

*Aleksandra Baszczyńska\**

## TEST SYMETRYCZNOŚCI LI

**Streszczenie.** Test symetryczności rozkładu zmiennej losowej zaproponowany przez Li w 1997 roku jest testem wykorzystującym metodę jądrową. Dlatego też w procedurze wnioskowania statystycznego pojawia się problem wyboru odpowiednich parametrów metody jądrowej: funkcji jądra i parametru wygładzania. Autorka analizuje wpływ wyboru funkcji jądra i parametru wygładzania na procedurę weryfikacji hipotezy o symetryczności rozkładu zmiennej losowej. Test symetryczności Li jest wykorzystany w analizie wskaźnika rozwoju społecznego (Human Development Index).

**Słowa kluczowe:** test symetryczności, metoda jądrowa.

### I. WPROWADZENIE

Symetryczność rozkładu zmiennej losowej może podlegać weryfikacji przy wykorzystaniu, między innymi, nieparametrycznych testów statystycznych. Analiza symetryczności rozkładu stanowić może odrębne zagadnienie badawcze, jak również może być elementem analizy wstępnej związanej ze sprawdzaniem założeń wymaganych przy wyborze metody badawczej. Gdyż symetryczność rozkładu zmiennej losowej jest wymagana w wielu procedurach wnioskowania statystycznego (por. Hollander, 1976).

Test statystyczny zaproponowany przez Li w 1997 roku jest jądrowym testem symetryczności. Metoda jądrowa, zastosowana po raz pierwszy w estymacji funkcji gęstości, jest metodą znajdującą zastosowanie również, na przykład, w estymacji dystrybuanty zmiennej losowej, estymacji kwantyli, oraz w weryfikacji hipotez statystycznych (zgodności rozkładu, niezależności dwóch lub więcej zmiennych losowych). W metodzie jądrowej konieczny jest wybór odpowiedniej funkcji jądra oraz parametru wygładzania. Hardle (1990) przedstawia następujące funkcje jądra  $K(u)$ :

| Funkcja jądra | $K(u)$                      |
|---------------|-----------------------------|
| 1             | 2                           |
| Jednostajna   | $\frac{1}{2} I( u  \leq 1)$ |
| Trójkątna     | $(1 -  u ) I( u  \leq 1)$   |

---

\* Dr, Katedra Metod Statystycznych, Uniwersytet Łódzki

| 1              | 2                                                             |
|----------------|---------------------------------------------------------------|
| Epanecznikowa  | $\frac{3}{4}(1-u^2)I( u  \leq 1)$                             |
| Kwadratowa     | $\frac{15}{16}(1-u^2)^2 I( u  \leq 1)$                        |
| Trójkwadratowa | $\frac{35}{32}(1-u^2)^3 I( u  \leq 1)$                        |
| Gaussowska     | $\frac{1}{\sqrt{2\pi}} \exp\left(-\frac{1}{2}u^2\right)$      |
| Kosinusowa     | $\frac{\pi}{4} \cos\left(\frac{\pi}{2}u\right) I( u  \leq 1)$ |

gdzie  $I(A)$  jest indykatorem:  $I(A)=1$  gdy  $A$  zachodzi, natomiast  $I(A)=0$  gdy  $A$  nie zachodzi.

Powyższe funkcje jądra są symetryczne i spełniają warunek  $\int_{-\infty}^{+\infty} K(u)du = 1$ .

Są to tzw. klasyczne funkcje jądra.

Metodami wyboru parametru wygładzania  $h$  najczęściej stosowanymi w praktyce ze względu na ich prostotę są zasady praktyczne Silvermana (por. Silverman, 1996):

| Praktyczna zasada                 | $h$                                                                 |
|-----------------------------------|---------------------------------------------------------------------|
| Praktyczna zasada I               | $1,06 \min\left(\hat{\sigma}, \frac{\hat{R}}{1,34}\right) n^{-1/5}$ |
| Zmodyfikowana praktyczna zasada I | $1,06\hat{\sigma}n^{-1/z}$                                          |

gdzie:

$\hat{\sigma}$  jest odchyleniem standardowym szacowanym z próby  $X_1, \dots, X_n$ ,

$\hat{R}$  jest rozstępem międzykwartkowym szacowanym z próby  $X_1, \dots, X_n$ ,

$z$  jest dowolną liczbą naturalną,

$X_1, \dots, X_n$  jest próba prostą wylosowaną z populacji o nieznannej funkcji gęstości.

