \frac{\partial R}{\partial \alpha} = \frac{\beta}{(\alpha + \beta)^2} [\zeta - c - \beta r(\zeta)]$$

$$(8) \quad \frac{\partial R}{\partial \beta} = \frac{\alpha}{(\alpha + \beta)^2} [c - \zeta - \alpha r(\zeta)]$$

$$(9) \quad \frac{\partial R}{\partial \zeta} = \frac{-c\beta + (1 - c)\alpha - \alpha\beta r(\zeta)}{\alpha + \beta}$$

Ustalmy teraz ζ . Rozpatrzmy następujące przypadki.

1. Niech $\zeta \leq c$. Mamy wtedy $\frac{\partial R}{\partial \alpha} \leq 0$ na mocy (7). Dla dowolnego, ustalonego β funkcja R przyjmuje maksimum $\sup_\alpha R(\alpha, \beta, \delta_\zeta)$ dla $\alpha = \alpha_*$.

Jeśli przy tym $\zeta + \alpha_* r(\zeta) \leq c$, to $\frac{\partial R}{\partial \beta} \geq 0$ na mocy (8) i maksimum $\sup_{\alpha, \beta} R(\alpha, \beta, \delta_\zeta)$ jest osiągnięte w punkcie (α_*, β^*) . Jeśli $\zeta + \alpha_* r(\zeta) \geq c$, to maksimum jest osiągnięte w (α_*, β_*) .

2. Niech $\zeta \geq c$. Podobne rozważania prowadzą do wniosku, że funkcja R osiąga maksimum w punkcie (α_*, β_*) , jeśli $\zeta - \beta_* r(\zeta) \leq c$ oraz w punkcie (α^*, β_*) , jeśli $\zeta - \beta_* r(\zeta) \geq c$.

Oznaczmy przez z_* i z^* punkty spełniające równania

$$(11) \quad z_* + \alpha_* r(z_*) = c, \quad z^* - \beta_* r(z^*) = c.$$

Punkty takie istnieją i spełniają nierówność $z_* \leq z^*$. Z faktu, że funkcja r jest wklęsła wynika, że dla $\zeta \leq z_*$ zachodzi nierówność $\zeta + \alpha_* r(\zeta) \leq c$, dla $z_* \leq \zeta \leq z^*$ mamy $\zeta - \beta_* r(\zeta) \leq c \leq \zeta + \alpha_* r(\zeta)$, zaś dla $\zeta \geq z^*$ mamy $\zeta - \beta_* r(\zeta) \geq c$. Reasumując,

$$(12) \quad \max_{\alpha, \beta} R(\alpha, \beta, \delta_\zeta) = \begin{cases} R(\alpha_*, \beta^*, \delta_\zeta) & \text{dla } \zeta \leq z_*, \\ R(\alpha_*, \beta_*, \delta_\zeta) & \text{dla } z_* \leq \zeta \leq z^*, \\ R(\alpha^*, \beta_*, \delta_\zeta) & \text{dla } \zeta \geq z^*, \end{cases}$$

Wprowadźmy następujące oznaczenia:

$$A_1^- = \frac{\partial R}{\partial \zeta}(\alpha_*, \beta^*, z_*-), \quad A_1^+ = \frac{\partial R}{\partial \zeta}(\alpha_*, \beta_*, z_*+), \\ A_2^- = \frac{\partial R}{\partial \zeta}(\alpha_*, \beta_*, z^*-), \quad A_2^+ = \frac{\partial R}{\partial \zeta}(\alpha^*, \beta_*, z^*+)$$

($\frac{\partial R}{\partial \zeta}(\alpha, \beta, \zeta \pm)$ oznaczają tu pochodne lewo- i prawostronne funkcji $\zeta \rightarrow R(\alpha, \beta, \delta_\zeta)$). Łatwo zauważyć, że $A_1^- \leq A_1^+ \leq A_2^- \leq A_2^+$. Możemy teraz wypowiedzieć następujące twierdzenie.

Twierdzenie 3. *Jeżeli układ zmiennych losowych (I, Z_1, Z_2) spełnia założenia (I) i (III') to maksimum ryzyka (2) jest najmniejsze dla funkcji decyzyjnej δ_ζ (postaci (4)), gdzie ζ jest określone następująco:*

1) jeśli $A_1^- \geq 0$ to ζ jest dowolnym punktem spełniającym nierówność

$$\rho(\zeta+) \leq \frac{1-c}{\beta^*} - \frac{c}{\alpha_*} \leq \rho(\zeta-)$$

2) jeśli $A_1^- \leq 0$ i $A_1^+ \geq 0$ to $\zeta = z_*$,

3) jeśli $A_1^+ \leq 0$ i $A_2^- \geq 0$ to ζ jest punktem spełniającym nierówność

$$\rho(\zeta+) \leq \frac{1-c}{\beta_*} - \frac{c}{\alpha_*} \leq \rho(\zeta-),$$

