

SZYMON ŁUKASIK¹

IDENTYFIKACJA ROZKŁADU W SYSTEMACH RZECZYWISTYCH ZA POMOCĄ ESTYMATORÓW JĄDROWYCH

1. Wprowadzenie

Estymacja funkcji gęstości rozkładu zmiennej losowej jest jednym z głównych zagadnień z zakresu stosowania metod statystycznych. Tradycyjne metody estymacji polegają na założeniu z góry ustalonego typu rozkładu zmiennej losowej, a następnie określeniu parametrów definiujących ten rozkład. Z tego też powodu ogólnie nazywa się je *metodami parametrycznymi*.

Istnieje bogata literatura przedmiotowa opisująca algorytmy pomocne przy ich stosowaniu (np. [5,6]), jednak w obecnych czasach możliwości ich użycia do modelowania systemów rzeczywistych stają się coraz bardziej niewystarczające.

Oznaczenia

X	– rozważana zmienna losowa
n	– wymiar zmiennej losowej X
$x_1, x_2, \dots, x_m$	– próba losowa zmiennej X
m	– liczność elementów próby
$f(x)$	– funkcja gęstości rozkładu zmiennej losowej X
$\hat{f}(x)$	– estymator funkcji gęstości rozkładu skonstruowany na podstawie próby X

¹ mgr inż. Szymon Łukasik, Zakład Automatyki, Politechnika Krakowska.

Głównym powodem takiego stanu jest właśnie konieczność ograniczenia się w praktyce do kilkunastu dostępnych typów rozkładów. Sytuacja staje się jeszcze bardziej złożona w przypadku rozważania wielowymiarowych zmiennych losowych.

Wymienionej powyżej wady nie posiadają *metody nieparametryczne*, takie jak np. najprostsze *histogramy* czy, będące tematem niniejszego artykułu, *estymatory jądrowe* (ang. *Kernel Density Estimators – KDE*). W metodach tych nie określa się a priori typu rozkładu charakteryzującego badane zmienne. Doboru funkcji opisujących rozkład i ich współczynników dokonuje się stosując odpowiednie kryteria optymalizacyjne. Kolejnym przyczyną coraz częstszego sięgania po metody nieparametryczne jest gwałtowny wzrost mocy obliczeniowej współczesnych komputerów, umożliwiający rozwiązywanie w coraz krótszym czasie coraz bardziej złożonych problemów.

Celem niniejszego artykułu jest przedstawienie zarysu wiedzy na temat estymatorów jądrowych (praktycznie w polskiej literaturze, poza pozycjami [7,8], pomijanymi) oraz zaprezentowanie praktycznego zastosowania tychże estymatorów do identyfikacji rozkładu konkretnych systemów rzeczywistych.

Rozdział 2 niniejszej publikacji zawiera podstawowe informacje teoretyczne na temat estymatorów jądrowych, doboru *jądra* – funkcji stanowiącej o postaci estymatora oraz *współczynnika wygładzania* – zapewniającego „skalowanie” jąder dla rozważanego problemu.

Rozdział 3 opisuje praktyczną implementację metody estymatorów jądrowych wykorzystywaną w niniejszej publikacji oraz prezentuje rezultaty identyfikacji kilku wybranych systemów rzeczywistych uzyskane opisywaną we wcześniejszym rozdziale metodyką.

2. Estymatory jądrowe

Estymator jądrowy funkcji gęstości rozkładu prawdopodobieństwa zmiennej losowej X definiuje się jako:

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

gdzie:

$h > 0$ – tzw. współczynnik wygładzania.

$K : \mathbb{R}^n \rightarrow [0, \infty)$ – funkcja spełniająca warunki:

$$\int_{\mathbb{R}^n} K(x) dx = 1 \quad (2)$$

$$K(x) = K(-x) \quad \forall \quad x \in \mathbb{R}^n \quad (3)$$

$$K(0) \geq K(x) \quad \forall \quad x \in \mathbb{R}^n \quad (4)$$

Odwzorowanie K jest nazywane *jądrem* (ang. *kernel*) przez analogię do jądra operatora całkowego.

Na jakość estymacji jądrowej duży wpływ mają: wybór postaci jądra oraz zastosowanie odpowiedniej wartości parametru wygładzania. Zagadnienia te zostaną one omówione pokrótce w kolejnych podrozdziałach.

