PRACE NAUKOWE AKADEMII EKONOMICZNEJ WE WROCŁAWIU
Nr 1123 — 2006
Ekonometria 17

Robert Kapłon

NIEJEDNORODNOŚĆ OBSERWACJI W MODELACH LOGITOWYCH

1. Wstęp

Jedną z pierwszych analiz wymagających potraktowania materiału statystycznego jako zbioru niejednorodnych obserwacji pochodzących z dwóch populacji przeprowadził w roku 1894 Karl Pearson. Zaproponował on model oparty na mieszance dwóch rozkładów normalnych, o różnych średnich i wariancjach. Takie ujęcie zagadnienia pozwoliło, w oparciu na danych dotyczących długości 1000 krabów, wyprowadzić wniosek o istnieniu dwóch podgatunków tych skorupiaków [9].

Chociaż upłynął wiek od propozycji Pearsona, to jednak stosunkowo niedawno, bo od pojawienia się pracy Dempstera, Laida i Rubina w 1978 r., metody uwzględniające heterogeniczność obserwacji, takie jak mieszanki rozkładów (*FMD*) i analiza klas ukrytych (*LCA*), zyskały duże zainteresowanie. Wspomniani autorzy przezwyciężyli problem estymacji parametrów metodą największej wiarygodności, wykorzystując koncepcję algorytmu EM. Tym samym możliwości aplikacyjne znacznie wzrosły i trudno wskazać obszar dziedziny naukowej, w której wskazane metody pozwalające rezygnować z założenia o jednorodności próby byłyby pomijane.

Należy odnotować, że mieszanki rozkładów i analiza klas ukrytych wyrosły na gruncie innych koncepcji. Jak wspomniano wcześniej, modelowanie na podstawie *FMD* miało na celu uwzględnienie faktu, że obserwacje pochodzą z różnych populacji. Istota *LCA* sprowadza się natomiast do badania związków między danymi dyskretnymi – a więc mierzonymi w skali nominalnej lub porządkowej. Zakłada się zatem, że między obserwowanymi zmiennymi zależności istnieją do czasu wprowadzenia tzw. dodatkowej zmiennej ukrytej. Jej obecność sprawia, że zbiór zmiennych jest już niezależny, a zmienna ukryta wyjaśnia istotę tych związków.

Ważną cechą charakterystyczną *LCA* – która czyni tę analizę podobną do analizy skupień – jest możliwość zidentyfikowania wzajemnie wyłączających się klas (por. [15]). To z kolei sprawia, że jest ona – podobnie jak *FMD* – często

wykorzystywaną metodą klasyfikacji, pozwalającą uwzględnić niejednorodność obserwacji. Czy wobec tego należałoby stwierdzić za Vermuntem [14], że „mieszanki rozkładów to inna nazwa dla analizy klas ukrytych"?

Wydaje się, że rozstrzygnięć należy poszukiwać w odpowiednich kryteriach – wyprowadzonych z istoty i koncepcji obu metod – pozwalających odróżnić te podejścia. W konsekwencji umożliwiłoby to określenie sytuacji, w których obie metody mają zastosowanie.

Właśnie taki cel, zorientowany na wskazanie podobieństw i różnic między metodami, przyświeca niniejszej pracy. W celu kompletności rozważań przedstawiono cały proces modelowania – od budowy modelu, poprzez estymację parametrów i weryfikację. Podstawą wyprowadzanych wniosków będą modele logitowe, które na gruncie badań marketingowych bardzo często wykorzystywane są do modelowania zachowań (decyzji wyboru) konsumentów.

2. Wielomianowy, warunkowy model logitowy

W modelach logitowych wykorzystuje się dwa typy obserwacji. Pierwszy ma charakter dyskretny i związany jest z wyborem badanych produktów. Rozważając w tym wypadku wektor losowy $\mathbf{Y}_i = \left[Y_{i1} \;\; Y_{i2} \;\; \ldots \;\; Y_{ij}\right]^{\mathrm{T}}$ oraz jego realizację postaci $\mathbf{y}_i^* = \left[y_{i1}^* \;\; y_{i2}^* \;\; \ldots \;\; y_{ij}^*\right]^{\mathrm{T}}$, przyjmuje się, że

$$y_{ij}^* = \begin{cases} 1, & \text{gdy konsument } i \text{ zakupi produkt } j, \\ 0, & \text{w przeciwnym wypadku.} \end{cases} \tag{1}$$

Obserwacje drugiego typu:

$$\mathbf{X}_i = \left[x_{ija}\right]_{j \times A},\ i = 1, 2, \ldots, N,\ j = 1, 2, \ldots, J,\ a = 1, 2, \ldots, A$$

są nielosowe i odpowiadają ocenie, jaką konsument i dał atrybutowi a marce j.

