
Warunkowa wielowymiarowa identyfikacja wartości parametrów przy niesymetrycznych i skorelowanych stratach błędów estymacji

Piotr Kulczycki¹, Małgorzata Charytanowicz²

¹ Politechnika Krakowska, Katedra Automatyki i Technik Informatycznych

Instytut Badań Systemowych PAN, Centrum Informatycznych Metod Analizy Danych, kulczycki@ibspan.waw.pl

² Katolicki Uniwersytet Lubelski, Instytut Matematyki i Informatyki

Instytut Badań Systemowych PAN, Centrum Informatycznych Metod Analizy Danych, mchmat@ibspan.waw.pl

Wiele współczesnych wyspecjalizowanych zadań badawczych wymaga uwzględnienia wpływu nieuniknionych w praktyce błędów identyfikacji parametrów, występujących w modelu. Pozytywne wyniki może również przynieść dokonanie pomiaru, a następnie uwzględnienie konkretnych (np. aktualnych) wartości tych czynników, które w istotny sposób oddziałują na wartości owych parametrów. Niniejsza praca jest poświęcona zagadnieniu estymacji wektora parametrów, gdy straty wynikłe z ich niedoszacowania i przeszacowania są niesymetryczne i wzajemnie skorelowane. Zagadnienie rozpatrywane jest z dodatkowym aspektem warunkowym, przy czym poszczególne współrzędne zmiennych warunkujących mogą być ciągłe, binarne, dyskretne lub kategoriowe (uporządkowane i nieuporządkowane). Końcowym wynikiem jest kompletny algorytm wyznaczania wartości estymatora optymalnego w sensie minimum wartości oczekiwanej strat, z użyciem wielowymiarowej niesymetrycznej funkcji kwadratowej, przy dowolnym rozkładzie zmiennych opisujących i warunkujących.

1. Wstęp

Właściwa identyfikacja (estymacja) wartości parametrów występujących w modelu zawsze miała podstawowe znaczenie w nietrywialnych zagadnieniach nauki i problemach aplikacyjnych [20]. We współczesnych zadaniach poszczególne parametry nierzadko przestały jednak opisywać konkretne wielkości fizyczne, lecz raczej reprezentują cały zakres złożonych zjawisk, upraszczanych w ramach modelu do jednego parametru, istniejącego jedynie w sensie formalnym. W tej sytuacji ich identyfikacja nie może być rozumiana w klasycznym sensie [13], jako wyznaczanie wartości estymatora możliwie jak najbliższej wyimaginowanych „prawdziwych” aczkolwiek nieznanymi wartości parametrów (gdyż takie w ogóle nie istnieją), lecz bardziej poprzez odpowiednie uwzględnienie skutków implikowanych przez poszczególne wartości na działanie rozważanego systemu. W szczególności, w wielu współczesnych zadaniach aplikacyjnych, dodatnie i ujemne błędy estymacji mogą skutkować

różnymi następstwami, zarówno w sensie ilościowym jak i jakościowym. Znaczenie tego aspektu rośnie w miarę jak zwiększa się złożoność i wyrafinowanie współczesnych modeli oraz wymagania stawiane im przez zadania, dla których są konstruowane [11].

Potrzeba uwzględnienia odmiennych implikacji dodatnich i ujemnych błędów estymacji (tj. odpowiednio przeszacowania oraz niedoszacowania badanej wielkości) prowadzi wprost do koncepcji niesymetrycznej postaci funkcji strat [2]. Istotność takiego problemu została wcześniej zauważona i rozważona dla prostych przypadków pojedynczego parametru (por. klasyczny artykuł [23]). Ciekawe porównanie podstawowych typów stosowanych wówczas funkcji strat, a to: niesymetrycznej liniowej, niesymetrycznej kwadratowej oraz liniowo-eksponencjalnej LINEX (w przybliżeniu liniowej dla ujemnych błędów estymacji oraz eksponencjalnej dla dodatnich albo odwrotnie) można znaleźć w artykule [15]. Należy także nadmienić o badaniach dotyczących estymacji pojedynczego parametru z wielomianową niesymetryczną funkcją strat [12], przedstawionych w artykule [9] dodatkowo w wersji warunkowej, to znaczy gdy szacowana wielkość jest istotnie zależna od czynników warunkujących. W praktyce inżynierskiej, czynnik taki stanowi często aktualna temperatura. Jeśli konkretne wartości tego typu czynników są metrologicznie dostępne, to uwzględnienie ich może znacząco uściślić stosowany model. W niniejszej pracy badania te zostaną uogólnione na przypadek wielowymiarowy, gdy identyfikowanych jest kilka parametrów, traktowanych jako wektor, a straty wynikłe z ich przeszacowania i niedoszacowania mogą być niesymetryczne i skorelowane.

Prezentowana tu koncepcja opiera się na ujęciu bayesowskim, co umożliwia minimalizację wartości oczekiwanej strat wynikłych z nieuniknionych w praktyce błędów estymacji.

Do wyznaczenia charakterystyk probabilistycznych użyto nieparametrycznej metody estymatorów jądrowych, co uniezależnia opracowaną procedurę od postaci rozkładów charakteryzujących zarówno identyfikowane parametry, jak i czynniki warunkujące.

Podsumowując, cel niniejszej publikacji stanowi przedstawienie algorytmu pozwalającego wyznaczyć wartości estymowanych parametrów, optymalne w sensie minimum wartości oczekiwanej strat, gdy straty wynikłe z dodatnich i ujemnych błędów estymacji są różne i dodatkowo skorelowane pomiędzy poszczególnymi parametrami. Procedura opracowana jest w ujęciu warunkowym, co pozwala uściślić wynik o ustaloną (np. aktualną) wartość czynników warunkujących. Rozkłady zarówno estymowanych parametrów, jak i tych czynników, mogą być dowolne.

Pełny opis algorytmu jest przedmiotem publikacji [10], gdzie podany został w kompletnej postaci, tzn. wraz z cytowanym materiałem bibliograficznym może być on bezpośrednio użyty, bez szczegółowej znajomości teoretycznych aspektów, żmudnych badań i analitycznych obliczeń.

2. Preliminaria matematyczne: statystyczne estymatory jądrowe

Niech dana będzie n -wymiarowa zmienna losowa X , o rozkładzie posiadającym gęstość f . Jej estymator jądrowy $\hat{f} : \mathbb{R}^n \rightarrow [0, \infty)$, wyznacza się na podstawie eksperymentalnie uzyskanych wartości m -elementowej próby losowej

$$x_1, x_2, \dots, x_m \quad (1)$$

i w podstawowej postaci definiowany jest jako

$$\hat{f}(x) = \frac{1}{mh^n} \sum_{i=1}^m K\left(\frac{x - x_i}{h}\right), \quad (2)$$

gdzie $m \in \mathbb{N} \setminus \{0\}$, współczynnik $h > 0$ zwany jest parametrem wygładzania, natomiast mierzalną funkcję $K : \mathbb{R}^n \rightarrow [0, \infty)$ spełniającą warunek $\int_{\mathbb{R}^n} K(x) dx = 1$, symetryczną względem zera i mającą w tym punkcie słabe maksimum globalne, określa się mianem jądra.