2.1. Wybór postaci jądra

W praktyce wymagania wobec funkcji jądra przedstawione na początku niniejszego rozdziału spełnia wiele funkcji analitycznych (najczęściej stosowane wymieniono w [8]). Wybór jądra przeprowadza się zatem biorąc pod uwagę rozmaite dodatkowe kryteria wynikające z rozważanego problemu. Przykładowo, gdy wymagane jest by rozkład posiadał pochodne dowolnie wysokiego rzędu można stosować jądro normalne, gdy natomiast pożądanym jest ograniczenie nośnika zmiennej losowej można użyć jądra typu dwuwagowego lub Epanechnikowa.

Przy wyborze typu jądra istotna jest również jego efektywność tj. stopień, w jakim minimalizowana jest przy jego zastosowaniu wartość scałkowanego błędu średniokwadratowego MISE (ang. *Mean Integrated Square Error*). Najwyższą efektywność posiada jądro Epanechnikowa. Warto nadmienić jednak, że spadek efektywności w przypadku stosowania innych standardowych typów jąder jest niewielki np. w przypadku jądra normalnego wynosi on 5%, tak więc bez wielkiej szkody dla przebiegu estymacji przy wyborze funkcji jądra można kierować się innymi, wspomnianymi wyżej, przesłankami. Przedstawione w pracy wyniki uzyskano przy użyciu estymatorów z jądrem normalnym.

W przypadku wielowymiarowym stosuje się jądra dwóch rodzajów: jądra radialne lub jądra produktowe.

Nazwa jądra radialnego wywodzi się od symetrii radialnej, którą to jądro wykazuje. Definiuje się je następująco:

$$K(x) = c_n K^1\left(\left\|\frac{x - x_i}{h}\right\|\right) \quad (5)$$

gdzie: K^1 oznacza jednowymiarową funkcję jądra, a c_n - stałą zapewniającą spełnienie przez jądro K warunku (2).

Z kolei jądro produktowe stanowi iloczyn (ang. *product*) jąder jednowymiarowych:

$$K(x) = K\left(\begin{bmatrix} x_1 \\ x_2 \\ \vdots \\ x_n \end{bmatrix}\right) = K^1(x_1) \cdot K^1(x_2) \cdot \dots \cdot K^1(x_n) \quad (6)$$

Każde z jąder $K^1(x_i)$ może być definiowane przez funkcję różnego typu (choć ze względów praktycznych nie jest to zalecane). Jądra te wyróżniają również przypisane do

nich współczynniki wygładzania h_i (zatem czynnik normujący h^n w definicji (1) przyjmuje postać $h^n = \prod_{i=1}^n h_i$).

Mimo iż większą efektywność (o 2% przykładowo - dla jąder Epanechnikowa i wymiaru $n = 2$) wykazuje jądro radialne, to w przypadku wielowymiarowej zmiennej losowej można zalecać stosowanie jądra produktowego. Sugestia ta wynika z praktycznych zalet tej postaci jądra, takich jak: łatwość wyznaczenia współczynnika wygładzania czy też bezproblemowe przeprowadzanie estymacji w przypadku, gdy zmienna losowa posiada różne skale poszczególnych współrzędnych (wybór jądro radialnego w takiej sytuacji wiąże się z przeprowadzeniem dodatkowej transformacji liniowej).

2.2. Wyznaczanie parametru wygładzania

Kluczowe znaczenie dla uzyskania wysokiej jakości estymacji ma właściwy dobór współczynnika wygładzania. Zbyt duża jego wartość powoduje nadmierne „rozmycie” funkcji gęstości rozkładu prawdopodobieństwa, zbyt mała – jego przesadne „wyostrenie”.

Większość opracowanych, w celu wyznaczenia optymalnej wartości współczynnika wygładzania, metod opiera się o kryterium minimalizacji MISE. Do najczęściej stosowanych należą: metoda podstawień (ang. *plug-in*) i krzyżowego uwiarygodniania (ang. *cross validation*).