Sformułowanie modelu wymaga przyjęcia założenia, że użyteczność U_{ij}, jaką dla konsumenta i ma produkt j, można przedstawić jako addytywną formułę

$$U_{ij} = \boldsymbol{\beta}^{\mathrm{T}} \mathbf{x}_{ij} + \varepsilon_{ij}, \tag{2}$$

w której ε_{ij} jest składnikiem losowym reprezentującym tę część użyteczności, która nie została zaobserwowana; natomiast $\boldsymbol{\beta}$ jest wektorem parametrów. Wykorzystując informację empiryczną o realizacji zmiennej losowej Y_{ij}, definiuje się prawdopodobieństwo wyboru produktu j (por. [13])

$$\Pr(Y_{ij}=1)=\Pr(U_{ij})>\max_{m\neq j}[U_{im}]. \tag{3}$$

Ponadto, jeżeli $F=(\varepsilon_{i1},\varepsilon_{i2},\ldots,\varepsilon_{iJ})$ jest dystrybuantą oraz $F_1=\partial F/\partial\varepsilon_{i1}$, to można zdefiniować prawdopodobieństwo wyboru pierwszej marki

$$\Pr(Y_{ij}=1)=\int F_1(\varepsilon_{i1},\varepsilon_{i1}+\beta^{\mathrm{T}}\mathbf{x}_{i1}-\beta^{\mathrm{T}}\mathbf{x}_{i2},\ldots,\varepsilon_{i1}+\beta^{\mathrm{T}}\mathbf{x}_{i1}-\beta^{\mathrm{T}}\mathbf{x}_{iJ})d\varepsilon_{i1}. \tag{4}$$

Aby otrzymać wielomianowy model logitowy, przyjmuje się, że składniki losowe są niezależne, o jednakowym rozkładzie Gumbela. Biorąc funkcję gęstości tego rozkładu

$$f(\varepsilon_{ij})=e^{-\varepsilon_{ij}}\exp\left(-e^{-\varepsilon_{ij}}\right) \tag{5}$$

oraz wzór (4), otrzymuje się wielomianowy model logitowy:

$$\Pr(Y_{ij}=1)=\frac{\exp(\beta^{\mathrm{T}}\mathbf{x}_{ij})}{\sum_{m}\exp(\beta^{\mathrm{T}}\mathbf{x}_{im})},\quad m=1,2,\ldots,J. \tag{6}$$

3. Modele logitowe ze zmienną ukrytą

Koncepcja modeli ze zmiennymi ukrytymi pozwala na uwzględnienie heterogeniczności zachowań konsumentów. W wypadku zaprezentowanego modelu logitowego dotyczą one wyboru produktu oraz jego oceny (atrybutów). Z tymi zachowaniami związana jest zmienna ukryta Θ. Nie jest ona bezpośrednio obserwowalna, a więc nie posiadamy jej realizacji. Rola jej sprowadza się do podziału badanej zbiorowości na wzajemnie rozłączne i bezpośrednio nieobserwowalne klasy (segmenty). Po uwzględniieniu tej zmiennej funkcja rozkładu wektora $\mathbf{y}_i$ ma postać:

$$p(\mathbf{y}_i|\Psi,\mathbf{X}_i)=\int_{R_\Theta} p(\mathbf{y}_i|\Theta;\Psi^*,\mathbf{X}_i)dH_\pi(\Theta), \tag{7}$$

gdzie: $H_\pi(\Theta)$ jest miarą prawdopodobieństwa zmiennej ukrytej, natomiast Ψ i Ψ^* są wektorami parametrów. Taka reprezentacja pozwala uwzględnić zarówno ciągły, jak i dyskretny charakter tej zmiennej. Najczęściej jednak – zarówno w analizie klas ukrytych, jak i mieszankach rozkładów – przyjmuje się jej dyskretną naturę (por. [12]). Niech wobec tego zmienna ukryta Θ przyjmuje następujące wartości θ_1, θ_2, ..., θ_S odpowiednio z prawdopodobieństwami $\pi(\theta_1)$, $\pi(\theta_2)$, ..., $\pi(\theta_S)$. Prawdopodobieństwo przynależności do klasy s będzie więc równe:

$$\Pr(\Theta = \theta_s) = \pi(\theta_s),\ s = 1, 2, \ldots, S. \tag{8}$$

Po uwzględnieniu powyższych założeń wzór (7) można zapisać w postaci:

$$p(\mathbf{y}_i|\Psi, \mathbf{X}_i) = \sum_{s=1}^{S} \pi(\theta_s) p(\mathbf{y}_i|\Theta = \theta_s; \Psi_s, \mathbf{X}_i), \tag{9}$$

która w swojej ogólności wyraża ideę zarówno analizy klas ukrytych (*LCA*), jak i mieszanek rozkładów (*FMD*)[1]. Kluczowym obszarem zainteresowania jest więc rozstrzygnięcie, kiedy i w jakich warunkach można mówić o każdym z tych modeli.