W praktyce, wybór postaci jądra nie ma istotnego znaczenia z punktu widzenia statystyki i dzięki temu możliwe jest uwzględnienie przede wszystkim własności otrzymanego estymatora, korzystnych

dla konkretnego zastosowania. Szczegółowe rozważania można znaleźć w publikacjach [7] – sekcja 3.1.3, [21] – sekcje 2.7 i 4.5. W prezentowanej metodzie stosowane będzie jednowymiarowe jądro Cauchy’ego

$$K(x) = \frac{2}{\pi} \frac{1}{(1+x^2)^2}, \quad (3)$$

w przypadku n -wymiarowym uogólnione do jądra produktowego

$$K(x) = K \left(\begin{bmatrix} x_1 \\ x_2 \\ \vdots \\ x_n \end{bmatrix} \right) = K_1(x_1)K_2(x_2) \dots K_n(x_n), \quad (4)$$

gdzie K_i oznaczają omawiane wcześniej jednowymiarowe jądro (3), przy czym jako występujące w podstawowym wzorze (2) wyrażenie h^n należy teraz przyjąć iloczyn parametrów wygładzania $h_1 \cdot h_2 \cdot \dots \cdot h_n$, określonych dla poszczególnych współrzędnych.

Duże znaczenie dla jakości estymacji ma natomiast wyznaczenie wartości parametru wygładzania h . Szczęśliwie – z punktu widzenia praktyki – opracowano wiele procedur dogodnych do obliczania tej wartości na podstawie posiadanej próby losowej (1). Ich przegląd znaleźć można w monografiach [7], [17], [21]. W szczególności dla jednowymiarowego przypadku dogodną okazuje się metoda podstawień, inaczej *plug-in* [7] – sekcja 3.1.5, [21] – sekcja 3.6.1. Metoda ta może być stosowana także w przypadku n -wymiarowym, gdy używane jest jądro produktowe (4), odrębnie wobec każdej współrzędnej.

W praktyce stosuje się również dodatkowe procedury generalnie poprawiające jakość estymatora (2), jak też dopasowujące jego własności do szczegółowych uwarunkowań problemu. Dla metody przedstawionej w niniejszym opracowaniu, szczególnie polecana jest modyfikacja parametru wygładzania, indywidualizująca jego wartość dla poszczególnych jąder [7] – sekcja 3.1.6, [17] – sekcja 5.3.1.

Powyższa koncepcja zostanie uogólniona na przypadek warunkowy. Poza podstawową (określaną czasem mianem opisującej) n_Y -wymiarową zmienną losową Y , niech dana będzie także n_W -wymiarowa zmienna losowa W , zwana dalej warunkującą zmienną losową. Ich zestawienie $X = \begin{bmatrix} Y \\ W \end{bmatrix}$ jest zmienną losową o wymiarze $n_Y + n_W$. Załóżmy, że rozkłady zmiennej X i w konsekwencji również W , posiadają gęstości, oznaczane poniżej odpowiednio jako $f_X : \mathbb{R}^{n_Y+n_W} \rightarrow [0, \infty)$ oraz $f_W : \mathbb{R}^{n_W} \rightarrow [0, \infty)$. Niech także dana będzie tzw. wartość warunkująca, czyli ustalona wartość warunkującej zmiennej losowej, $w^* \in \mathbb{R}^{n_W}$ taka iż

$$f_W(w^*) > 0. \quad (5)$$

Wtedy funkcja $f_{Y|W=w^*} : \mathbb{R}^{n_Y} \rightarrow [0, \infty)$ określona wzorem

$$f_{Y|W=w^*}(y) = \frac{f_X(y, w^*)}{f_W(w^*)} \quad \text{dla każdego } y \in \mathbb{R}^{n_Y} \quad (6)$$

stanowi warunkową gęstość prawdopodobieństwa zmiennej losowej Y przy wartości warunkującej w^* . Warunkową gęstość prawdopodobieństwa $f_{Y|W=w^*}$ można zatem traktować jako „klasyczną” gęstość, której postać została w zagadnieniu praktycznym uściślona poprzez podanie w^* – konkretnej wartości, jaką w danej sytuacji przyjęła zmienna warunkująca W .

Założmy zatem, iż dysponuje się próbą losową

$$\begin{bmatrix} y_1 \\ w_1 \end{bmatrix}, \begin{bmatrix} y_2 \\ w_2 \end{bmatrix}, \dots, \begin{bmatrix} y_m \\ w_m \end{bmatrix}, \quad (7)$$

pozyskaną ze zmiennej $X = \begin{bmatrix} Y \\ W \end{bmatrix}$. Jej poszczególne elementy interpretowane są jako wartości y_i uzyskane poprzez pomiar zmiennej losowej Y , gdy zmienna warunkująca W przyjmowała odpowiednio wartości w_i . Korzystając z metodyki przedstawionej w początkowej części niniejszej sekcji, na podstawie próby (7) wyznaczyć wtedy można \hat{f}_X , czyli jądrowy estymator gęstości prawdopodobieństwa zmiennej losowej X , natomiast na podstawie próby

$$w_1, w_2, \dots, w_m \quad (8)$$

uzyskuje się \hat{f}_W – jądrowy estymator gęstości zmiennej warunkującej W . Jądrowy estymator warunkowej gęstości prawdopodobieństwa zmiennej losowej Y przy wartości warunkującej w^* , definiowany jest wtedy – w naturalnej konsekwencji formuły (6) – jako funkcja $\hat{f}_{Y|W=w^*} : \mathbb{R}^{n_Y} \rightarrow [0, \infty)$ dana wzorem

$$\hat{f}_{Y|W=w^*}(y) = \frac{\hat{f}_X(y, w^*)}{\hat{f}_W(w^*)}. \quad (9)$$

W przypadku stosowania przy konstrukcji estymatora \hat{f}_W jądra o dodatnich wartościach, nierówność $\hat{f}_W(w^*) > 0$ implikowana przez warunek (5) jest spełniona dla dowolnego $w^* \in \mathbb{R}^{n_W}$.