Metoda podstawień jest zalecana w przypadku estymacji jednowymiarowej zmiennej losowej oraz dla zmiennej wielowymiarowej, gdy estymator konstruowany jest w oparciu o jądro produktowe. Natomiast metodę krzyżowego uwiarygodniania warto stosować dla zmiennych wielowymiarowych i estymatora z jądrem radialnym. Obie przedstawione metody szczegółowo omawia monografia [8].

Dla lepszego dopasowania estymatora do rzeczywistego rozkładu można również zastosować *adaptację parametru wygładzania* (inaczej *modyfikację parametru wygładzania* [8]). Polega ona na wstępnym wyznaczeniu h jedną z w/w. metod oraz wartości estymatora dla poszczególnych elementów próby $\hat{f}(x_i)$, a następnie określeniu dla każdego x_i współczynnika modyfikującego s_i , zależnego od stosunku wartości estymatora $\hat{f}(x_i)$ do średniej geometrycznej wartości estymatorów dla $i=1...m$. Współczynniki te modyfikują (przez przemnożenie) wartość obecnego w liczniku (1) parametru wygładzania h . Tak skonstruowany estymator cechuje zindywidualizowanie wpływu, jaki ma parametr wygładzania na jądra stowarzyszone z poszczególnymi elementami próby losowej.

3. Przykłady zastosowań w identyfikacji systemów rzeczywistych

Wyniki przedstawione w publikacji uzyskano przy użyciu programu *KDEstim*, napisanego wraz z mgr inż. Zenonem Cygankiem.

Program pozwala na wyznaczenie estymatora jądrowego rozkładu zmiennej losowej wielowymiarowej, określenie optymalnej wartości współczynnika wygładzania metodami

krzyżowego uwiarygodniania i podstawień, a także wizualizację i zapis otrzymanych wyników. Umożliwia on również uzyskanie estymatora innych charakterystyk probabilistycznych: dystrybucyj i kwantyla (zgodnie z opisanymi w [8] procedurami). Przy stworzeniu aplikacji wykorzystano środowisko programistyczne *Microsoft Visual Studio*.

Przykładowe zastosowania opisywanego narzędzia w estymacji nieparametrycznej dla wybranych systemów rzeczywistych przedstawiono w kolejnych podrozdziałach.

3.1. Analiza wybranych rozkładów z zakresu fizyki wysokich energii

Obiektem rozważań w tej części publikacji jest identyfikacja rozkładu zmiennych losowych, mierzonych w eksperymencie fizyki wysokich energii, metodą estymatorów jądrowych oraz porównanie otrzymanych wyników z efektami działania histogramów - celem weryfikacji zaproponowanej metody badawczej i podkreślenia zalet jakie ona posiada.

Przedstawione próby zostały uzyskane przy pomocy detektora ZEUS na akceleratorze HERA w DESY w Hamburgu. Dane doświadczalne zgromadzono w dedykowanych pomiarach mających na celu precyzyjne wyznaczenie wartości przekroju czynnego na fotoprodukcję [1,4].

W procesie fotoprodukcji rzeczywiste fotony oddziałując z materią (w tym wypadku z protonami) produkują hadrony w stanie końcowym. Przypadki oddziaływań, dla których obserwowano pozytron (antycząstkę elektronu) w dedykowanym kalorymetrze elektronowym [1] i jednocześnie obserwowano znaczącą aktywność w głównym kalorymetrze detektora ZEUS, były akceptowane i zapisywane na dysku. W czasie trwania eksperymentu wybrano szerszą klasę oddziaływań (spełniających łagodniejsze żądania) do zapisania na dysku. Miało to na celu z jednej strony możliwość kalibracji aparatury doświadczalnej a z drugiej kontrolę procesów składających się na tło. W szczególności dane te zawierały znaczny procent danych pochodzących z procesu bremsstrahlungu czy też promieniowania w stanie początkowym (patrz: [4]). Dane te pochodzą z 1996 roku, gdy akcelerator HERA zderzał ze sobą wiązkę pozytronów o energii 27,5 GeV z wiązką protonów o energii 820 GeV.