Punktem wyjścia do dalszych rozważań jest przywołanie cech charakterystycznych dla analizy klas ukrytych. U jej podstaw leży dyskretny charakter zmiennej zależnej $\mathbf{Y}_i$ oraz zmiennej ukrytej Θ. Przyjęte wcześniej założenia, pozwalające zdefiniować prawdopodobieństwo (9), odnoszą się właśnie do dyskretnej natury zmiennych. Chociaż dla mieszanek rozkładów zmienna zależna zazwyczaj jest ciągła (por. [9]), to jednak w literaturze przedmiotu nie traktuje się tego jako warunku koniecznego. W konsekwencji dopuszcza się również zmienne dyskretne. To z kolei sprawia, że ta cecha (zmienne ciągłe a dyskretne) nie pozwala odróżnić *LCA* od *FMD*.

Fundamentalnym założeniem analizy klas ukrytych jest aksjomat o tzw. lokalnej niezależności. Występuje ona wtedy, kiedy między obserwowanymi zmiennymi zależności istnieją do czasu wprowadzenia tzw. dodatkowej zmiennej ukrytej. Jej obecność sprawia, że zbiór zmiennych jest już niezależny, a zmienna ukryta wyjaśnia istotę tych związków [5]. Formalnie zapisuje się to w postaci:

$$\Pr\left(\mathbf{Y}_i = \mathbf{y}_i^* \middle| \Theta = \theta_s; \Psi_s, \mathbf{X}_i\right) = \Pr\left(Y_{i1} = y_{i1}^* \middle| \Theta = \theta_s; \Psi_s, \mathbf{X}_i\right) \cdot \ldots \cdot \Pr\left(Y_{iJ} = y_{iJ}^* \middle| \Theta = \theta_s; \Psi_s, \mathbf{X}_i\right). \tag{10}$$

W kontekście decyzji konsumenckich w zakresie wyboru produktów założenie (10) implikuje możliwość niezależnego (warunkowego) wyboru kilku wyrobów, przy czym nie więcej niż J. Oznacza to, że analiza klas ukrytych nie może być podstawą do uwzględnienia niejednorodności obserwacji w sformułowanym wielomianowym modelu logitowy (6). Uwaga ta pozwala w istotny sposób dokonać rozróżnienia i sformułować taki model, w którym może być wykorzystana tylko koncepcja *FMD*.

Aby zbudować taki model, przyjęto założenie dotyczące liczby nabywanych produktów – konsument wybiera jeden wyrób spośród J rozważanych. Następnie określono strategię decyzyjną – konsument wybierze ten produkt, który w jego

[1] Przyjęte skróty pochodzą z języka angielskiego: *LCA – latent lass analysis*, *FMD – finite mixture distributions*.

osądzie będzie miał najwyższą użyteczność w segmencie s. Model ten, z prawdopodobieństwem warunkowym wyboru produktów, przedstawiono poniżej[2].

Model I. Wielomianowy model logitowy oparty na mieszankach rozkładów

Założenia:

$$\forall_{i=1,2,\ldots,N}\left(\sum_j y_{ij}^* = 1 \wedge y_{ij}^* \in \{0,1\}\right).$$

Strategia decyzyjna konsumenta i:

$$\Pr\left(U_{ij|s} > \max_{m\neq j}\left[U_{im|s}\right]\right) = \int_{-\infty}^{\infty} f\left(\varepsilon_{ij|s}\right)\prod_{\substack{m=1\\ m\neq j}}^{J} F\left(\varepsilon_{ij|s} + \beta_s^{\mathrm{T}}\mathbf{x}_{ij} - \beta_s^{\mathrm{T}}\mathbf{x}_{im}\right)d\varepsilon_{ij|s} = \frac{\exp\left(\beta_s^{\mathrm{T}}\mathbf{x}_{ij}\right)}{\sum_m \beta_s^{\mathrm{T}}\mathbf{x}_{im}}.$$

Prawdopodobieństwo warunkowe zaobserwowania wektora $\mathbf{y}_i$:

$$p\left(\mathbf{y}_i|\Theta = \theta_s; \mathsf{M}_1, \Psi_s, \mathbf{X}_i\right) = \prod_{j=1}^{J}\left[P_i^s\left(j; \mathsf{M}_1\right)\right]^{y_{ij}^*};\ P_i^s\left(j; \mathsf{M}_1\right) = \frac{\exp\left(\beta_s^{\mathrm{T}}\mathbf{x}_{ij}\right)}{\sum_m \beta_s^{\mathrm{T}}\mathbf{x}_{im}}.$$

Kontynuując wątek dotyczący liczby nabywanych produktów, warto nadmienić, że dość często konsumenci decydują się na wybór kilku marek lub, co bardziej prawdopodobne, dokonują tego w pewnym przedziale czasowym. Ta sytuacja dotyczy przede wszystkim dóbr konsumpcyjnych szybko zbywalnych. Przykładowo w jednym z badań respondenci mieli wybrać tę markę kawy, którą zakupiliby. Zaznaczono, że mogą zdecydować się na wybór więcej niż jednej z nich. Okazało się, decyzja o zakupie tylko jednej marki kawy była rzadka; najczęściej wskazywano na dwie lub trzy [6].