## II. TEST SYMETRYCZNOŚCI LI

Niech  $x_1, \dots, x_n$  będzie realizacją  $n$ -elementowej próby prostej  $X_1, \dots, X_n$  wylosowanej z populacji o nieznannej funkcji gęstości  $f(x)$ .

Hipoteza zerowa oraz hipoteza alternatywna w teście symetryczności Li sformułowane są w następujący sposób:

$$H_0: f(x) = f(-x), \quad (1)$$

$$H_1: f(x) \neq f(-x). \quad (2)$$

Li zaproponował następującą miarę dopasowania:

$$I = \frac{1}{2} \int_{-\infty}^{+\infty} (f(x) - f(-x))^2 dx = \int_{-\infty}^{+\infty} (f(x) - f(-x)) dF(x) \quad (3)$$

Jako estymator miary (3) wykorzystywany jest:

$$\begin{aligned} \tilde{I} = \int_{-\infty}^{+\infty} (\hat{f}(x) - \hat{f}(-x)) d\hat{F}(x) &= \frac{1}{n^2 h} \sum_{i=1}^n \sum_{j=1, j \neq i}^n \left[ K\left(\frac{x_i - x_j}{h}\right) - K\left(\frac{x_i + x_j}{h}\right) \right] + \\ &+ \frac{1}{n^2 h} \sum_{i=1}^n \left[ K(0) - K\left(\frac{2x_i}{h}\right) \right] = \tilde{I}_1 + \tilde{I}_2 \end{aligned} \quad (4)$$

gdzie  $\hat{f}(x) = \frac{1}{nh} \sum_{i=1}^n K\left(\frac{x_i - x}{h}\right)$  jest estymatorem jądrowym gęstości z funkcją jądra  $K(u)$  i parametrem wygładzania  $h$ .

Statystyki testowe w teście symetryczności Li są następujące:

$$T = n\sqrt{h} \frac{\left( \tilde{I} - \frac{K(0)}{nh} \right)}{\hat{\sigma}_1}, \quad (5)$$

$$T_1 = n\sqrt{h} \frac{\tilde{I}_1}{\hat{\sigma}_1} \quad (6)$$

gdzie:

$$\hat{\sigma}_1 = \frac{1}{n} \sum_{i=1}^n \hat{f}(x) \int_{-\infty}^{+\infty} K^2(u) du$$

Można wykazać (por. Pagan, 1999), że przy prawdziwości hipotezy zerowej statystyki (5) i (6) mają rozkład normalny standaryzowany.

W celu zbadania własności testu symetryczności Li przeprowadzono badanie symulacyjne. Rozważano następujące warianty populacji, z której pochodzi próba prosta:

- wariant I: rozkład normalny  $N(0,1)$  – rozkład symetryczny,
- wariant II: rozkład gamma  $(0.5, 2)$  – rozkład asymetryczny J-kształtny,
- wariant III: rozkład gamma  $(15, 2)$  – rozkład umiarkowanie asymetryczny,
- wariant IV: rozkład gamma  $(3, 1)$  – rozkład silnie asymetryczny,
- wariant V: rozkład gamma  $(100, 2)$  – rozkład o słabej asymetrii.

Z powyższych populacji losowano w sposób niezależny próby o liczebnościach 10, 20, 30, 40, 50, 60, 70, 80, 90 oraz 100. Na podstawie próby wyznaczano wartości statystyki testowej (5) i (6). Losowanie powtarzano 1000 razy. Zliczano te przypadki gdzie na poziomie istotności  $\alpha=0.1$  oraz  $\alpha=0.05$  odrzucono hipotezę zerową o symetryczności rozkładu populacji.

Liczbę powtórzeń w badaniu ustalono na 1000.