4) jeśli $A_2^- \leq 0$ i $A_2^+ \geq 0$ to $\zeta = z^*$,

5) jeśli $A_2^+ \leq 0$ to ζ jest punktem spełniającym nierówność

$$\rho(\zeta+) \leq \frac{1-c}{\beta_*} - \frac{c}{\alpha^*} \leq \rho(\zeta-).$$

Dowód. Zauważmy, że funkcja (12) wyrażająca maksimum ryzyka jest wypukła i A_1^-, A_1^+, A_2^- oraz A_2^+ są jej pochodnymi jednostronnymi w punktach z_* i z^* . Tezę twierdzenia otrzymujemy natychmiast, badając przebieg tej funkcji na przedziałach $[0, z_*]$, $[z_*, z^*]$ i $[z^*, 1]$ przy wykorzystaniu wzoru (9). ■

Rozważmy teraz analogiczne zadanie przy dodatkowym założeniu jednorodności ekspertów. Innymi słowy poszukujemy funkcji decyzyjnej minimum w przypadku, kiedy rodzina $\mathcal{P}$ rozkładów wektora (I, Z_1, Z_2) składa się ze wszystkich rozkładów na $\{0, 1\} \times \mathbb{R}^2$ spełniających oprócz założeń (I) i (III') również założenie (II). Wiemy na podstawie Twierdzenia 2, że $\{P \in \mathcal{P} : P|_Z = Q\}$ jest w tym przypadku rodziną jednoparametrową; gęstości f i g są postaci $f(z) = 1 + \alpha r'(z)$, $g(z) = z - (1/\alpha)r'(z)$, zaś $\pi = 1/(1 + \alpha^2)$. Parametr α przebiega przedział $\alpha_* = k \leq \alpha \leq -1/s = \alpha^*$, zaś $r' = \rho$ jest znaną funkcją (pochodną funkcji r określonej wzorem (3.14)). Nasze zadanie przybiera teraz postać: znaleźć punkt $\zeta \in [0, 1]$ w którym osiągany jest kres dolny

$$(13) \quad \inf_{0 \leq \zeta \leq 1} \sup_{\alpha_* \leq \alpha \leq \alpha^*} R(\alpha, 1/\alpha, \delta_\zeta),$$

gdzie $R(\cdot, \cdot, \cdot)$ jest funkcją daną wzorem (6). Ze wzorów (7) i (8) wynika, że

$$(14) \quad \frac{d}{d\alpha} R(\alpha, 1/\alpha, \delta_\zeta) = \frac{2/\alpha}{(\alpha + 1/\alpha)^2} \left[\zeta - c + \frac{1}{2} r(\zeta)(\alpha - 1/\alpha) \right].$$

Ustalmy teraz ζ . Rozpatrzmy następujące przypadki.

1. Jeśli $\zeta + \frac{1}{2} r(\zeta)(\alpha_* - 1/\alpha_*) \leq c$ to $\frac{d}{d\alpha} R \leq 0$ i funkcja R przyjmuje maksimum w punkcie α_* .

2. Jeśli $\zeta + \frac{1}{2} r(\zeta)(\alpha_* - 1/\alpha_*) \leq c \leq \zeta + \frac{1}{2} r(\zeta)(\alpha^* - 1/\alpha^*)$ to funkcja R przyjmuje maksimum w punkcie $\alpha = \alpha(\zeta)$ będącym rozwiązaniem równania $\zeta + \frac{1}{2} r(\zeta)(\alpha - 1/\alpha) = c$.

3. Jeśli $c \leq \zeta + \frac{1}{2} r(\zeta)(\alpha_* - 1/\alpha_*)$ to $\frac{d}{d\alpha} R \geq 0$ i funkcja R przyjmuje maksimum w punkcie α^* .

Oznaczmy przez $z_\#$ i $z^\#$ punkty spełniające równania

$$(15) \quad z_\# + \frac{1}{2} r(z_\#)(\alpha^* - 1/\alpha^*) = c, \quad z^\# + \frac{1}{2} r(z^\#)(\alpha_* - 1/\alpha_*) = c.$$

Punkty takie istnieją i spełniają nierówność $z_* \leq z_\# \leq z^\# \leq z^*$. Dla $\zeta \leq z_\#$, $z_\# \leq \zeta \leq z^\#$ i $\zeta \geq z^\#$ zachodzą odpowiednio przypadki 1, 2 i 3 wymienione powyżej. Funkcję maksimum ryzyka możemy zatem wyrazić następująco.

$$(16) \quad \max_{\alpha} R(\alpha, 1/\alpha, \delta_\zeta) = \begin{cases} R(\alpha_*, 1/\alpha_*, \delta_\zeta) & \text{dla } \zeta \leq z_\#, \\ R(\alpha(\zeta), 1/\alpha(\zeta), \delta_\zeta) & \text{dla } z_\# \leq \zeta \leq z^\#, \\ R(\alpha^*, 1/\alpha^*, \delta_\zeta) & \text{dla } \zeta \geq z^\#. \end{cases}$$