Jeżeli do konstrukcji estymatorów \hat{f}_X i \hat{f}_W stosuje się jądro produktowe (4), używając w estymatorze \hat{f}_X wobec współrzędnych które odpowiadają wektorowi W oraz w estymatorze \hat{f}_W parami tych samych dodatnich jąder, to formuła jądrowego estymatora warunkowej gęstości prawdopodobieństwa staje się szczególnie dogodna dla praktycznych zastosowań. Wzór (9) można wówczas zapisać w postaci

$$\hat{f}_{Y|W=w^*}(y) = \hat{f}_{Y|W=w^*} \left(\begin{bmatrix} y_1 \\ y_2 \\ \vdots \\ y_{n_Y} \end{bmatrix} \right) = \quad (10)$$

$$= \frac{\frac{1}{h_1 h_2 \dots h_{n_Y}} \sum_{i=1}^m K_1 \left(\frac{y_1 - y_{i,1}}{h_1} \right) \dots K_{n_Y} \left(\frac{y_{n_Y} - y_{i,n_Y}}{h_{n_Y}} \right) K_{n_Y+1} \left(\frac{w_1^* - w_{i,1}}{h_{n_Y+1}} \right) \dots K_{n_Y+n_W} \left(\frac{w_{n_W}^* - w_{i,n_W}}{h_{n_Y+n_W}} \right)}{\sum_{i=1}^m K_{n_Y+1} \left(\frac{w_1^* - w_{i,1}}{h_{n_Y+1}} \right) \dots K_{n_Y+n_W} \left(\frac{w_{n_W}^* - w_{i,n_W}}{h_{n_Y+n_W}} \right)},$$

przy czym $h_1, h_2, \dots, h_{n_Y+n_W}$ reprezentują odpowiednio parametry wygładzania przyporządkowane kolejnym współrzędnym zmiennej losowej X , a poszczególne współrzędne wektorów w^*, y_i i w_i oznaczone zostały jako

$$w^* = \begin{bmatrix} w_1^* \\ w_2^* \\ \vdots \\ w_{n_W}^* \end{bmatrix} \quad \text{oraz} \quad y_i = \begin{bmatrix} y_{i,1} \\ y_{i,2} \\ \vdots \\ y_{i,n_Y} \end{bmatrix}, \quad w_i = \begin{bmatrix} w_{i,1} \\ w_{i,2} \\ \vdots \\ w_{i,n_W} \end{bmatrix} \quad \text{dla } i = 1, 2, \dots, m. \quad (11)$$

Zdefiniujmy tzw. parametry warunkujące d_i ($i = 1, 2, \dots, m$) następującym wzorem:

$$d_i = K_{n_Y+1} \left(\frac{w_1^* - w_{i,1}}{h_{n_Y+1}} \right) K_{n_Y+2} \left(\frac{w_2^* - w_{i,2}}{h_{n_Y+2}} \right) \dots K_{n_Y+n_W} \left(\frac{w_{n_W}^* - w_{i,n_W}}{h_{n_Y+n_W}} \right). \quad (12)$$

Dzięki założeniu o dodatniej wartości jąder $K_{n_Y+1}, K_{n_Y+2}, \dots, K_{n_Y+n_W}$, parametry te są także dodatnie. Wówczas jądrowy estymator warunkowej gęstości prawdopodobieństwa (10) można przed-

stawić w postaci

$$\hat{f}_{Y|W=w^*}(y) = \hat{f}_{Y|W=w^*} \left(\begin{bmatrix} y_1 \\ y_2 \\ \vdots \\ y_{n_Y} \end{bmatrix} \right) = \quad (13)$$

$$= \frac{1}{h_1 h_2 \dots h_{n_Y} \sum_{i=1}^m d_i} \sum_{i=1}^m d_i K_1 \left(\frac{y_1 - y_{i,1}}{h_1} \right) K_2 \left(\frac{y_2 - y_{i,2}}{h_2} \right) \dots K_{n_Y} \left(\frac{y_{n_Y} - y_{i,n_Y}}{h_{n_Y}} \right).$$

Wartość parametru d_i charakteryzuje „odległość” danej wartości warunkującej w^* od wartości zmiennej warunkującej dla której pozyskany został i -ty element próby losowej, czyli od w_i . Wówczas estymator (13) można zinterpretować jako kombinację liniową jąder przyporządkowanych poszczególnym elementom próby losowej otrzymanej dla zmiennej Y , przy czym współczynniki powyższej kombinacji charakteryzują reprezentatywność tych elementów dla ustalonej wartości w^* .

Szczegółowy materiał dotyczący estymatorów jądrowych dostępny jest w monografiach [7], [17], [18], [21].

3. Algorytm

Zapiszmy parametry, których wartości podlegają estymacji, w postaci wektora $y \in \mathbb{R}^{n_Y}$. Będą one traktowane jako wartości n_Y -wymiarowej zmiennej losowej Y . Niech także dana będzie n_W -wymiarowa warunkująca zmienna losowa W . Zakłada się również dostępność pozyskanych metrologicznie pomiarów wektora parametrów y , tzn. $y_1, y_2, \dots, y_m \in \mathbb{R}^{n_Y}$, otrzymanych – odpowiednio – dla wartości $w_1, w_2, \dots, w_m \in \mathbb{R}^{n_W}$ czynnika warunkującego. I wreszcie, niech $w^* \in \mathbb{R}^{n_W}$ oznacza ustaloną wartość warunkującą. Celem pracy jest wyznaczenie estymatora wektora parametrów, oznaczanego jako $\hat{y}_{w^*} \in \mathbb{R}^{n_Y}$, optymalnego w sensie minimum wartości oczekiwanej strat wynikłych z błędów estymacji, dla wartości warunkującej w^* . W celu rozwiązania niniejszego problemu zostanie użyta bayesowska reguła decyzyjna [2]. Dla przejrzystości prezentacji, przedstawiony będzie przypadek dwuwymiarowy ($n_Y = 2$). Sama idea może być jednak z powodzeniem przeniesiona na wyższe wymiary, aczkolwiek kosztem naturalnego w takich sytuacjach powiększenia złożoności problemu. Warto jednak zauważyć, że w praktycznych przypadkach korelacje błędów estymacji nie muszą występować pomiędzy wszystkimi parametrami, więc stosowana formuła może być znacząco uproszczona wobec ogólnej postaci, w zależności od konkretnych uwarunkowań zadania.

