

Aleksandra Baszczyńska

Uniwersytet Łódzki

UWAGI O TEŚCIE ZGODNOŚCI LI WYKORZYSTUJĄCYM METODĘ JĄDROWĄ

1. Wstęp

Nieparametryczny jądrowy estymator funkcji gęstości, zaproponowany przez Rosenblatta, określony jest dla próby prostej $X_1, X_2, \dots, X_n$ w następujący sposób [5]:

$$\hat{f}(x) = \frac{1}{nh} \sum_{i=1}^n K\left(\frac{x - X_i}{h}\right), \quad (1)$$

gdzie $K(u)$ jest funkcją zwaną funkcją jądra, spełniającą warunek $\int_{-\infty}^{+\infty} K(u) du = 1$,

natomiast $h > 0$ jest parametrem wygładzania.

Zarówno funkcja jądra, jak i parametr wygładzania są parametrami metody jądrowej, właściwy wybór tych parametrów zapewnia odpowiednią dobroć estymacji. W nieparametrycznej estymacji funkcji gęstości najczęściej stosowane są następujące funkcje jądra: Gaussowska, Epanechnikova, równomierna, dwukwadratowa, trójkątna, trójkwadratowa oraz kosinusowa. Zagadnienie wyboru funkcji jądra w estymacji funkcji gęstości jest rozważane w literaturze [1]. Wybór parametru wygładzania może być dokonany np. metodą Silvermana lub metodą *cross-validation* [7].

Metoda jądrowa może być również wykorzystywana w weryfikacji hipotez statystycznych. Istnieją testy statystyczne, w których nieparametryczny jądrowy estymator funkcji gęstości stosowany jest w konstrukcji statystyki testowej. Są to m.in. testy zgodności (w tym normalności), testy niezależności oraz testy symetryczności (por. [6]).

Test zgodności zaproponowany przez Li pozwala na weryfikację hipotezy o zgodności rozkładów dwóch populacji. W konstrukcji statystyki testowej wykorzystywana jest metoda jądrowa, stąd wynika potrzeba oceny wpływu parametrów metody jądrowej na wynik procedury testowej.

Badanie własności testu zgodności Li będzie dokonane metodami symulacyjnymi. Ponadto zostanie przeprowadzona weryfikacja hipotezy o takim samym rozkładzie wskaźnika percepcji korupcji w latach 1996 i 2005. Oprócz testu zgodności Li wykorzystane będą inne nieparametryczne testy dla zmiennych połączonych, takie jak: test znaków, test rangowanych znaków (zwany również testem kolejności par Wilcozona), test znaków Fischera oraz test Friedmana.

2. Test zgodności Li

Niech $f(x)$ oraz $g(x)$ będą nieznanymi ograniczonymi i ciągłymi funkcjami gęstości 2 populacji. Hipotezy zerowa i alternatywna w teście zgodności Li mają postać:

$$\begin{aligned} H_0 : f(x) &= g(x), \\ H_1 : f(x) &\neq g(x). \end{aligned} \quad (2)$$

Zagadnienie polega zatem na weryfikacji hipotezy o identyczności rozkładów 2 populacji, co może być wykorzystane do stwierdzenia, czy zmiany w rozkładach badanych populacji są statystycznie znaczące.

Próby $\{x_i\}$, $\{y_i\}$ dla $i = 1, \dots, n$ pochodzą z 2 populacji o rozkładach z funkcjami gęstości, odpowiednio, $f(x)$ oraz $g(x)$. Liczebności prób pochodzących z badanych populacji są takie same. Ponadto $\{x_i\}$ i $\{y_i\}$ mogą być zależne, w sensie obserwacji pochodzących z badań panelowych (np. badanie rozkładów dochodów w 2 regionach, badanie rozkładów wskaźników ekonomiczno-społecznych w 2 momentach czasowych).

Statystyka testowa jest oparta na mierze odległości $I = \int_{-\infty}^{+\infty} (\hat{f}(x) - \hat{g}(x))^2 dx$,

gdzie $\hat{f}(x) = \frac{1}{nh} \sum_{i=1}^n K\left(\frac{x_i - x}{h}\right)$ i $\hat{g}(x) = \frac{1}{nh} \sum_{i=1}^n K\left(\frac{y_i - x}{h}\right)$ są nieparametrycznymi estymatorami jądrowymi funkcji gęstości odpowiednio pierwszej i drugiej populacji.