Jest to funkcja wypukła, jako kres górny rodziny funkcji wypukłych. Oznaczmy

$$B_1 = \frac{\partial R}{\partial \zeta}(\alpha_*, 1/\alpha_*, z_{\#}), \quad B_2 = \frac{\partial R}{\partial \zeta}(\alpha^*, 1/\alpha^*, z_{\#}).$$

Mamy $B_1 \leq B_2$. Jeśli $B_1 \geq 0$, to maksimum funkcji (16) leży w przedziale $[0, z_{\#}]$, jeśli $B_1 \leq 0 \leq B_2$ – w przedziale $[z_{\#}, z^{\#}]$, jeśli zaś $B_2 \leq 0$ – w przedziale $[z^{\#}, 1]$. Badanie przebiegu tej funkcji na przedziałach $[0, z_{\#}]$ i $[z^{\#}, 1]$ sprowadza się do bezpośredniego zastosowania wzoru (9). Zbadajmy teraz przebieg funkcji (16) na przedziale $[z_{\#}, z^{\#}]$. Załóżmy, dla uproszczenia, że funkcja ρ jest ciągła. Ponieważ $\frac{d}{d\zeta} R(\alpha(\zeta), 1/\alpha(\zeta), \delta_{\zeta}) = \frac{dR}{d\alpha} \frac{d\alpha}{d\zeta} + \frac{\partial R}{\partial \zeta}$, zaś $\frac{dR}{d\alpha} = 0$ dla $\alpha = \alpha(\zeta)$, więc w punkcie ζ , w którym funkcja przyjmuje minimum mamy $\frac{\partial R}{\partial \zeta} = 0$. Pochodna $\frac{\partial R}{\partial \zeta}$ jest dana wzorem (9).

Udowodniliśmy w ten sposób następujące twierdzenie.

Twierdzenie 4. *Jeżeli układ zmiennych losowych (I, Z_1, Z_2) spełnia założenia (I), (II), (III') i funkcja r (określona wzorem (3.14)) ma ciągłą pochodną ρ to maksimum ryzyka (2) jest najmniejsze dla funkcji decyzyjnej δ_{ζ} (postaci (4)), gdzie ζ jest określone następująco:*

1) jeśli $B_1 \geq 0$ to ζ jest rozwiązaniem równania

$$\rho(\zeta) = (1 - c)\alpha_* - \frac{c}{\alpha_*},$$

2) jeśli $B_1 \leq 0 \leq B_2$, to ζ jest punktem spełniającym układ równań

$$\begin{cases} \rho(\zeta) = (1 - c)\alpha - \frac{c}{\alpha}, \\ \zeta + \frac{1}{2}r(\zeta)(\alpha - 1/\alpha) = c, \end{cases}$$

3) jeśli $B_2 \leq 0$ to ζ jest rozwiązaniem równania

$$\rho(\zeta) = (1 - c)\alpha^* - \frac{c}{\alpha^*}.$$

Dla klasycznej funkcji strat wyznaczyliśmy zatem, w rozważanych przez nas modelach, minimaksowe reguły decyzyjne w klasie reguł zależnych od pojedynczej współrzędnej obserwowalnego wektora (Z_1, Z_2) . Zauważmy na zakończenie, że problem znalezienia minimaksowej funkcji decyzyjnej wśród funkcji obu współrzędnych prowadzi do pewnego zagadnienia wariacyjnego, którego autorzy tej pracy nie potrafili rozwiązać.

Bibliografia

- [1] V. Gafrikova, T. Kowalczyk (1988), *Discriminant analysis based on experts' opinions*, Journal of Statistical Planning and Inference (praca przedłożona do publikacji).
- [2] P. W. Holland, P. R. Rosenbaum (1986), *Conditional association and unidimensionality in monotone latent variable models*, Annals of Statistics 14, 1523–1543.

- [3] T. Kowalczyk, E. Pleszczyńska, W. Szczesny, *Pomiar zależności stochastycznej*, w: *Teoria i Praktyka Wnioskowania Statystycznego*, pod red. T. Bromka i E. Pleszczyńskiej, PWN, Warszawa, 1988.
- [4] C.R.Rao, *Modele Liniowe Statystyki Matematycznej*, PWN, Warszawa, 1982.

Abstract

On inference concerning binary latent trait

Let Z_1 and Z_2 be observable random variables. Assume they depend on latent trait U and are conditionally independent, given U . 1) How, and to what extent, the joint distribution of (U, Z_1, Z_2) can be recovered from that of (Z_1, Z_2) ? 2) Suppose that, knowing Z_1 and/or Z_2 , we are to make decision concerning U . What decision rule is the best? Both the problems are properly formalized and solved in the simple case of binary U .