Niech zatem szacowane parametry traktowane będą jako dwuwymiarowy wektor $\begin{bmatrix} y_1 \\ y_2 \end{bmatrix}$, natomiast ich estymatory oznaczmy przez $\begin{bmatrix} \hat{y}_1 \\ \hat{y}_2 \end{bmatrix}$. Funkcję strat $l : \mathbb{R}^2 \times \mathbb{R}^2 \rightarrow \mathbb{R}$, która zgodnie z założeniami teorii decyzji [2] określa straty występujące gdy przyjęto wartość $\begin{bmatrix} \hat{y}_1 \\ \hat{y}_2 \end{bmatrix}$, podczas gdy w rzeczywistości hipotetyczny stan wynosił $\begin{bmatrix} y_1 \\ y_2 \end{bmatrix}$, zdefiniowana zostanie w następującej kwadratowej

i niesymetrycznej postaci:

$$l\left(\begin{bmatrix} \hat{y}_1 \\ \hat{y}_2 \end{bmatrix}, \begin{bmatrix} y_1 \\ y_2 \end{bmatrix}\right) = \begin{cases} a_l(\hat{y}_1 - y_1)^2 + a_{ld}(\hat{y}_1 - y_1)(\hat{y}_2 - y_2) + a_d(\hat{y}_2 - y_2)^2 \\ \text{gdy } \hat{y}_1 - y_1 \leq 0 \text{ oraz } \hat{y}_2 - y_2 \leq 0 \\ a_r(\hat{y}_1 - y_1)^2 + a_{rd}(\hat{y}_1 - y_1)(\hat{y}_2 - y_2) + a_d(\hat{y}_2 - y_2)^2 \\ \text{gdy } \hat{y}_1 - y_1 \geq 0 \text{ oraz } \hat{y}_2 - y_2 \leq 0 \\ a_l(\hat{y}_1 - y_1)^2 + a_{lu}(\hat{y}_1 - y_1)(\hat{y}_2 - y_2) + a_u(\hat{y}_2 - y_2)^2 \\ \text{gdy } \hat{y}_1 - y_1 \leq 0 \text{ oraz } \hat{y}_2 - y_2 \geq 0 \\ a_r(\hat{y}_1 - y_1)^2 + a_{ru}(\hat{y}_1 - y_1)(\hat{y}_2 - y_2) + a_u(\hat{y}_2 - y_2)^2 \\ \text{gdy } \hat{y}_1 - y_1 \geq 0 \text{ oraz } \hat{y}_2 - y_2 \geq 0 \end{cases}, \quad (14)$$

gdzie $a_l, a_r, a_u, a_d > 0, a_{ld}, a_{ru} \geq 0$ oraz $a_{lu}, a_{rd} \leq 0$. Współczynnik a_l reprezentuje straty spowodowane niedoszacowaniem pierwszego parametru; (zauważmy bowiem, że jest on mnożony przez czynnik $(\hat{y}_1 - y_1)^2$ dotyczący błędu estymacji pierwszego parametru, a także iż występuje w tych opcjach w których obowiązuje warunek $\hat{y}_1 - y_1 \leq 0$ stanowiący o jego niedoszacowaniu). Analogicznie współczynnik a_p związany jest ze stratami spowodowanymi przeszacowaniem tego parametru (jest mnożony przez $(\hat{y}_1 - y_1)^2$ oraz występuje gdy $\hat{y}_1 - y_1 \geq 0$, co stanowi o przeszacowaniu). Analogicznie a_d oraz a_u związane są ze stratami wynikłymi z niedoszacowania i przeszacowania drugiego parametru. Z kolei, współczynniki $a_{ld}, a_{ru}, a_{lu}, a_{rd}$ reprezentują wzajemną korelację błędów estymacji obu parametrów. I tak kolejno: współczynnik a_{ld} charakteryzuje dodatkowe straty wynikłe z jednoczesnego niedoszacowania obu parametrów, a_{ru} z jednoczesnego ich przeszacowania, natomiast a_{lu} z niedoszacowania pierwszego i przeszacowania drugiego, a a_{rd} odwrotnie – z przeszacowania pierwszego i niedoszacowania drugiego. Warto również zauważyć, że w przypadku $a_{ld} = a_{rd} = a_{lu} = a_{ru} = 0$, co oznacza iż błędy wynikłe z estymacji obu parametrów są nieskorelowane, problem redukuje się do dwóch odrębnych jednowymiarowych kwadratowych zagadnień opracowanych w sekcji 3.2 artykułu [9]. Graficzna ilustracja postaci funkcji l widoczna jest na rysunku 1, w sekcji 4, wraz z przykładowym wynikiem symulacji numerycznej.

Załóżmy warunkową niezależność [3] estymowanych parametrów. Wówczas warunkowa gęstość $f_{Y|W=w^*}$ reprezentująca ich nieokreśloność może być zapisana jako iloczyn jednowymiarowych gęstości $f_{Y_1|W=w^*} : \mathbb{R} \rightarrow [0, \infty)$ oraz $f_{Y_2|W=w^*} : \mathbb{R} \rightarrow [0, \infty)$ odpowiadających poszczególnym składowym, czyli

$$f_{Y|W=w^*}(y_1, y_2) = f_{Y_1|W=w^*}(y_1)f_{Y_2|W=w^*}(y_2). \quad (15)$$

Przyjmuje się także założenia, że funkcje $f_{Y_1|W=w^*}$ i $f_{Y_2|W=w^*}$ są ciągłe, a także iż spełnione są warunki $\int_{-\infty}^{\infty} y_1 f_{Y_1|W=w^*}(y_1) dy_1 < \infty$ oraz $\int_{-\infty}^{\infty} y_2 f_{Y_2|W=w^*}(y_2) dy_2 < \infty$.

Szczegółowa analiza [10] wskazuje, że dla funkcji strat (14) minimum strat istnieje, jest jedyne i występuje dla wektora $\begin{bmatrix} \hat{y}_1 \\ \hat{y}_2 \end{bmatrix}$ będącego rozwiązaniem następującego układu równań:

$$\begin{aligned} & \int_{-\infty}^{\hat{y}_1} f_{Y_1|W=w^*}(y_1) dy_1 \left[(a_{ru} - a_{rd} - a_{lu} + a_{ld}) \left(\hat{y}_2 \int_{-\infty}^{\hat{y}_2} f_{Y_2|W=w^*}(y_2) dy_2 - \int_{-\infty}^{\hat{y}_2} y_2 f_{Y_2|W=w^*}(y_2) dy_2 \right) + \right. \\ & \quad \left. + (a_{rd} - a_{ld}) \left(\hat{y}_2 - \int_{-\infty}^{\infty} y_2 f_{Y_2|W=w^*}(y_2) dy_2 \right) \right] + 2a_l \left(\hat{y}_1 - \int_{-\infty}^{\infty} y_1 f_{Y_1|W=w^*}(y_1) dy_1 \right) + \\ & + 2(a_r - a_l) \left(\hat{y}_1 \int_{-\infty}^{\hat{y}_1} f_{Y_1|W=w^*}(y_1) dy_1 - \int_{-\infty}^{\hat{y}_1} y_1 f_{Y_1|W=w^*}(y_1) dy_1 \right) + a_{ld} \left(\hat{y}_2 - \int_{-\infty}^{\infty} y_2 f_{Y_2|W=w^*}(y_2) dy_2 \right) + \\ & \quad + (a_{lu} - a_{ld}) \left(\hat{y}_2 \int_{-\infty}^{\hat{y}_2} f_{Y_2|W=w^*}(y_2) dy_2 - \int_{-\infty}^{\hat{y}_2} y_2 f_{Y_2|W=w^*}(y_2) dy_2 \right) = 0 \end{aligned} \quad (16)$$