Zamodelowanie tej sytuacji wymaga wykorzystania innej strategii decyzyjnej niż ta, którą uwzględniono w pierwszym modelu. Można więc przyjąć, że jeżeli użyteczność produktu w segmencie s przekroczy pewien próg $\Delta_s = \delta_s + \varepsilon_s$, to taki produkt zostanie wybrany. Część deterministyczną reprezentuje δ_s, a losową – ε_s o rozkładzie Gumela (5).

Konsekwencją przyjętej strategii decyzyjnej będzie, po pierwsze, prawdopodobieństwo wyboru marki oparte na regresji logistycznej, a nie wielomianowym modelu logitowym. Po drugie, dokonywane wybory między produktami są niezależne, co jest równoważne z przyjęciem aksjomatu o lokalnej niezależności (10). Należałoby więc sądzić, że *LCA* będzie odpowiednia do uwzględnienia heterogeniczności zachowań konsumentów. Aby to rozstrzygnąć, sformułowano następujący model.

[2] Prezentowane modele będą oznaczone przez M_l – dla modelu l (l =1, 2, 3).

Model II. Model logitowy oparty na mieszankach rozkładów
Założenia:

$$\forall_{i=1,2,\ldots,N}\left(0 \le \sum_j y_{ij}^* \le J \wedge y_{ij}^* \in \{0,1\}\right).$$

Strategia decyzyjna konsumenta i:

$$\Pr\left(U_{ij|s} > \Delta_s\right) = \int_{-\infty}^{\infty} f\left(\varepsilon_{ij|s}\right) F\left(\varepsilon_{ij|s} + \beta_s^{\mathrm{T}} \mathbf{x}_{ij} - \delta_s\right) d\varepsilon_{ij|s} = \left[1 + \exp\left(-\beta_s^{\mathrm{T}} \mathbf{x}_{ij}\right)\right]^{-1}$$

i niech $\forall_{i,j} x_{ij0} = -1$, $\beta_{0s} = \delta_s$.

Prawdopodobieństwo warunkowe zaobserwowania wektora $\mathbf{y}_i$:

$$p\left(\mathbf{y}_i | \Theta = \theta_s; \mathsf{M}_2, \Psi_s, \mathbf{X}_i\right) = \prod_{j=1}^{J}\left[P_i^s\left(j; \mathsf{M}_2\right)\right]^{y_{ij}^*} \prod_{j=1}^{J}\left[1 - P_i^s\left(j; \mathsf{M}_2\right)\right]^{1-y_{ij}^*},$$

$$P_i^s\left(j; \mathsf{M}_2\right) = \left[1 + \exp\left(-\beta_s^{\mathrm{T}} \mathbf{x}_{ij}\right)\right]^{-1}.$$

Należy wyrazić przekonanie, że i w tym wypadku model opiera się na mieszankach rozkładów. Potwierdzeniem słuszności tej tezy jest pogląd, według którego „cechą wspólną modeli klas ukrytych jest założenie, że każda osoba w klasie ma identyczne prawdopodobieństwo zaobserwowania zmiennej na odpowiednim poziomie, tzn. osoby nie różnią się wewnątrz klasy" [11, s. 28]. Bezpośrednią implikacją tego stwierdzenia jest wymóg, aby prawdopodobieństwa wyboru marki w zadanej klasie były identyczne dla wszystkich konsumentów. Jak łatwo spostrzec, nie jest tak ze względu na różne oceny atrybutów x_{ija} dokonywane przez konsumentów. Dopiero przyjęcie zagregowanej oceny atrybutów spowoduje, że rozważane prawdopodobieństwa warunkowe będą identyczne dla wszystkich konsumentów. To z kolei pozwoli wykorzystać analizę klas ukrytych.