W statystyce testowej wykorzystano 7 funkcji jądra (jednostajną, trójkątną, Epanecznikową, kwadratową, trójkwadratową, Gaussowską i kosinusową) oraz parametr wygładzania  $1,06\hat{\sigma}n^{-1/5}$  dla  $z=5$  i 8. Przykładowe wyniki dla próby 30 elementowej są następujące:

Tablica 1. Liczba odrzuceń hipotezy zerowej o symetryczności rozkładu zmiennej losowej dla próby 30-elementowej oraz dla  $\alpha=0.1$

| Wariant | Parametr wygładzania | Funkcja jądra |      |      |      |      |      |      |
|---------|----------------------|---------------|------|------|------|------|------|------|
|         |                      | JED           | TKT  | EPA  | KW   | TKD  | GAU  | KOS  |
| I       | $z = 5$              | 73            | 70   | 82   | 65   | 35   | 85   | 40   |
|         | $z = 8$              | 68            | 65   | 70   | 57   | 28   | 76   | 38   |
| II      | $z = 5$              | 880           | 812  | 920  | 856  | 903  | 997  | 960  |
|         | $z = 8$              | 789           | 756  | 911  | 805  | 879  | 956  | 872  |
| III     | $z = 5$              | 987           | 890  | 1000 | 936  | 915  | 1000 | 935  |
|         | $z = 8$              | 819           | 780  | 1000 | 890  | 917  | 1000 | 998  |
| IV      | $z = 5$              | 1000          | 1000 | 1000 | 1000 | 1000 | 1000 | 1000 |
|         | $z = 8$              | 1000          | 1000 | 1000 | 1000 | 1000 | 1000 | 1000 |
| V       | $z = 5$              | 987           | 890  | 1000 | 876  | 850  | 1000 | 950  |
|         | $z = 8$              | 907           | 814  | 1000 | 817  | 876  | 1000 | 911  |

gdzie: JED oznacza funkcję jądra jednostajną, TKT oznacza funkcję jądra trójkątną, EPA oznacza funkcję jądra Epanecznikową, KW oznacza funkcję jądra kwadratową, TKD oznacza funkcję jądra trójkwadratową, GAU oznacza funkcję jądra Gaussowską, KOS oznacza funkcję jądra kosinusową.

Źródło: Obliczenia własne.

Analogiczne wyniki otrzymano dla innych wielkości prób.

Na podstawie powyższych wyników można stwierdzić, iż w rozważanych wariantach badania symulacyjnego zastosowanie funkcji jądra Gaussowskiej i Epanecznikowa dała przybliżone wyniki. Może to oznaczać podobne własności wygładzania w przypadku tych dwóch funkcji jądra. W rozważanych sytuacjach funkcje jądra inne niż Gaussowska i Epanecznikowa spowodowały mniejsze liczby odrzuconych hipotez o symetryczności rozkładu populacji, z której losowane były próby. Oznaczać to może lepsze własności testu wykorzystującego metodę jądrową (z funkcją jądra inną jądro najczęściej stosowane w praktyce, czyli Gaussowskie i Epanecznikowa. Dla rozkładu silnie asymetrycznego (wariant IV) w każdym przypadku odrzucano hipotezę zerową o symetryczności. Wystąpienie słabej asymetrii (wariant V) spowodował jedynie w przypadku funkcji jądra Gaussowskiej i Epanecznikowa odrzucenie hipotezy o symetryczności.

Wybór parametru wygładzania ma również wpływ na wynik weryfikacji hipotezy (1). Dla zmodyfikowanej praktycznej zasady Silvermana dla  $z=8$  obserwujemy mniejszą liczbę odrzuceń hipotez niż w przypadku  $z=5$ . W przypadku każdej rozważanej w badaniu funkcji jądra praktyczna zasada (dla  $z=8$ ) dała lepsze wyniki niż klasyczna zasada (dla  $z=5$ ).

### **III. WERYFIKACJA HIPOTEZY O SYMETRYCZNOŚCI ROZKŁADU WSKAŹNIKA ROZWOJU SPOŁECZNEGO (HUMAN DEVELOPMENT INDEX)**

Test Li zastosowano do badania symetryczności rozkładu wskaźnika rozwoju społecznego (HDI) dla 2006 roku.

Wskaźnik rozwoju społecznego opracowany został przez Prof. Mahbuba ul Haq'a w 1990 roku. Od 1993 roku wykorzystywany jest przez ONZ (Program Narodów Zjednoczonych do spraw Rozwoju UNDP).