$$\begin{aligned}
& \int_{-\infty}^{\hat{y}_2} f_{Y_2|W=w^*}(y_2) dy_2 \left[(a_{ru} - a_{rd} - a_{lu} + a_{ld}) \left(\hat{y}_1 \int_{-\infty}^{\hat{y}_1} f_{Y_1|W=w^*}(y_1) dy_1 - \int_{-\infty}^{\hat{y}_1} y_1 f_{Y_1|W=w^*}(y_1) dy_1 \right) + \right. \\
& \quad \left. + (a_{lu} - a_{ld}) \left(\hat{y}_1 - \int_{-\infty}^{\hat{y}_1} y_1 f_{Y_1|W=w^*}(y_1) dy_1 \right) \right] + 2a_d \left(\hat{y}_2 - \int_{-\infty}^{\hat{y}_2} y_2 f_{Y_2|W=w^*}(y_2) dy_2 \right) + \\
& + 2(a_u - a_d) \left(\hat{y}_2 \int_{-\infty}^{\hat{y}_2} f_{Y_2|W=w^*}(y_2) dy_2 - \int_{-\infty}^{\hat{y}_2} y_2 f_{Y_2|W=w^*}(y_2) dy_2 \right) + a_{ld} \left(\hat{y}_1 - \int_{-\infty}^{\hat{y}_1} y_1 f_{Y_1|W=w^*}(y_1) dy_1 \right) + \\
& \quad + (a_{rd} - a_{ld}) \left(\hat{y}_1 \int_{-\infty}^{\hat{y}_1} f_{Y_1|W=w^*}(y_1) dy_1 - \int_{-\infty}^{\hat{y}_1} y_1 f_{Y_1|W=w^*}(y_1) dy_1 \right) = 0. \tag{17}
\end{aligned}$$

Wyznaczenie jego wartości nie jest wszakże w ogólnym przypadku zadaniem łatwym. Jednak jeżeli estymacja występujących powyżej gęstości jest odpowiednio dokonywana z użyciem opisanej w sekcji 2 metodyki statystycznych estymatorów jądrowych, to możliwe jest skonstruowanie do tego celu efektywnego algorytmu numerycznego.

I tak, dla każdego ustalonego $i = 1, 2, \dots, m$ zdefiniujmy funkcje $U_{1,i} : \mathbb{R} \rightarrow \mathbb{R}$, $U_{2,i} : \mathbb{R} \rightarrow \mathbb{R}$, $V_{1,i} : \mathbb{R} \rightarrow \mathbb{R}$ oraz $V_{2,i} : \mathbb{R} \rightarrow \mathbb{R}$, dane jako

$$U_{1,i}(\hat{y}_1) = \frac{1}{h_1} \int_{-\infty}^{\hat{y}_1} K\left(\frac{y_1 - y_{i,1}}{h_1}\right) dy_1 \tag{18}$$

$$U_{2,i}(\hat{y}_2) = \frac{1}{h_2} \int_{-\infty}^{\hat{y}_2} K\left(\frac{y_2 - y_{i,2}}{h_2}\right) dy_2 \tag{19}$$

$$V_{1,i}(\hat{y}_1) = \frac{1}{h_1} \int_{-\infty}^{\hat{y}_1} y_1 K\left(\frac{y_1 - y_{i,1}}{h_1}\right) dy_1 \tag{20}$$

$$V_{2,i}(\hat{y}_2) = \frac{1}{h_2} \int_{-\infty}^{\hat{y}_2} y_2 K\left(\frac{y_2 - y_{i,2}}{h_2}\right) dy_2. \tag{21}$$

Unormujmy także parametry warunkujące d_i , definiując dodatnie wielkości

$$d_i^* = \frac{d_i}{\sum_{i=1}^m d_i} \quad \text{dla każdego } i = 1, 2, \dots, m; \tag{22}$$

zauważmy iż

$$\sum_{i=1}^m d_i^* = 1. \tag{23}$$

Po wprowadzeniu powyższych notacji, uwzględniając postać jądrowego estymatora warunkowej zmiennej losowej (13), kryteria (16)-(17) mogą być zapisane w równoważnej postaci

$$\begin{aligned}
& \sum_{i=1}^m d_i^* U_{1,i}(\hat{y}_1) \left[(a_{ru} - a_{rd} - a_{lu} + a_{ld}) \sum_{i=1}^m d_i^* (\hat{y}_2 U_{2,i}(\hat{y}_2) - V_{2,i}(\hat{y}_2)) + (a_{rd} - a_{ld}) \left(\hat{y}_2 - \sum_{i=1}^m d_i^* y_{i,2} \right) \right] + \\
& + 2a_l \left(\hat{y}_1 - \sum_{i=1}^m d_i^* y_{i,1} \right) + 2(a_r - a_l) \sum_{i=1}^m d_i^* (\hat{y}_1 U_{1,i}(\hat{y}_1) - V_{1,i}(\hat{y}_1)) + a_{ld} \left(\hat{y}_2 - \sum_{i=1}^m d_i^* y_{i,2} \right) +
\end{aligned}$$

$$+(a_{lu} - a_{ld}) \sum_{i=1}^m d_i^* (\hat{y}_2 U_{2,i}(\hat{y}_2) - V_{2,i}(\hat{y}_2)) = 0 \quad (24)$$

$$\begin{aligned} & \sum_{i=1}^m d_i^* U_{2,i}(\hat{y}_2) \left[(a_{ru} - a_{rd} - a_{lu} + a_{ld}) \sum_{i=1}^m d_i^* (\hat{y}_1 U_{1,i}(\hat{y}_1) - V_{1,i}(\hat{y}_1)) + (a_{lu} - a_{ld}) (\hat{y}_1 - \sum_{i=1}^m d_i^* y_{i,1}) \right] + \\ & + 2a_d (\hat{y}_2 - \sum_{i=1}^m d_i^* y_{i,2}) + 2(a_u - a_d) \sum_{i=1}^m d_i^* (\hat{y}_2 U_{2,i}(\hat{y}_2) - V_{2,i}(\hat{y}_2)) + a_{ld} (\hat{y}_1 - \sum_{i=1}^m d_i^* y_{i,1}) + \\ & + (a_{rd} - a_{ld}) \sum_{i=1}^m d_i^* (\hat{y}_1 U_{1,i}(\hat{y}_1) - V_{1,i}(\hat{y}_1)) = 0. \end{aligned} \quad (25)$$