Li zaproponował, do weryfikacji hipotezy (2) o zgodności rozkładów, następującą statystykę testową:

$$T = n\sqrt{h} \frac{\left(I - \frac{2K(0)}{nh}\right)}{\hat{\sigma}_0}, \quad (3)$$

gdzie: $K(u)$ jest funkcją jądra, h jest parametrem wygładzania,

$$I = \frac{1}{n^2 h} \sum_{i=1}^n \sum_{j=1, j \neq i}^n \left[K\left(\frac{x_i - x_j}{h}\right) + K\left(\frac{y_i - y_j}{h}\right) - K\left(\frac{y_i - x_j}{h}\right) - K\left(\frac{x_i - y_j}{h}\right) \right] + \\ + \frac{1}{n^2 h} \sum_{i=1}^n \left[2K(0) - 2K\left(\frac{x_i - y_i}{h}\right) \right],$$

oraz

$$\hat{\sigma}_0^2 = \frac{1}{n^2 h} \sum_{i=1}^n \sum_{j=1}^n \left[K\left(\frac{x_i - x_j}{h}\right) + K\left(\frac{y_i - y_j}{h}\right) + 2K\left(\frac{x_i - y_j}{h}\right) \right] \int_{-\infty}^{+\infty} K^2(u) du.$$

Przy założeniu prawdziwości hipotezy zerowej, dla $h \rightarrow 0$, $nh \rightarrow \infty$, statystyka (3) ma rozkład normalny z wartością oczekiwaną 0 i odchyleniem standardowym 1.

Modyfikacje statystyki testowej (3) są następujące [6]:

$$T_1 = n\sqrt{h} \frac{I_1}{\hat{\sigma}_0}, \quad (4)$$

$$T_2 = n\sqrt{h} \frac{I_1}{\tilde{\sigma}_0}, \quad (5)$$

$$T_3 = n\sqrt{h} \frac{I_1}{\bar{\sigma}_0}, \quad (6)$$

gdzie:

$$I_1 = \frac{1}{n^2 h} \sum_{i=1}^n \sum_{j=1, j \neq i}^n \left[K\left(\frac{x_i - x_j}{h}\right) + K\left(\frac{y_i - y_j}{h}\right) - K\left(\frac{y_i - x_j}{h}\right) - K\left(\frac{x_i - y_j}{h}\right) \right],$$

$$\tilde{\sigma}_0^2 = \hat{\sigma}_0^2 - 4h \{ 2E_n[f^2(x)] - [E_n[f(x)]]^2 \},$$

$$E_n[f(x)] = \frac{1}{n} \sum_{i=1}^n f_n(x_i),$$

$$E_n[f^2(x)] = \frac{1}{n} \sum_{i=1}^n f_n^2(x_i),$$

$$\bar{\sigma}_0^2 = \frac{\hat{\sigma}_0^2 + \tilde{\sigma}_0^2}{2}.$$

W konstrukcji statystyk testowych (3), (4), (5) i (6) istnieje potrzeba wykorzystania metody jądrowej z funkcją jądra $K(u)$ i parametrem wygładzania h .

Dlatego tak ważnym zagadnieniem jest ocena wpływu parametrów metody jądrowej na wynik procedury testowej dla wszystkich 4 statystyk testowych zaproponowanych przez Li.

3. Badanie własności testu Li metodami symulacyjnymi

W celu zbadania własności testu Li przeprowadzono eksperymenty pozwalające określić „najlepszą” statystykę testową testu Li wśród rozważanych powyżej oraz określić wpływ parametrów metody jądrowej.

3.1. Etap I

Rozważano dwie populacje o rozkładach opisanych następującymi funkcjami gęstości:

$f(x)$ funkcja gęstości rozkładu normalnego $N(5;2)$,

$g(x)$ funkcja gęstości mieszaniny dwóch rozkładów ciągłych $g(x) = p_1 g_1(x) + p_2 g_2(x)$,

gdzie p_1 i p_2 są udziałami ($p_1, p_2 > 0$ i $p_1 + p_2 = 1$).