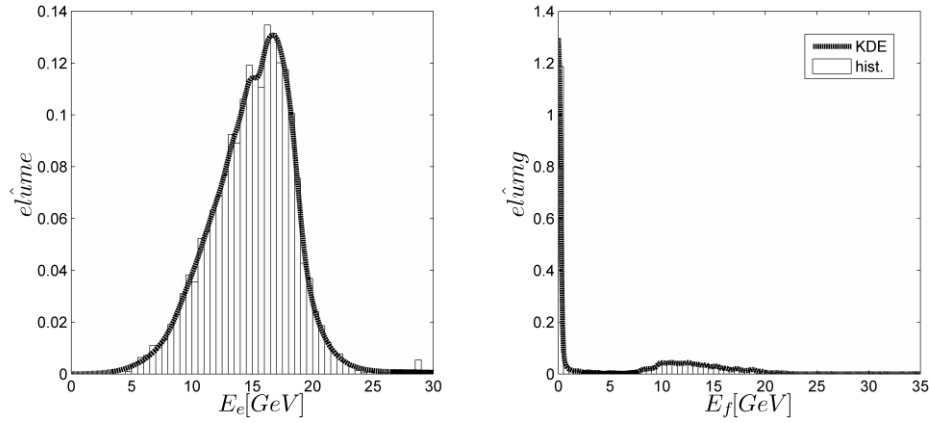
Z otrzymanych w ten sposób danych wyodrębniono próby następujących zmiennych losowych:

- $elume$ – energia elektronu, mierzona przez kalorymetr elektronów, wyrażona w GeV (liczność: 4395),
- $elumg$ – energia fotonu, mierzona przez kalorymetr fotonowy [1], wyrażona w GeV (liczność: 4395),

Rysunki 1–2 prezentują rozkłady prawdopodobieństwa wyżej wymienionych zmiennych losowych – uzyskane poprzez zastosowanie estymatorów jądrowych o współczynnikach wygładzania uzyskanych metodą plug-in. Zestawiono je z histogramami uzyskanymi wg. wzoru:

$$\hat{f}(x) = \frac{\#\{x_i \in H_k\}}{mh} \quad (7)$$

tzn. dla każdego przedziału H_k histogram definiuje się jako funkcję równą liczbie tych wartości próby losowej, które należą do tego przedziału (oznaczanej jako $\#\{x_i \in H_k\}$) podzielonej przez mh (normalizacja).

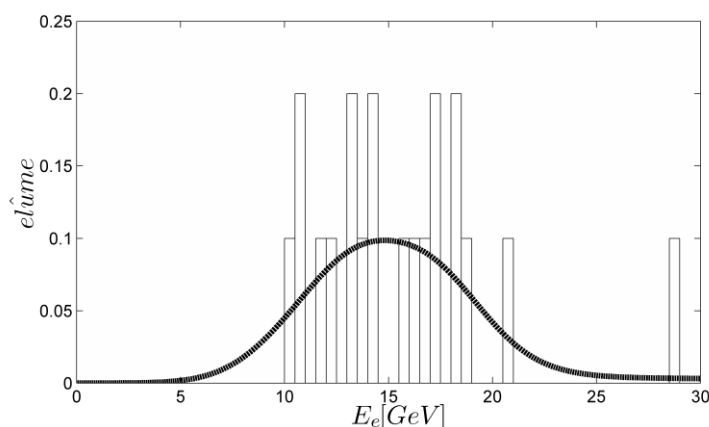


Rys. 1,2. Estymatory gęstości rozkładu zmiennych losowych $elume$ i $elumg$ uzyskane met. histogramów i estymatorami jądrowymi

Fig. 1,2. Density function estimates of $elume$ and $elumg$ variables distributions obtained using histograms and kernel estimators

Obie metody dają w tym przypadku bardzo podobne wyniki. Jeśli jednak do konstrukcji estymatora rozkładu użyta zostanie próba o znacznie zredukowanej liczności (20 elementów), a rozkład nie posiada bardzo wyraźnego maksimum (jak $elumg$) to przewaga zaproponowanej metody estymacji nad histogramami jest zauważalna – co pokazano na przykładzie rozkładu $elume$ zaprezentowanego na rys. 3.

Estymatory jądrowe znacznie lepiej oddają zatem charakter badanego rozkładu gdy dostępna jest skąpa ilość danych doświadczalnych (co często ma miejsce gdy istnieją problemy z powtarzalnością eksperymentu). Estymatory te nie posiadają również wad charakterystycznych dla histogramów: dużej zależności kształtu rozkładu od szerokości binów oraz punktu zaczepienia pierwszego z nich. Można więc zalecać, tam gdzie jest to możliwe, stosowanie, jako narzędzia identyfikacji gęstości rozkładów badanych zmiennych losowych, estymatorów jądrowych zamiast histogramów.