Oznaczając lewe strony powyższych równań jako $L_1(\hat{y}_1, \hat{y}_2)$ oraz $L_2(\hat{y}_1, \hat{y}_2)$, otrzymujemy następujące pochodne cząstkowe:

$$\begin{aligned} \frac{\partial L_1(\hat{y}_1, \hat{y}_2)}{\partial \hat{y}_1} &= \sum_{i=1}^m d_i^* \frac{1}{h_1 s_{i,1}} K\left(\frac{y_1 - y_{i,1}}{h_1 s_{i,1}}\right) \left[(a_{ru} - a_{rd} - a_{lu} + a_{ld}) \sum_{i=1}^m d_i^* (\hat{y}_2 U_{2,i}(\hat{y}_2) - V_{2,i}(\hat{y}_2)) + \right. \\ & \left. + (a_{rd} - a_{ld}) (\hat{y}_2 - \sum_{i=1}^m d_i^* y_{i,2}) \right] + 2(a_r - a_l) \sum_{i=1}^m d_i^* U_{1,i}(\hat{y}_1) + 2a_l \end{aligned} \quad (26)$$

$$\begin{aligned} \frac{\partial L_1(\hat{y}_1, \hat{y}_2)}{\partial \hat{y}_2} &= \sum_{i=1}^m d_i^* U_{1,i}(\hat{y}_1) \left[(a_{ru} - a_{rd} - a_{lu} + a_{ld}) \sum_{i=1}^m d_i^* U_{2,i}(\hat{y}_2) + (a_{rd} - a_{ld}) \right] + \\ & + (a_{lu} - a_{ld}) \sum_{i=1}^m d_i^* U_{2,i}(\hat{y}_2) + a_{ld} \end{aligned} \quad (27)$$

$$\begin{aligned} \frac{\partial L_2(\hat{y}_1, \hat{y}_2)}{\partial \hat{y}_1} &= \sum_{i=1}^m d_i^* U_{2,i}(\hat{y}_2) \left[(a_{ru} - a_{rd} - a_{lu} + a_{ld}) \sum_{i=1}^m d_i^* U_{1,i}(\hat{y}_1) + (a_{lu} - a_{ld}) \right] + \\ & + (a_{rd} - a_{ld}) \sum_{i=1}^m d_i^* U_{1,i}(\hat{y}_1) + a_{ld} \end{aligned} \quad (28)$$

$$\begin{aligned} \frac{\partial L_2(\hat{y}_1, \hat{y}_2)}{\partial \hat{y}_2} &= \sum_{i=1}^m d_i^* \frac{1}{h_2 s_{i,2}} K\left(\frac{\hat{y}_2 - y_{i,2}}{h_2 s_{i,2}}\right) \left[(a_{ru} - a_{rd} - a_{lu} + a_{ld}) \sum_{i=1}^m d_i^* (\hat{y}_1 U_{1,i}(\hat{y}_1) - V_{1,i}(\hat{y}_1)) + \right. \\ & \left. + (a_{lu} - a_{ld}) (\hat{y}_1 - \sum_{i=1}^m d_i^* y_{i,1}) \right] + 2(a_u - a_d) \sum_{i=1}^m d_i^* U_{2,i}(\hat{y}_2) + 2a_d. \end{aligned} \quad (29)$$

Wówczas rozwiązanie równań (16)-(17) może być efektywnie wyznaczone za pomocą wielowymiarowego algorytmu Newtona [19] jako granica dwuwymiarowego ciągu $\left\{ \begin{matrix} \hat{y}_{j,1} \\ \hat{y}_{j,2} \end{matrix} \right\}_{j=0}^{\infty}$ zdefiniowanego wzorami

$$\hat{y}_{0,1} = \frac{\sum_{i=1}^m d_i y_{i,1}}{\sum_{i=1}^m d_i} \quad (30)$$

$$\hat{y}_{0,2} = \frac{\sum_{i=1}^m d_i y_{i,2}}{\sum_{i=1}^m d_i} \quad (31)$$

$$\begin{bmatrix} \hat{y}_{j+1,1} \\ \hat{y}_{j+1,2} \end{bmatrix} = \begin{bmatrix} \hat{y}_{j,1} \\ \hat{y}_{j,2} \end{bmatrix} - \begin{bmatrix} \frac{\partial L_1(\hat{y}_{j,1}, \hat{y}_{j,2})}{\partial \hat{y}_1} & \frac{\partial L_1(\hat{y}_{j,1}, \hat{y}_{j,2})}{\partial \hat{y}_2} \\ \frac{\partial L_2(\hat{y}_{j,1}, \hat{y}_{j,2})}{\partial \hat{y}_1} & \frac{\partial L_2(\hat{y}_{j,1}, \hat{y}_{j,2})}{\partial \hat{y}_2} \end{bmatrix}^{-1} \begin{bmatrix} L_1(\hat{y}_{j,1}, \hat{y}_{j,2}) \\ L_2(\hat{y}_{j,1}, \hat{y}_{j,2}) \end{bmatrix} \quad \text{dla } j = 0, 1, \dots, \quad (32)$$

przy czym wszystkie wielkości występujące w powyższych zależnościach zostały podane powyżej, natomiast warunek stopu przyjmuje postać koniunkcji następujących nierówności:

$$|\hat{y}_{j,1} - \hat{y}_{j-1,1}| \leq 0.01\hat{\sigma}_1 \quad (33)$$

$$|\hat{y}_{j,2} - \hat{y}_{j-1,2}| \leq 0.01\hat{\sigma}_2, \quad (34)$$

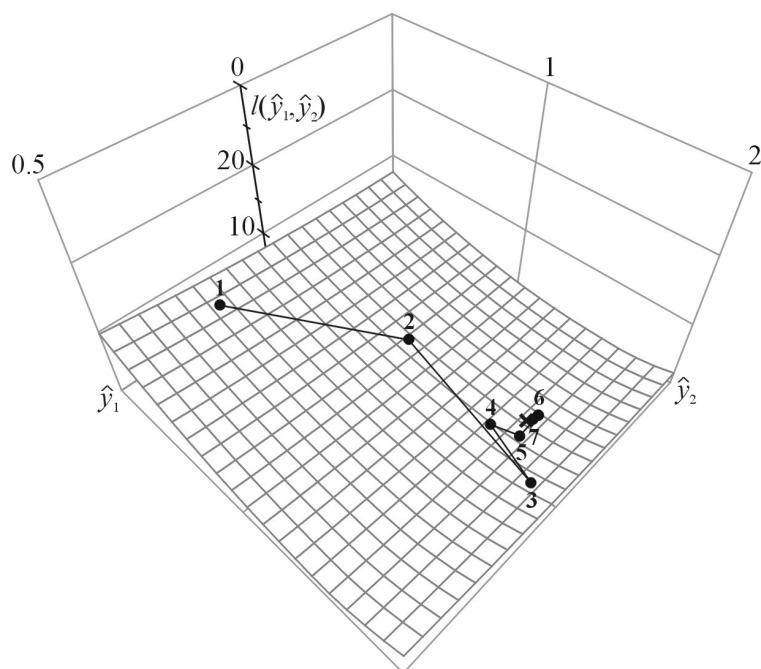
gdzie $\hat{\sigma}_1$ oraz $\hat{\sigma}_2$ oznaczają estymatory odchyłeń standardowych dla poszczególnych współrzędnych wektora Y .