Analizie poddano następujące warianty mieszaniny rozkładów populacji drugiej.

Wariant I:

$g_1(x)$ jest funkcją gęstości rozkładu normalnego $N(5;2)$,

natomiast $g_2(x)$ jest funkcją gęstości rozkładu normalnego $N(9;1,5)$.

Wariant II:

$g_1(x)$ jest funkcją gęstości rozkładu normalnego $N(5;2)$,

natomiast $g_2(x)$ jest funkcją gęstości rozkładu normalnego $N(1;0,75)$.

Wariant III:

$g_1(x)$ jest funkcją gęstości rozkładu normalnego $N(5;2)$,

natomiast $g_2(x)$ jest funkcją gęstości rozkładu normalnego $N(5;1)$.

Wariant IV:

$g_1(x)$ jest funkcją gęstości rozkładu normalnego $N(5;2)$,

natomiast $g_2(x)$ jest funkcją gęstości rozkładu normalnego $N(6;2)$.

Wariant V:

$g(x)$ jest funkcją gęstości rozkładu normalnego $N(5;2)$,

W eksperymencie $p_2 g_2(x)$ traktowane jest jako zanieczyszczenie rozkładu z funkcją gęstości $g(x)$, gdzie p_2 jest udziałem zanieczyszczenia populacji drugiej. W wariantach V rozważane populacje mają taki sam rozkład (taka sama funkcja gęstości oraz takie same parametry rozkładów 2 populacji).

W ten sposób określone zostały populacje, których rozkłady będą rozważane w hipotezie (2) testu Li. Udziały zanieczyszczenia populacji drugiej p_2 w badaniu przybiera wartości od 0,01 do 0,3, co powoduje stosunkowo niewielkie różnice między rozkładami badanych populacji.

Celem badania jest analiza wpływu wielkości zanieczyszczenia (poprzez p_2) i charakteru zanieczyszczenia (poprzez $g_2(x)$) na wartość statystyki testowej testu zgodności Li, a co za tym idzie, na wynik procedury testowej. Do weryfikacji hipotezy zerowej mówiącej o zgodności 2 rozkładów będą wykorzystywane 4 statystyki testowe: (3), (4), (5), (6).

Z każdej tak określonej populacji (odpowiedni wariant badania oraz udział zanieczyszczenia w populacji drugiej) pobierano próby proste (liczebność 15, 50, 120). Na podstawie wyników prób obliczano statystykę testową testu zgodności Li oraz podejmowano odpowiednią decyzję. Oczywiście w każdym z przypadków, gdy wprowadzano zanieczyszczenia populacji drugiej, oczekiwaną decyzją było odrzucenie hipotezy zerowej mówiącej o takim samym rozkładzie badanych populacji. Eksperyment powtarzano 1000 razy.

W etapie I we wszystkich wymienionych wariantach udział p_2 wynosił 0,01, 0,02, ..., 0,29, 0,3 (krok 0,01).

W każdym wariancie oraz dla każdej wartości p_2 wyznaczano następujące charakterystyki: średnie wartości T , T_1 , T_2 , T_3 , I ; odchylenie standardowe, minimalną wartość, maksymalną wartość powyższych charakterystyk oraz wyznaczano liczbę przypadków (na 1000), w których odrzucano hipotezę zerową (gdy statystyki testowe T , T_1 , T_2 , T_3 wpadały do obszaru odrzucenia) mówiącą o takim samym rozkładzie badanych populacji (czyli decyzji właściwych). Liczba przypadków, gdy statystyki testowe wpadały do obszaru odrzucenia, była określana przy poziomie istotności 0,1 i 0,05. Charakterystykę I można traktować jako miarę bliskości estymatorów funkcji gęstości porównywanych populacji.