Rys. 3. Estymator gęstości rozkładu zmiennej losowej *elume* uzyskany met. histogramów i estymatorami jądrowymi dla zredukowanej liczności próby

Fig. 3. Density function estimate of *elume* variable distribution obtained using histograms and kernel estimators for reduced sample size

3.2. Wykrywanie uszkodzeń silnika asynchronicznego

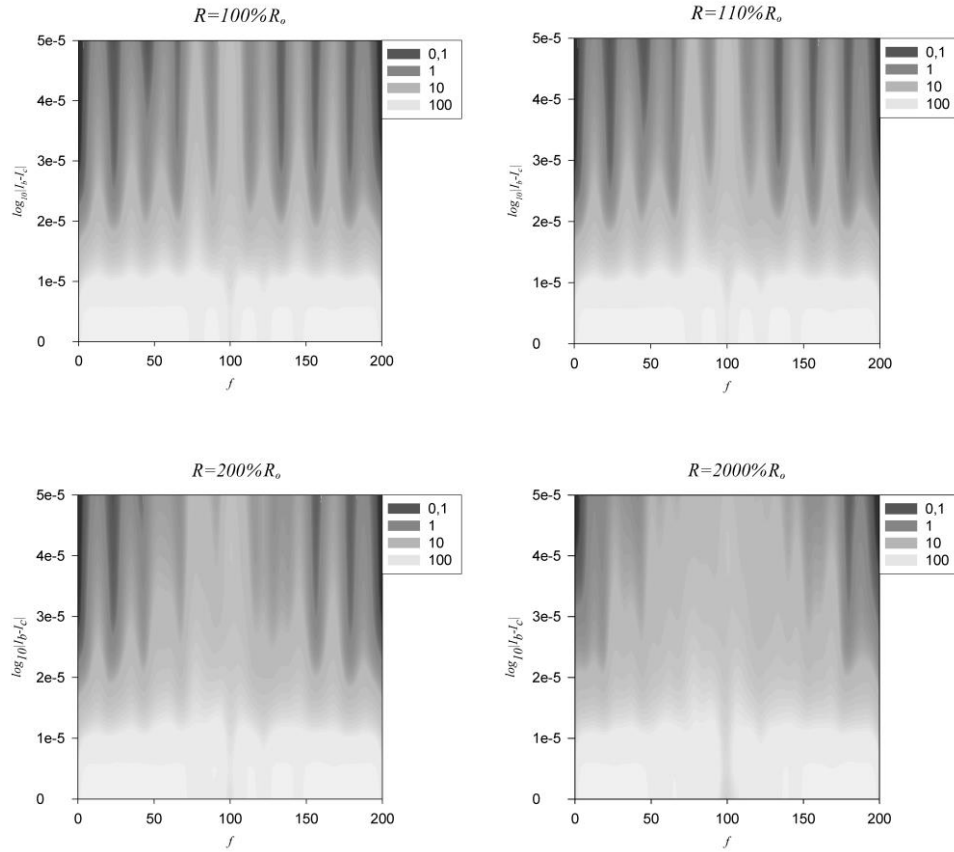
Jedną z możliwości wykorzystania estymatorów jądrowych jest ich użycie do detekcji i diagnozy uszkodzeń. Niniejszy podrozdział poświęcony zostanie wynikom badań przydatności estymacji jądrowej w zastosowaniu do wykrycia uszkodzenia i próbie oszacowania jego wielkości w przypadku trójfazowego silnika asynchronicznego klatkowego.