4. Wnioski końcowe i posumowanie

W niniejszej pracy przedstawiono algorytm do wyznaczania warunkowej wartości wektora parametrów, gdy straty wynikłe z dodatnich i ujemnych błędów estymacji są niesymetryczne i wzajemnie skorelowane. Ujęcie warunkowe pozwala w praktyce na uściślenie modelu poprzez uwzględnienie odpowiedniej (aktualnej) wartości warunkujących czynników. Użycie bayesowskiej reguły decyzyjnej zapewnia minimum wartości oczekiwanej strat, natomiast metodyka statystycznych estymatorów jądrowych uniezależnia procedurę od postaci rozkładów zmiennych opisujących i warunkujących. Przedstawiona koncepcja ma charakter uniwersalny i może być stosowana do szerokiego zakresu zagadnień nauki oraz inżynierii, ekonomii i zarządzania, nauk społecznych, biomedycyny i innych pokrewnych dyscyplin.

Poprawne działanie algorytmu zostało potwierdzone ze pomocą licznych badań numerycznych przy danych sztucznie generowanych dla celów ilustracyjnych, a także przebiegów symulacyjnych oraz aplikacji do zadań praktycznych z zakresu inżynierii sterowania i medycyny. Na rysunku 1 pokazano przykładową graficzną ilustrację funkcji strat (14), z zaznaczonymi kolejnymi krokami uzyskiwanymi z użyciem algorytmu Newtona (30)-(32), zbieżnymi do wartości estymatora minimalizującego wartość oczekiwaną strat.

Przede wszystkim, generalną regułą były przesunięcia wartości wyznaczanego estymatora w kierunkach związanych z mniejszymi stratami wynikłymi z błędów estymacji, określanymi funkcją strat (14). Otóż, zwiększanie wartości parametru a_l względem a_r , a w konsekwencji powiększenie wartości tej funkcji dla dodatnich błędów estymacji pierwszego parametru, czyli jego przeszacowania, skutkowało zwiększaniem wartości wyznaczanego estymatora tego parametru. W konsekwencji redukuje to prawdopodobieństwo owego przeszacowania. Odwrotny efekt występuje przy zwiększaniu wartości parametru a_r względem a_l : wartość wyznaczanego estymatora ulegała pomniejszeniu, co redukuje prawdopodobieństwo niedoszacowania. Im iloraz a_l/a_r bardziej różnił się od jedności, tym powyższe efekty były wyrazistsze. Analogiczne zależności występowały dla parametrów a_d oraz a_u przy estymacji wartości drugiego parametru. Z kolei zwiększenie wartości parametru a_{ld} skutkowało równomiernym powiększeniem wartości estymatorów obu parametrów, redukując w obu przypadkach prawdopodobieństwo przeszacowania. Przeciwne skutki implikowały zmiany parametru a_{ru} . I wreszcie, zwiększenie wartości bezwzględnej parametru a_{rd} redukuje prawdopodobieństwo niedoszacowania pierwszego parametru i przeszacowania drugiego, poprzez pomniejszenie wartości wyznaczanego estymatora dla pierwszego i powiększenie dla drugiego z nich. Przeciwne zjawiska dotyczą parametru a_{lu} . Warto dodać, że znaczenie mają tu nie same wartości poszczególnych parametrów, lecz relacje (proporcje) między nimi. O wielkości zmiany wartości estymatorów stanowią rozkłady reprezentowane przez gęstości $f_{Y_1|W=w^*}$ oraz $f_{Y_2|W=w^*}$.



Rysunek 1: Funkcja strat (14) i przebieg algorytmu Newtona (30)-(32); liczbami 1-7 oznaczono kolejne kroki, natomiast symbol \times wskazuje położenie poszukiwanej wartości optymalnej estymatora.

Ujęcie warunkowe skutkowało odpowiednią korektą wartości estymatora, zgodnie z charakterem korelacji między zmiennymi opisującymi i warunkującymi. Jeżeli parametr był dodatnio skorelowany z czynnikiem warunkującym, to zwiększenie/zmniejszenie wartości warunkującej powodowało powiększenie/pomniejszenie wartości estymatora tego parametru. Odwrotne efekty występowały przy ujemnej korelacji. Zależność ta może być bardziej złożona, zgodnie z potencjalnie dowolną formą zależności gęstości warunkowych $f_{Y_1|W=w^*}$ oraz $f_{Y_2|W=w^*}$ od wartości warunkujących w^* .

Akceptowalną jakość rozwiązań uzyskiwano już dla liczebności próby rzędu 50-100 gdy wartość warunkująca ulokowana była w okolicy głównej wartości modalnej zmiennej warunkującej, 100-200 w odległości odchylenia standardowego oraz 200-500 na drugim odchyleniu standardowym. Biorąc pod uwagę złożony wielowymiarowy charakter zagadnienia, w praktyce nie wydaje się to nadmiernym wymaganiem. Dzięki uśredniającym własnościom estymatorów jądrowych, algorytm okazał się odporny na nikłą liczbę lub wręcz brak danych z otoczenia wartości warunkującej. I wreszcie, w miarę wzrostu liczebności próby losowej wyznaczone wartości estymatora zmierzały do wartości teoretycznej, a odchylenie standardowe błędu sukcesywnie malało do zera. Powyższe asymptotyczne własności mają istotne znaczenie z aplikacyjnego punktu widzenia, gdyż najczęściej umożliwiają uzyskanie żądanej dokładności, aczkolwiek pod warunkiem posiadania dostatecznej liczebności próby losowej, jak również możliwości obliczeniowych. W praktyce stanowi to o konieczności kompromisu dotyczącego powyższych aspektów.

Procedura została tu przedstawiona w swej podstawowej wersji, łatwej do zaimplementowania i dogodnej obliczeniowo. Obrazowa interpretacja umożliwia jednak wprowadzanie indywidualnych modyfikacji oraz generalizacji, dostosowanych do nietypowych uwarunkowań poszczególnych zagadnień aplikacyjnych. Przede wszystkim możliwe jest uwzględnienie zmiennych warunkujących odmiennych od ciągłych (rzeczywistych), a to: binarnych [7] – sekcja 3.1.8, [17] – sekcja 6.1.4, dyskretnych [1], [22], kategoriowych (także uporządkowanych) [5], [14], [16], jak też różnorodnych ich złożań. Ta własność powinna być szczególnie podkreślona, biorąc pod uwagę specyfikę współczesnej analizy i eksploracji danych, coraz częściej wykorzystującej różne konfiguracje atrybutów odmiennych rodzajów. Także metoda Newtona może być zastosowana w jednej ze swych licznych modyfikacji [4], [6], używanych

zwłaszcza w celu zmniejszenia liczby iteracji lub powiększenia obszaru zbieżności.