Wyniki eksperymentu dla przykładowego udziału $p_2 = 0,05$ (5% zanieczyszczenie populacji drugiej) oraz dla parametrów metody jądrowej następującej postaci:

$$K(u) = \frac{1}{\sqrt{2\pi}} e^{-\frac{1}{2}u^2} \quad (\text{funkcja jądra Gaussowska}),$$

$$h = 1,06\hat{\sigma}n^{-\frac{1}{5}} \quad (\text{parametr wygładzania wyznaczony metodą Silvermana}), \text{ gdzie}$$

$$\hat{\sigma} = \sqrt{\frac{1}{n-1} \sum_{i=1}^n (x_i - \bar{x})^2},$$

przedstawiono w tab. 1.

Tabela 1. Charakterystyki testowe dla udziału zanieczyszczenia $p_2 = 0,05$.

Funkcja jądra Gaussowska. Parametr wygładzania wyznaczony metodą Silvermana

Wariant $g_2(x)$	n	T	T_2	I	$LOT_{0,1}$	$LOT_{0,05}$	$LOT_{2,0,1}$	$LOT_{2,0,05}$
I $N(9;1,5)$	50	-0,38	0,01	0,0118	18	13	51	27
	120	-0,28	0,09	0,0066	27	14	76	34
II $N(1;0,75)$	50	-0,35	0,05	0,0123	24	11	56	34
	120	-0,24	0,14	0,0068	29	17	71	43
III $N(5;1)$	50	-0,41	-0,12	0,0112	15	10	44	28
	120	-0,36	0,01	0,0060	22	10	49	29
IV $N(6;2)$	50	-0,41	-0,02	0,0112	18	10	47	30
	120	-0,35	0,02	0,0061	20	10	51	31

Źródło: obliczenia własne.

Wyniki etapu I eksperymentu wskazują wyraźnie na przewagę statystyki T_2 nad pozostałymi modyfikacjami statystyk testowych testu Li. Statystyka ta może być traktowana jako najbardziej wrażliwa, gdyż przybierała w każdym z rozważanych wariantów, dla określonego poziomu zanieczyszczenia, wartości największe. Oznacza to, że jako pierwsza spośród 4 statystyk testowych wpadała do obszaru odrzucenia (czyli wskazywała na niezgodność rozkładów populacji). Jest ona statystyką testową najlepszą (w sensie najbardziej wrażliwą) spośród rozważanych. Należy zwrócić uwagę, że wystąpiły takie przypadki etapu I eksperymentu, gdy statystyka testowa T_2 wpadała do obszaru odrzucenia (czyli była podstawą do słusznej decyzji), podczas gdy pozostałe statystyki testowe znajdowały się w obszarze przyjęcia, przy takich samych warunkach wstępnych eksperymentu. Można zatem określić hierarchię wrażliwości statystyk testu Li: T_2 , T_3 , T_1 oraz T .

W miarę wzrostu liczebności próby miara odległości I jest coraz mniejsza, wraz ze wzrostem liczebności próby zwiększa się również liczba odrzuceń hipotezy zerowej (czyli liczba decyzji prawidłowych) dla wszystkich rozważanych statystyk testu Li.

3.2. Etap II

Punktem wyjścia etapu II są wyniki dotyczące takich przypadków (takich udziałów zanieczyszczenia), gdzie sąsiednie udziały zanieczyszczenia wskazywały na różne decyzje dotyczące zgodności rozkładów dwóch populacji (na podstawie statystyki T_2). Analizowano jedynie warianty I i II, gdyż w wyniku etapu I uzyskano informację, że nawet dla 30% zanieczyszczenia populacji drugiej statystyka testowa nie wpadała do obszaru odrzucenia. Badano, która z funkcji jądra wykorzystywana w konstrukcji statystyki testowej daje podobne wyniki jak funkcja jądra Gaussowska stosowana w etapie I eksperymentu. Wyniki dla prób 50-elementowych oraz dla funkcji jądra: Gaussowskiej, Epanechnikova, równo-

miernej, dwukwadratowej, trójkątnej, trójkwadratowej i kosinusowej, przedstawiono w tab. 2 i 3.

Tabela 2. Wartości statystyki T_2 testu Li dla wybranych udziałów zanieczyszczenia p_2 dla wariantu I zanieczyszczenia populacji drugiej

Funkcja jądra	Udział zanieczyszczenia p_2			
	0,26	0,27	0,30	0,31
Gaussowska	1,15	1,36	1,59	1,84
Epanechnikowa	0,96	1,14	1,34	1,55
Równomierna	1,07	1,27	1,50	1,73
Dwukwadratowa	0,86	1,02	1,20	1,39
Trójkątna	0,90	1,06	1,24	1,44
Trójkwadratowa	0,79	0,93	1,10	1,27
Kosinusowa	0,94	1,12	1,31	1,52

Źródło: obliczenia własne.