Zebrane dane zostały wygenerowane na podstawie modelu matematycznego opisanego szczegółowo w [9]. Zawierają one wyniki pomiarów wartości składowych harmonicznych prądu dla faz I_a , I_b i I_c uzwojenia stojana, (połączonego w gwiazdę bez przewodu zerowego) wraz numerami f tych harmonicznych (w zakresie 0 do 200). Dane te dotyczą 4 stanów silnika:

- 1) silnik nieuszkodzony
- 2) silnik o rezystancji jednego pręta zwiększonej o 10%
- 3) silnik o rezystancji jednego pręta zwiększonej o 100%
- 4) silnik o rezystancji jednego pręta zwiększonej o 1900%

Dla każdego z 4 w/w. stanów wykorzystano 120 przypadków o różnych wartościach ekscentryczności statycznej i dynamicznej przy stałym poślizgu $s = 0,5$. Analizie poddano funkcję gęstości rozkładu dwuwymiarowej zmiennej losowej X o próbie zdefiniowanej następująco: $x_i = [f_i, \log_{10}(|I_{bi} - I_{ci}| + 1)]$.

Otrzymane w ten sposób wyniki porównano dla 4 wyodrębnionych stanów silnika (rys.4-7).



Rys. 4,5,6,7. Estymator gęstości rozkładu zmiennej losowej X dla 4 stanów silnika (gradient szarości w skali logarytmicznej)

Fig. 4,5,6,7. Density function estimate of X variable distribution for 4 engine states (grey gradient in logarithmic scale)

Można zauważyć, że rozkład dla silnika nieuszkodzonego posiada uporządkowaną strukturę: najbardziej prawdopodobne są różnice prądów w zakresie pasma $0 - 2 \cdot 10^{-5}$, dla większych różnic pojawiają się „prążki” skojarzone z wybranymi zakresami częstotliwości (w rozważanym zakresie można wyróżnić 9 takich równomiernie rozłożonych prążków). W miarę wzrostu uszkodzenia prążki ulegają stopniowemu rozmyciu, coraz mniej stają się też prawdopodobne małe różnice prądów. W przypadku skrajnie powiększonej rezystancji pręta (rys. 7) prążki praktycznie nie występują, a rozkład uzyskuje znacznie mniej uporządkowaną strukturę.

Obserwacje te mogą posłużyć do konstrukcji układu monitorującego stan silnika, pozwalającego na wykrycie uszkodzenia jednego z jego prętów. Dalsze badania dotyczące

omawianego tematu obejmować powinny analizę wpływu uszkodzeń kilku prętów, występowania ekscentryczności oraz weryfikację sformułowanych wniosków przy użyciu danych rzeczywistych.

3.3. Analiza wyników sondażu PGSS

Współczesna analiza systemowa zajmuje się często układami, które reprezentują zjawiska i procesy z dziedzin spoza szeroko rozumianych dyscyplin technicznych – takich jak: socjologia, zarządzanie i marketing, handel itp.

Przedmiotem rozważań w niniejszym podrozdziale jest zastosowanie estymatorów jądrowych w celu uzyskania gęstości rozkładu prawdopodobieństwa zmiennej losowej, której próba pobrana została z ogólnodostępnych danych socjologicznych stanowiących wyniki Polskich Generalnych Sondaży Społecznych (w skrócie PGSS) [2].

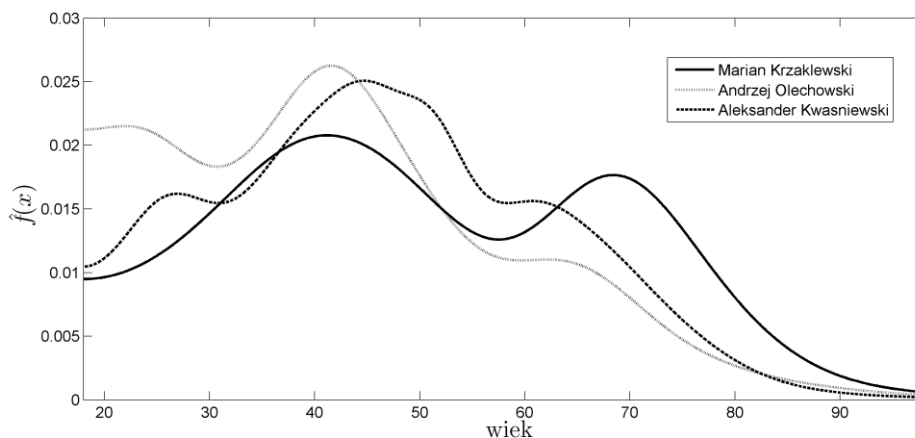
Sondaże PGSS, przeprowadzano kolejno w latach 1993, 1994, 1995, 1997, 1999, 2002. Ich obszerne podsumowanie stanowi publikacja [3]. Ogólnym celem omawianego Sondażu jest „systematyczny pomiar trendów i skutków zmian społecznych w Polsce” [3].