Ważną cechą przedstawionego algorytmu jest możliwość jego dekompozycji na dwie fazy. Pierwsza zawiera procedury wyznaczania parametrów wygładzania i jego modyfikacji. Złożoność obliczeniowa wynosi tu $O((n_Y + n_W)m^2)$, ale obliczenia te mają miejsce tylko jeden raz na początku obliczeń. Drugą fazę stanowi bezpośrednie wyznaczenie wartości estymatora. Złożoność obliczeniowa jest tu liniowa względem m, n_Y oraz n_W , a algorytm Newtona najczęściej wymagał 10-15 iteracji. Implikuje to mały czas obliczeń, w większości praktycznych zadań pozwalając na pracę w czasie rzeczywistym. Uzyskano to w dużym stopniu dzięki odpowiedniemu doborowi postaci jądra (3), umożliwiającemu uzyskanie analitycznych postaci funkcji (18)-(21).

Szczegółowy opis metodyki zaprezentowanej w niniejszej pracy oraz badań weryfikacyjnych zawarty jest w artykule [10], będącym obecnie w fazie publikowania. Sam algorytm został tam przedłożony w zakresie kompletnym – wraz z materiałem zawartym w cytowanej literaturze może być bezpośrednio stosowany bez szczegółowej wiedzy przedmiotowej oraz dodatkowych badań.

I wreszcie na koniec pomniejsza osobista uwaga. Prezentując przypadek jednowymiarowy (wielomianowy) na poprzedniej konferencji KKA'2011 [8] napisaliśmy, że koncepcja ta „może być uogólniona na przypadek wielowymiarowy, czyli gdy identyfikowany jest wektor parametrów warunkowo skorelowanych. Jednak wówczas zarówno same analityczne kryteria na optymalne wartości parametrów, jak i ich późniejsza numeryczna implementacja, stają się skomplikowane w stopniu przewyższającym obecne możliwości ich praktycznego stosowania”. Na szczęście, po wprowadzeniu założenia dotyczącego warunkowej niezależności identyfikowanych parametrów, przy jednoczesnym zachowaniu korelacji strat wywołanych błędami ich estymacji, powyższe stwierdzenie okazało się zbyt pesymistyczne. Co więcej, przeprowadzone symulacje wskazują, że proponowany algorytm jest dość odporny na to założenie i w praktyce może być z powodzeniem stosowany także gdy nie jest ono bezwzględnie spełnione.

Podziękowania

Niniejszym pragniemy podziękować naszemu koledze dr inż. Aleksandrowi Mazgajowi, z którym rozpoczynaliśmy prezentowane badania. Za Jego zgodą, niniejszy tekst zawiera także wyniki tamtych, początkowych wspólnych badań.

Literatura

- [1] Ahmad I.A., Cerrito P.B., Nonparametric estimation of joint discrete-continuous probability densities with applications, *Journal of Statistical Planning and Inference*, Vol. 41, pp. 349–364 (1994)
- [2] Berger J.O., *Statistical Decision Theory*, Springer-Verlag, New York (1980)
- [3] Dawid A.P., Conditional Independence in Statistical Theory, *Journal of the Royal Statistical Society*, Series B, Vol. 41, pp. 1–31 (1979)
- [4] Deuffhard P., *Newton Methods for Nonlinear Problems. Affine Invariance and Adaptive Algorithms*, Springer, Berlin (2004)
- [5] Gaosheng J., Rui L., Zhongwen L., Nonparametric Estimation of Multivariate CDF with Categorical and Continuous Data, *Advances in Econometrics*, Vol. 25, pp. 291–318 (2009)
- [6] Kelley C.T., *Solving Nonlinear Equations with Newton's Method*, SIAM, Philadelphia (2003)
- [7] Kulczycki P., *Estymatory jądrowe w analizie systemowej*, WNT, Warszawa (2005)

- [8] Kulczycki P., Charytanowicz M., Warunkowa bayesowska identyfikacja parametryczna z niesymetrycznymi stratami, *XVII Krajowa Konferencja Automatyki*, Kielce, 19-22 czerwca 2011, pp. 215-226; postproceedings: *Postępy automatyki i robotyki*, K. Malinowski, R. Dindorf – red., Vol. 1, pp. 107-122, Komitet Automatyki i Robotyki PAN, WPŚ, Kielce (2011)
- [9] Kulczycki P., Charytanowicz M., Conditional Parameter Identification with Different Losses of Under- and Overestimation, *Applied Mathematical Modelling*, Vol. 37, pp. 2166–2177 (2013)
- [10] Kulczycki P., Charytanowicz M., An Algorithm for Conditional Multidimensional Parameter Identification with Asymmetric and Correlated Losses of Under- and Overestimations, in press (2014)
- [11] Kulczycki P., Hryniewicz O., Kacprzyk J. – red., *Techniki informacyjne w badaniach systemowych*, WNT, Warszawa (2007)
- [12] Kulczycki P., Mazgaj A., Parameter Identification for Asymmetrical Polynomial Loss Function, *Information Technology and Control*, Vol. 38, pp. 51–60 (2009); Errata: Vol. 38, pp. 167–168 (2009)
- [13] Lehmann E.L., *Theory of Point Estimation*, Wiley, New York (1983)
- [14] Li Q., Racine J.S., Nonparametric estimation of conditional CDF and quantile functions with mixed categorical and continuous data, *Journal of Business and Economic Statistics*, Vol. 26, pp. 423–434 (2008)
- [15] McCullough B.D., Optimal Prediction with a General Loss Function, *Journal of Combinatorics, Information and System Sciences*, Vol. 25, pp. 207–221 (2000)
- [16] Ouyang D., Li Q., Racine J.S., Cross-validation and the estimation of probability distributions with categorical data, *Journal of Nonparametric Statistics*, Vol. 18, pp. 69–100 (2006)
- [17] Silverman B.W., *Density Estimation for Statistics and Data Analysis*, Chapman and Hall, London (1986)
- [18] Simonoff J.S., *Smoothing Methods in Statistics*, Springer, New York (1996)
- [19] Stoer J., Bulirsch R., *Introduction to Numerical Analysis*, Springer-Verlag, New York (2002)
- [20] Walter E., Pronzato I., *Identification of Parametric Models*, Springer-Verlag, Berlin (1997)
- [21] Wand M.P., Jones M.C., *Kernel Smoothing*, Chapman and Hall, London (1995)
- [22] Wang M., Ryzin J., A class of smooth estimators for discrete distributions, *Biometrika*, Vol. 68, pp. 301–309 (1981)
- [23] Zellner A. Bayesian Estimation and Prediction Using Asymmetric Loss Function, *Journal of the American Statistical Association*, Vol. 81, pp. 446–451 (1985)