Tabela 3. Wartości statystyki T_2 testu Li dla wybranych udziałów zanieczyszczenia p_2 dla wariantu II zanieczyszczenia populacji drugiej

Funkcja jądra	Udział zanieczyszczenia p_2			
	0,20	0,21	0,24	0,25
Gaussowska	1,09	1,34	1,62	1,93
Epanechnikowa	1,01	1,26	1,52	1,81
Równomierna	1,10	1,36	1,65	1,97
Dwukwadratowa	0,92	1,13	1,38	1,65
Trójkątna	0,95	1,18	1,43	1,70
Trójkwadratowa	0,85	1,06	1,28	1,52
Kosinusowa	1,00	1,23	1,50	1,79

Źródło: obliczenia własne.

W wariancie I i II eksperymentu statystyka testowa T_2 przybiera podobne wartości jedynie dla funkcji jądra Gaussowskiej i równomiernej. Potwierdza to wcześniejsze analizy wpływu funkcji jądra na procedurę statystyczną dokonane w przypadku estymacji funkcji gęstości [1]. Świadczy to również o tym, że wybór parametrów metody jądrowej (funkcji jądra) ma znaczący wpływ na wynik procedury testowej. Funkcja jądra Gaussowska i równomierna wykorzystana w statystyce testowej testu Li zapewnia w wariancie I i II największą liczbę właściwych decyzji dotyczących odrzucenia hipotezy zerowej.

3.3. Etap III

W wariancie V eksperymentu rozkłady 2 analizowanych populacji są takie same. Decyzja prawidłowa polega na stwierdzeniu braku podstaw do odrzucenia hipotezy zerowej. Rozważano liczbę odrzuceń hipotezy zerowej (na 1000) dla

wszystkich funkcji jądra z etapu II oraz dla poziomu istotności 0,1 i 0,05. Wyniki przedstawiono w tab. 4.

Tabela 4. Wybrane charakterystyki testowe testu Li dla wariantu V

Funkcja jądra	LOT _{2,0,1}	LOT _{2,0,05}	I
Gaussowska	42	23	0,0112
Epanechnikova	85	44	0,0252
Równomierna	73	47	0,0149
Dwukwadratowa	81	48	0,0328
Trójkątna	79	48	0,0355
Trójkwadratowa	80	47	0,0392
Kosinusowa	82	44	0,0266

Źródło: obliczenia własne.

Najmniejsza liczba nieprawidłowych decyzji występuje dla funkcji jądra Gaussowskiej. Potwierdza to prawidłowość wyboru tej właśnie funkcji jądra w etapie I eksperymentu. Również dla funkcji jądra Gaussowskiej miara odległości I jest najmniejsza, co świadczy o dobrych własnościach estymatorów jądrowych wykorzystanych w procedurze testowej testu Li. Należy zwrócić uwagę, że we wszystkich rozważanych eksperymentach metoda wyboru parametru wygładzania jest taka sama (metoda Silvermana). Kierunki dalszych badań są związane z innymi metodami wyboru parametru wygładzania, np. z metodą *cross-validation*.

4. Wykorzystanie wybranych testów dla zmiennych połączonych do badania zgodności rozkładów wskaźnika percepcji korupcji

Badania zgodności rozkładów wskaźnika percepcji korupcji w latach 1996 oraz 2005 dokonano, stosując wybrane testy dla zmiennych połączonych. Dane o wielkości wskaźnika pochodzą ze stron internetowych organizacji Transparency International Polska i dotyczą 53 państw, które brały udział w badaniu zarówno w 1996 jak i 2005 r. Wartości wskaźnika z lat 1996 i 2005 stanowiące wartości prób są zatem powiązane, gdyż odnoszą się do elementów połączonych parami. Hipoteza o zgodności rozkładów wskaźnika percepcji korupcji w 2 badanych latach jest weryfikowana za pomocą następujących testów: testu zgodności Li ze statystyką testową T_2 (funkcja jądra Gaussowska i metoda Silvermana wyznaczenia parametru wygładzania), testu znaków, testu rangowanych znaków (zwanego również testem kolejności par Wilcoxona), testu znaków Fischera oraz testu Friedmana [2; 3; 4]. Weryfikacji hipotezy dokonano przy następujących poziomach istotności $\alpha = 0,1$, $\alpha = 0,05$ oraz $\alpha = 0,01$. Wyniki przedstawione są w tab. 5.