Do analizy w niniejszej publikacji wybrano wyniki uzyskane w roku 2002, a konkretnie próbę dla zmiennej losowej *AGE* określającą wiek respondenta. Odniesiono ją do wartości przyjmowanych przez zmienną losową *WHO00_1*. Określa ona odpowiedź respondenta na pytanie: „Na kogo oddał(a) Pan(i) głos w wyborach prezydenckich w 2000 r.?”.

Po przeniesieniu danych z formatu SPSS do postaci obsługiwanej przez program *KDEstim* wydzielono klasy odpowiadające poszczególnym kandydatom, na których respondenci oddali swój głos. Każda z klas zawiera próbę losową reprezentującą wiek respondentów (w roku wyborczym) którzy zagłosowali na daną osobę ubiegającą się o urząd Prezydenta Rzeczypospolitej Polskiej w 2000 roku. Próby o liczności mniejszej niż 70 zostały odrzucone.

Dla każdej z klas wyznaczono gęstość rozkładu prawdopodobieństwa stosując estymatory jądrowe o współczynnikach wygładzania wyznaczonych metodą *plug-in*. Aby oddać rzeczywistą własność populacji zastosowano ograniczenie lewostronnie nośnika zmiennej losowej dla wartości 18 (czynne prawo wyborcze przysługuje w Polsce obywatelom polskim, którzy ukończyli 18 lat). Metodologia ograniczeń nośnika zmiennej losowej opisana została szczegółowo w [8].

Uzyskane wyniki dla trzech liderów sondażu przedstawia rys. 8.



Rys. 8. Estymatory gęstości rozkładu zmiennej losowej *AGE* dla 3 liderów sondażu

Fig. 8. Density function estimates of *AGE* variable distributions for 3 poll leaders

Dokładna analiza otrzymanych wyników leży poza zakresem merytorycznym niniejszej publikacji. Można jednak poczynić parę interesujących obserwacji. Nietrudno zauważyć, że maksima rozkładów dla wszystkich kandydatów występują dla wieku w przedziale 40-50 lat. W przypadku Mariana Krzaklewskiego maksimum to jest najmniej zarysowane – relatywnie sporą część jego elektoratu stanowią ludzie starsi (drugie maksimum lokalne dla wieku ok. 70 lat). Natomiast wśród potencjalnych wyborców Andrzeja Olechowskiego występuje największy odsetek ludzi młodych.

Dowodem na reprezentatywność danych sondażowych oraz skuteczność zastosowanej metody badawczej może być porównanie zaprezentowanych rezultatów z uzyskanymi poprzez tzw. exit-polls czyli sondażu prowadzonego w czasie wyborów w którym wyborcy wychodzący z lokali wyborczych odpowiadają na pytanie na kogo głosowali. Skróć wynikiów takiego badania przedstawiono w tabeli 1.

Tabela 1

Wyniki sondażu exit-polls w trakcie wyborów prezydenckich w 2000 r.
[podano za INTERIA.PL]

Wiek	Marian Krzaklewski	Aleksander Kwaśniewski	Andrzej Olechowski
18-24	10,2 %	15,8 %	18,9 %
25-39	23,0 %	26,7 %	31,3 %
40-59	38,5 %	40,9 %	38,4 %
60 i więcej	28,2 %	16,6 %	11,4 %

Rezultat tak przeprowadzonej weryfikacji jest pozytywny. Potwierdzone zostają zaobserwowane wcześniej tendencje. Estymatory jądrowe mogą być zatem skutecznie stosowane do celów analizy danych socjologicznych zebranych w ramach sondaży PGSS.

5. Podsumowanie

W niniejszej publikacji przedstawiono koncepcję estymatorów jądrowych – jednych z podstawowych narzędzi współczesnej nieparametrycznej analizy danych. Zaprezentowano również przykłady z różnych dziedzin nauki, które dowodzą skuteczności opisywanej metody w identyfikacji rozkładów zmiennych losowych oraz jej przewagi nad powszechnie stosowanymi obecnie histogramami. Uzyskane wyniki, zwłaszcza w dziedzinie wykrywania uszkodzeń silnika asynchronicznego, zachęcają do kontynuacji badań w zawężonym, bardziej wyspecjalizowanym zakresie.