Model III. Model logitowy oparty na analizie klas ukrytych
Założenia:

$$\forall_{i,i^*=1,2,\ldots,N}\left(0 \le \sum_j y_{ij}^* \le J \wedge y_{ij}^* \in \{0,1\} \wedge \mathbf{x}_{ij}\mathbf{x}_{i^*j}\right).$$

Strategia decyzyjna konsumenta i:

$$\Pr\left(U_{ij|s} > \Delta_s\right) = \int_{-\infty}^{\infty} f\left(\varepsilon_{ij|s}\right) F\left(\varepsilon_{ij|s} + \beta_s^{\mathrm{T}} \mathbf{x}_{ij} - \delta_s\right) d\varepsilon_{ij|s} = \left[1 + \exp\left(-\beta_s^{\mathrm{T}} \mathbf{x}_{ij}\right)\right]^{-1}$$

i niech $\forall_{i,j} x_{ij0} = -1$, $\beta_{0s} = \delta_s$.

Prawdopodobieństwo warunkowe zaobserwowania wektora $\mathbf{y}_i$:

$$p(\mathbf{y}_i|\Theta = \theta_s; \mathsf{M}_3, \Psi_s, \mathbf{X}_i) = \prod_{j=1}^{J}\left[P_i^s(j; \mathsf{M}_3)\right]^{y_{ij}^*} \prod_{j=1}^{J}\left[1 - P_i^s(j; \mathsf{M}_3)\right]^{1-y_{ij}^*},$$

$$P_i^s(j; \mathsf{M}_3) = \left[1 + \exp\left(-\beta_s^{\mathrm{T}} \mathbf{x}_{ij}\right)\right]^{-1}.$$

4. Estymacja parametrów modeli

W wypadku analizy klas ukrytych i mieszanek rozkładów do estymacji parametrów podchodzi się w identyczny sposób: wykorzystując najczęściej algorytm EM. Odwołując się do niego (por. [3]), funkcję wiarygodności

$$L(\Psi|\mathbf{Y}; \mathbf{X}, \mathsf{M}_l) = \prod_{i=1}^{N}\sum_{s=1}^{S}\pi(\theta_s)p(\mathbf{y}_i|\theta_s; \Psi_s, \mathbf{X}_i, \mathsf{M}_l) \tag{11}$$

traktuje się jako tę, która opiera się na niekompletnym zbiorze danych, dlatego wprowadza się dodatkową zmienną ukrytą

$$z_{is} = \begin{cases} 1, & \text{gdy konsument } i \in C_s, \\ 0, & \text{w przeciwnym wypadku,} \end{cases} \tag{12}$$

o wielomianowym rozkładzie. Wtedy dla kompletnego zbioru danych

$$(\mathbf{y}_1^*\mathbf{z}_1), (\mathbf{y}_2^*\mathbf{z}_2), \ldots, (\mathbf{y}_i^*\mathbf{z}_i), \ldots, (\mathbf{y}_N^*\mathbf{z}_N),$$

funkcja wiarygodności ma postać:

$$\log L_c(\Psi|\mathbf{Y}, \mathbf{Z}; \mathbf{X}, \mathsf{M}_l) = \sum_{i=1}^{N}\sum_{s=1}^{S} z_{is} \log \pi(\theta_s) + \sum_{i=1}^{N}\sum_{s=1}^{S} z_{is} \log p(\mathbf{y}_i|\theta_s; \Psi_s, \mathbf{X}_i, \mathsf{M}_l). \tag{13}$$

Funkcja (13) staje się podstawą do wykorzystania algorytmu EM, na który składają się dwa kroki:

Krok E: W iteracji $(k + 1)$ obliczyć $Q(\Psi|\Psi^{(k)})$, przyjmując za nieznany wektor parametrów Ψ początkowe wartości równe $\Psi^{(k)}$, gdzie

$$Q(\Psi|\Psi^{(k)}) = E_{\Psi^{(k)}}\left[\log L_c(\Psi|\mathbf{Y}, \mathbf{Z}; \mathbf{X}, \mathsf{M}_l)\middle|\mathbf{Y}, \Psi^{(k)}\right]. \tag{14}$$

Jeżeli się przyjmie za

$$E_{\Psi^{(k)}}\left[Z_{is}|\mathbf{Y}\right]=\kappa_{is}^{(k)},$$

to łatwo pokazać, że

$$\kappa_{is}^{(k)}=\frac{\pi(\theta_s)p\left(\mathbf{y}_i|\theta_s;\Psi_s^{(k)},\mathbf{X}_i,\mathbf{M}_l\right)}{\sum_s \pi(\theta_s)p\left(\mathbf{y}_i|\theta_s;\Psi_s^{(k)},\mathbf{X}_i,\mathbf{M}_l\right)}.$$

Biorąc to pod uwagę, funkcja (14) przekształca się do następującej postaci:

$$Q\left(\Psi|\Psi^{(k)}\right)=\sum_{i=1}^{N}\sum_{s=1}^{S}\kappa_{is}^{(k)}\log\pi(\theta_s)+\sum_{i=1}^{N}\sum_{s=1}^{S}\kappa_{is}^{(k)}\log p\left(\mathbf{y}_i|\theta_s;\Psi_s,\mathbf{X}_i,\mathbf{M}_l\right) \quad (15)$$