Określa on poziom życia, uwzględnia zarówno poziom ubóstwa, stopień analfabetyzmu oraz poziom opieki zdrowotnej. W 2006 roku do obliczenia syntetycznego wskaźnika HDI wykorzystywane były następujące wskaźniki cząstkowe:

- oczekiwana długość życia przy narodzeniu,
- średnia ważona odsetka dorosłych umiejących czytać i pisać oraz odsetka skolaryzacji na wszystkich poziomach edukacji.
- produkt krajowy brutto per capita.

Liderami rankingu w poprzednich latach były państwa wysoko rozwinięte (Norwegia, Kanada, Japonia).

Wskaźnik Rozwoju Społecznego w 2006 roku wyznaczono dla 177 państw: 175 członków ONZ oraz Hongkong i Autonomię Palestyńską. Nie wszystkie państwa dostarczyły odpowiednich danych, dlatego w rankingu zabrakło 16 państw członkowskich ONZ, m.in. Afganistanu, Andory, Iraku, Korei Północnej, Liechtensteinu, Monako, Serbii i Czarnogóry, Somalii. W rankingu Polska zajęła 36. miejsce, w grupie krajów wysoko rozwiniętych. Nasi sąsiedzi: Niemcy 20., Czechy 31., Słowacja 42., Ukraina 78., Białoruś 67., Litwa 39., Rosja – 62. miejsce. Spośród 32 państw słabo rozwiniętych, 30 leży w Afryce.

W badaniu przyjęto założenie, że wartość HDI jest zmienną losową. Analizując dane dotyczące badań HDI w okresie kilkunastu lat, można zauważyć, że liczba państw biorących udział w badaniu ulega zmianom. Jest to spowodowane, między innymi, zmianami politycznymi (powstanie nowych państw) oraz tym, że niektóre państwa rezygnują z udziału w badaniach. Należy spodziewać się również zmian w liczbie i strukturze państw biorących udział w badaniu w przyszłości. Coroczny raport dotyczący rankingu państw z punktu widzenia wielkości HDI wskazuje na brak symetrii rozkładu analizowanej zmiennej losowej. Odrzucenie hipotezy zerowej o symetryczności rozkładu HDI jest, zatem oczekiwaną decyzją w procesie wykorzystania testu symetryczności Li w analizie HDI.

W celu weryfikacji hipotezy o symetryczności rozkładu wskaźnika rozwoju społecznego w 2006 roku losowo pobrano próby o liczebności 10, 30, 50, oraz 100. Przy liczbie powtórzeń 1000 zliczano liczbę przypadków kiedy test symetryczności Li odrzucał hipotezę o symetryczności rozkładu. Dla badanych liczebności prób liczba odrzuconych hipotez wynosiła 1000. Również gdy 177 państw traktowana była jak próba, weryfikacja hipotezy o symetryczności była odrzucona na poziomie istotności 0,05.

Powyższa analiza przeprowadzona przy zastosowaniu metod symulacyjnych oraz przykład zastosowania testu Li w badaniu symetryczności HDI świadczy o bardzo dobrych własnościach analizowanego testu. Może on być traktowany jako wygodne narzędzie statystyczne w analizach związanych z badaniem rozkładu zmiennej losowej. Metoda jądrowa staje się coraz bardziej popularna, szczególnie w badaniach ekonomiczno-społecznych i finansowych, a co za tym idzie, istnieje oprogramowanie związane z metodą jądrową. Warte podkreślenia jest zarówno: łatwość obliczeń oraz uniwersalność metody jądrowej.

#### BIBLIOGRAFIA

- Hardle W., (1990), *Smoothing Techniques with Implementation in S*, Springer-Verlag  
Hollander M., Wolfe D., (1976), *Nonparametric Statistical Methods*, Wiley Series in Probability and Statistics  
Pagan A., Ullah A., (1999), *Nonparametric Econometrics*, Cambridge University Press  
Silverman B. W., (1996), *Density Estimation for Statistics and Data Analysis*, Chapman & Hall  
<http://www.economicdynamics.org/codes/razzak.prg>

---

*Aleksandra Baszczyńska***SYMMETRY TEST LI****Abstract**

Symmetry test proposed by Li in 1997, is a test that uses kernel method. In the statistical inference using kernel methods, we need to determine the kernel function and the smoothing parameter. The author analyzes the influence of choice of the kernel function and smoothing parameter for a procedure of verification the hypothesis of random variable distribution symmetry. Li symmetry test is also used in the analysis of Human Development Index.