Warto jednak przy okazji nadmienić, że tak przeprowadzona estymacja posiada jedną praktyczną wadę – dużą złożoność obliczeniową, która sprawia, że otrzymanie wyniku dla licznej (np. 100 000) próby wymaga długiego czasu obliczeń. Przedmiotem dalszych badań prowadzonych przez autora niniejszej publikacji będzie więc również możliwość przeprowadzenia procesu estymacji w środowisku równoległym.

Podziękowanie

Autor składa serdeczne podziękowania prof. Piotrowi Kulczyckiemu, prof. Januszowi Chwastowskiemu, prof. Piotrowi Maleckiemu, dr Maciejowi Sułowiczowi, dr Konradowi Weinrebowi, a także serdecznemu współpracownikowi mgr Zenonowi Cygankowi za ich pomoc w trakcie tworzenia niniejszego artykułu.

Literatura

- [1] Andruszków J. et al., „*Luminosity measurement in the ZEUS experiment*”, Acta Phys. Polonica vol. B32, pp. 2025-2057, 2001.
- [2] Cichomski B. (kierownik programu), Jerzyński T., Zieliński M., „*Polskie Generalne Sondaże Społeczne: skumulowany komputerowy zbiór danych 1992-2002*”, Instytut Studiów Społecznych, Uniwersytet Warszawski, Warszawa, 2003.
- [3] Cichomski B. (kierownik programu), Jerzyński T., Zieliński M., „*Polskie Generalne Sondaże Społeczne: struktura skumulowanych wyników badań 1992-2002*”, Instytut Studiów Społecznych, Uniwersytet Warszawski, Warszawa, 2003.
- [4] Chwastowski J. (praca habilitacyjna), „*Energy Evolution of the Total Cross Sections – Significance of the HERA γp Measurements*”, Report no 1940/PH, Instytut Fizyki Jądrowej im. Henryka Niewodniczańskiego w Krakowie, 2004.
- [5] Greń J., *Statystyka matematyczna*, PWN, Warszawa, 1987.
- [6] Koronacki J., Mielniczuk J., *Statystyka*, PWN, Warszawa, 2001.
- [7] Kulczycki P., *Wykrywanie uszkodzeń w systemach zautomatyzowanych metodami statystycznymi*, Alfa, Warszawa, 1998.

- [8] Kulczycki P., „*Estymatory jądrowe w analizie systemowej*”, WNT, Warszawa, 2005.
- [9] Sobczyk T.J., Weinreb K., Węgiel T., Sułowicz M., „*Theoretical Study of Effects Due to Rotor Eccentricities in Induction Motors*”, IEEE International Symposium on Diagnostics for Electrical Machines, Power Electronics and Drives (SDEMPED'99), pp.289-295, Gijon, 1999.

Streszczenie

Metody nieparametryczne znajdują coraz szersze zastosowanie w zagadnieniach współczesnej analizy i eksploracji danych. W artykule przedstawiono najpopularniejsze narzędzie powyższych metod – estymatory jądrowe. Poza koncepcją tychże estymatorów zaprezentowano także praktyczne aspekty estymacji oraz przykłady jej zastosowań do wyznaczania rozkładów z dziedziny fizyki wysokich energii, wykrywania uszkodzeń silnika asynchronicznego oraz analizy danych socjologicznych.

Słowa kluczowe: estymacja nieparametryczna, statystyczne estymatory jądrowe, fizyka jądrowa, wykrywanie uszkodzeń, silniki asynchroniczne, Polskie Generalne Sondaże Społeczne

Abstract

Nonparametric methods find increasing number of applications in the area of data analysis and data exploration. In this paper the most popular tool of those methods was presented – kernel estimators. Beside of kernel estimators' concept, practical estimation aspects and examples of applications in determining distributions from nuclear physics experiments, induction motors' damage detection and analysis of sociological data were shown as well.

Key words: nonparametric estimation, statistical kernel estimators, nuclear physics, damage detection, induction motors, Polish General Social Survey