Krok M: Należy maksymalizować wartość oczekiwaną, która została obliczona w poprzednim kroku E, czyli

$$\Psi^{(k+1)}=\arg\max_{\Psi} Q\left(\Psi|\Psi^{(k)}\right).$$

W wypadku estymatorów prawdopodobieństwa przynależności do klas możliwe jest podanie ich postaci analitycznej. Odwołując się do (15), należy obliczyć odpowiednie pochodne cząstkowe. Po przyrównaniu ich do zera otrzymuje się następujące oszacowanie:

$$\hat{\pi}(\theta_s)=\frac{1}{N}\sum_{i=1}^{N}\kappa_{is}^{(k)}.$$

Pozostałe parametry można oszacować, wykorzystując w tym celu jeden z algorytmów optymalizacji nieliniowej.

5. Kryteria weryfikacji

Otrzymany model należy poddać weryfikacji statystycznej. Wśród najczęściej wykorzystywanych testów, mających zastosowanie w modelach opartych na zarówno LCA, jak i na FMD, wyróżnia się test Walda i test oparty na ilorazie wiarygodności.

Niech $H(\Psi)$ będzie macierzą Hessianu. Sprawdzeniem hipotezy $H_0: \Psi = \Psi_0$ w teście Walda jest następującą statystyka

$$W=\left(\hat{\Psi}-\Psi_0\right)^{\mathrm{T}}H\left(\hat{\Psi}\right)\left(\hat{\Psi}-\Psi_0\right), \quad (16)$$

która asymptotycznie zbiega do rozkładu χ^2 z K stopniami swobody, gdzie K jest liczbą estymowanych parametrów. Zaletą tego testu jest możliwość łącznego

wnioskowania o parametrach, bez konieczności przeprowadzania dodatkowych estymacji. Tej cechy brakuje testowi opartemu na ilorazie wiarygodności, jednak w porównaniu z testem Walda jest bardziej dokładny w wypadku prób o małym i średnim rozmiarze. Co więcej, zgodnie z badaniami W.W. Haucka i A. Donnera [4], gdy różnica między wartością parametru estymowanego a parametru przyjętego powiększa się, wtedy wartość statystyki W bardzo szybko zbiega do zera; również maleje jego moc. Tych niepożądanych własności nie ma test oparty na ilorazie wiarygodności[3]. Przyjęta statystyka testowa

$$LR = -2\left[\log L\left(\Psi_0 \middle| \mathbf{Y}; \mathbf{X}, \mathbf{M}_l\right) - \log L\left(\hat{\Psi} \middle| \mathbf{Y}; \mathbf{X}, \mathbf{M}_l\right)\right] \tag{17}$$

ma również asymptotyczny rozkład χ^2 z liczbą stopni swobody równą liczbie nałożonych ograniczeń na parametry.

Przedstawione statystyki testowe (16) i (17) pozwalają podjąć decyzję dotyczącą tylko wyboru między dwoma modelami hierarchicznymi. Bardziej dokładne rozstrzygnięcia – zmierzające do ustalenia, czy proponowany model jest dobrze dopasowany do danych empirycznych – możliwe są dzięki statystykom Pearsona, G^2 i Cressiego-Reada. Oznaczając odpowiednio przez $\hat{f}_v$ i f_v częstości oczekiwane i zaobserwowane, statystyki Pearsona i G^2 mają następującą postać:

$$\chi^2 = \sum_{v=1}^{2^J} \left(f_v - \hat{f}_v\right)^2 \Big/ \hat{f}_v ,$$

$$G^2 = 2\sum_{v=1}^{2^J} f_v \log\left(f_v / \hat{f}_v\right),$$

aczkolwiek tylko w wypadku modeli opartych na *LCA* można je wykorzystać. Tym samym, rozważając kolejny etap modelowania, jakim jest weryfikacja modelu, nie można mówić o *FMD* jako ogólniejszym podejściu w stosunku do *LCA*.

Nasuwa się również pytanie dotyczące tego, która ze statystyk (χ^2 czy G^2) jest lepszym miernikiem oceny dopasowania modelu. Wątpliwości rodzą się wtedy, gdy różnica ich wartości jest znaczna. Okazuje się, że jeżeli występują małe oczekiwane częstotliwości, to wartości obu statystyk mogą być wypaczone, powodując te różnice. Z tego powodu dla małych częstotliwości oczekiwanych proponuje się statystykę Cressiego-Reada o postaci:

$$CR(\lambda) = 2\left[\lambda(\lambda+1)\right]^{-1} \sum_{v=1}^{2^J} f_v \left[\left(f_v / \hat{f}_v\right)^{\lambda} - 1\right].$$

[3] Wspomniani autorzy przeprowadzili badania z wykorzystaniem modelu logitowego.