Tabela 5. Charakterystyki testowe oraz obszary krytyczne do badania zgodności rozkładów wskaźnika percepcji korupcji

Nazwa testu	Wartość statystyki testowej	Obszar krytyczny $\alpha = 0,1$	Obszar krytyczny $\alpha = 0,05$	Obszar krytyczny $\alpha = 0,01$
Test Li	0,0725	$(1,29; +\infty)$	$(1,64; +\infty)$	$(2,33; +\infty)$
Test znaków	53	$(-\infty; 20)$	$(-\infty; 18)$	$(-\infty; 17)$
Test rangowanych znaków	-0,678	$(-\infty; -1,64)$ $(1,64; +\infty)$	$(-\infty; -1,96)$ $(1,96; +\infty)$	$(-\infty; -2,58)$ $(2,58; +\infty)$
Test znaków Fischera	0,6868	$(-\infty; -1,64)$ $(1,64; +\infty)$	$(-\infty; -1,96)$ $(1,96; +\infty)$	$(-\infty; -2,58)$ $(2,58; +\infty)$
Test Friedmana	0,4717	$(2,706; +\infty)$	$(3,841; +\infty)$	$(6,635; +\infty)$

Źródło: obliczenia własne.

Weryfikacja hipotezy o zgodności rozkładów wskaźnika percepcji korupcji w latach 1996 oraz 2005 prowadziła do następującego wniosku: przy wszystkich analizowanych poziomach istotności brak jest podstaw do odrzucenia hipotezy zerowej. Test zgodności Li oraz pozostałe nieparametryczne testy dla zmiennych połączonych dały takie same wyniki. Różnica między rozkładami wskaźnika percepcji korupcji w latach 1996 i 2005 jest nieistotna statystycznie.

Literatura

- [1] Baszczyńska A., *Some Remarks on the Choice of the Kernel Function in Density Estimation*, Acta Universitatis Lodzensis, Folia Oeconomica 194, Łódź 2005, s. 143-149.
- [2] Domański C., *Statystyczne testy nieparametryczne*, PWE, Warszawa 1979.
- [3] Domański C., *Teoretyczne podstawy testów nieparametrycznych i ich zastosowanie w naukach ekonomiczno-społecznych*, Acta Universitatis Lodzensis, Wydawnictwo Uniwersytetu Łódzkiego, Łódź 1986.
- [4] Domański C., *Testy statystyczne*, PWE, Warszawa 1990.
- [5] Domański C., Pruska K., *Nieklasyczne metody statystyczne*, PWE, Warszawa 2000.
- [6] Pagan A., Ullah A., *Nonparametric Econometrics*, Cambridge University Press, Cambridge 1999.
- [7] Silverman B.W., *Density Estimation for Statistics and Data Analysis*, Chapman and Hall, London 1996.
- [8] Strona internetowa Transparency International Polska <http://www.transparency.pl>.

A NOTE ON THE LI GOODNESS OF FIT TEST USING KERNEL METHOD

Summary

Kernel method is used, first of all, in nonparametric estimation of density function. There are statistical tests where nonparametric kernel estimator of density function is applied in construction of test's statistic. They are, among others, goodness of fit tests (including normality tests), independence tests and symmetry ones.

In the paper, construction and properties' examination of goodness of fit test, proposed by Li, are presented. On the basis of simulation study the trial of choosing the best test's statistic of Li test is made. The analysis of influence of kernel method's parameter (kernel function) on the result of test's procedure is also regarded.

Li test and some others nonparametric goodness of fit tests are used in examination of consistence of distributions of index of perception corruption in two years: 1996 and 2005.