W zależności od przyjętych wartości λ otrzymuje się statystykę G^2 i Pearsona odpowiednio dla $\lambda = 0$ (w granicy) i $\lambda = 1$. N. Cressie i T.R.C. Read proponują, aby przyjąć $\lambda = 2/3$.

W modelach opartych zarówno na mieszankach rozkładów, jak i analizie klas ukrytych pojawia się problem związany z rozstrzygnięciem, który z dwóch modeli o różnej liczbie klas lepiej dopasowany jest do danych. Pośrednio można to rozwiązać, odwołując się do kryteriów informacji [2; 7]. W tym celu wykorzystuje się kryterium informacji Akaikego (*AIC*)

$$AIC = -2 \log L\left(\hat{\Psi} \middle| \mathbf{Y}; \mathbf{X}, \mathsf{M}_l\right) + 2K.$$

Zostaje wybrany ten model, dla którego wartość *AIC* jest najmniejsza. Alternatywnym podejściem jest użycie statystyki opartej na ilorazie wiarygodności *LR*. Wtedy dla modelu o *df* stopniach swobody

$$AIC^* = -LR - 2df\,.$$

Warto zauważyć, że dla dwóch porównywanych modeli, dla których logarytm funkcji wiarygodności różni się nieznacznie, na podstawie kryterium *AIC* wybiera się model z mniejszą liczbą estymowanych parametrów.

Kolejnym wykorzystywanym miernikiem jest kryterium informacyjne *BIC* (*Bayesian information criteria*). Ma ono, w odróżnieniu od *AIC*, własności asymptotyczne, tzn. uwzględnia wielkość próby. Podobnie jak w kryterium informacyjnym Akaikego, *BIC* można wyrazić w dwojaki sposób:

$$BIC = -2 \log L\left(\hat{\Psi} \middle| \mathbf{Y}; \mathbf{X}, \mathsf{M}_l\right) + K \log N\,,$$

$$BIC^* = LR - df \log N\,.$$

Należy odnotować, że *AIC* ma tendencję do wyboru modeli, które są nazbyt rozbudowane, nawet wtedy, gdy próba jest duża. Z kolei *BIC* preferuje modele mniej złożone, tzn. z mniejszą liczbą parametrów.

6. Porównanie metod

Przedstawiona problematyka uwzględnienia heterogeniczności w modelach logitowych daje podstawę, by sądzić, że *FMD* jest podejściem bardziej elastycznym od *LCA*, a co więcej, podejściem, które skupia w sobie *LCA*. Chociaż formalne rozstrzygnięcia (por. wzór (9)) przemawiają za takim wnioskiem, to jednak koncepcje, na których zrodziły się te metody, były odmienne. Jak wspomniano we wstępie, mieszanki rozkładów zostały wykorzystane przez wzgląd na istnienie obserwacji pochodzących z różnych populacji. Od samego początku były więc przeznaczone

do uwzględnienia niejednorodności obserwacji. Z kolei analiza klas ukrytych pozwala na badanie zależności, jakie występują między zmiennymi mierzonymi na skalach słabych. Istota tych związków jest pozorna, gdyż istnieje do momentu wprowadzenia zmiennej ukrytej. Jej obecność sprawia, że zbiór zmiennych jest już niezależny, a zmienna ukryta wyjaśnia te związki.

Tym samym pierwotne przeznaczenie *LCA* było inne niż *FMD*. Dopiero później, ze względu na podobieństwo do analizy skupień, analiza klas ukrytych coraz częściej była wykorzystywana jako metoda klasyfikacji (por. VerMag). W odniesieniu do marketingu mówi się o metodzie segmentacji, pozwalającej otrzymać segmenty (klasy) ukryte. Określenie to (segmenty ukryte) jest używane również w stosunku do modeli, które nie są oparte na *LCA*, lecz na mieszankach rozkładów – tak jak w sformułowanym modelu pierwszym i drugim. Przykładem mogą być prace [10; 16]. Daje to podstawę, by sądzić, że nastąpiło przeniesienie terminologiczne. O jego słuszności (lub niesłuszności) trudno rozprawiać bez dokładnej definicji zmiennej ukrytej. Choć intuicyjnie jest ona wyczuwana, niemniej jednak w literaturze pojawia się wiele propozycji. Ze względu na swoją obszerność zagadnienie to nie będzie rozważane w niniejszym opracowaniu.

Tabela 1. Porównanie analizy klas ukrytych i mieszanek rozkładów

Punkt widzenia		Modele oparte na	
statystyki	marketingu	*LCA*	*FMD*
zmienna objaśniana (dyskretna)	wybór produktów	kilka marek	jedna marka
zmienne objaśniające	ocena atrybutów	identyczna dla każdego konsumenta	różni się między konsumentami
prawdopodobieństwa warunkowe	szanse wyboru produktów przez konsumentów należących do tej samej klasy	takie same	różne
model logitowy	strategia decyzyjna	regresja logistyczna	warunkowy, wielomianowy model logitowy
kryteria weryfikacji	kryteria weryfikacji	χ^2 Pearsona, G^2, *CR*, *LR*, *W*, *AIC*, *AIC**, *BIC* i *BIC**	*LR*, *W*, *AIC*, *AIC**, *BIC* i *BIC**

Źródło: opracowanie własne.

Przyjmując wcześniej przyjęty podział (modele oparte na *LCA* i *FMD*), w tab. 1 przedstawiono porównanie obu podejść. Przyjęte kryteria stanowią punkt widzenia statystyki i marketingu.

Literatura

[1] Bozdogan H., *Model Selection and Akaike's Information Criterion: The General Theory and its Analytical Extensions*, „Psychometrika" 1987, vol. 52, nr 3, s. 345-370.

[2] Dayton C.M., *Latent Class Scaling Analysis*, Sage University Papers Series on Quantitative Applications in the Social Sciences, 07-126, Thousand Oaks, CA Sage1998.

[3] Dempster A.P., Laird N.M., Rubin D.B., *Maximum Likelihood from Incomplete Data via the EM Algorithm*, „Journal of the Royal Statistical Society" 1977, ser. B, nr 1(39), s. 1-22.

[4] Hauck Jr. W.W., Donner A., *Wald's Test as Applied to Hypothesis in Logit Analysis*, „Journal of the American Statistical Association", 1977, 72(360), s. 851-853.

[5] Kapłon R., *Analiza danych dyskretnych za pomocą metody LCA*, [w:] *Taksonomia 9*, Prace Naukowe Akademii Ekonomicznej we Wrocławiu nr 942, AE, Wrocław 2002.

[6] Kapłon R., *Mapy pozycjonowania wyrobów przy wykorzystaniu analizy czynnikowej klas ukrytych*, [w:] *Taksonomia 10*, Prace Naukowe Akademii Ekonomicznej we Wrocławiu, AE, Wrocław 2003.

[7] Kass R.E., Raftery A.E., *Bayes Factors*, „Journal of the American Statistical Association" 1995, vol. 90, s. 773-793.

[8] McFadden D., *Conditional Logit Analysis of Qualitative Choice Behavior*, [w:] *Frontiers in Econometrics*, red. P. Zarembka, Academic Press, New York 1974.

[9] McLachlan G.J., Peel. D., *Finite Mixture Models*, Wiley, New York 2000.

[10] Oh M.S., Choi J.W., Kim D.G., *Bayesian Inference And model Selection in Latent Class Logit Models with Parameter Constraints: An Application to Market Segmentation*, „Journal of Applied Statistics" 2003, vol. 30, nr 2, s. 191-204.

[11] Rost J., Langeheine R., *A Guide Through Latent Structure Models For Categorical Data*, [w:] *Applications Of Latent Trait And Latent Class Models In The Social Science*, red. J. Rost, R. Langeheine, Waxmann, Berlin 1997.

[12] Titterington D.M., Smith A.F.M., Markov U.E., *Statistical Analysis of Finite Mixture Distributions*, John Wiley & Son, New York 1985.

[13] Train K.E., *Qualitative Choice Analysistr*, MIT Press, Cambridge 1986.

[14] Vermunt J.K., *Finite Mixture Model*, [w:] *Encyclopedia of Research Methods for the Social Sciences*, red. M. Lewis-Beck, A. Bryman, T.F. Liao, Sage Publications, New Bury Park 2004.

[15] Vermunt J.K., Magidson J., *Latent Class Cluster Analysis*, [w:] *Applied Latent Class Models*, red. J. Hagenaars, A. McCutcheon, Cambridge University Press, Cambridge 2002.

[16] Zenor M.J., Srivastava R.K., *Inferring Market Structure with Aggregate Data: A Latent Segment Logit Approach*, „Journal of Marketing Research" 1993, vol. 30, Issue 3, s. 369-379.

HETEROGENEITY IN LOGIT MODELS

Summary

In order to model heterogeneity in the logit models one can use latent class analysis as well as finite mixture distribution. Frequently, these two names are used interchangeably. Taking into account the fact that these methods were originally meant for different goals, the aim of this paper is to compare these methods – pointing out similarities and differences.

Dr inż. Robert Kaplon jest pracownikiem Instytutu Organizacji i Zarządzania Politechniki Wrocławskiej.