A C T A U N I V E R S I T A T I S L O D Z I E N S I S
FOLIA OECONOMICA 285, 2012

Aleksandra Baszczyńska[*]

SOME REMARKS ON THE SYMMETRY KERNEL TEST[1]

Abstract. The paper presents chosen statistical tests used to verify the hypothesis of the symmetry of random variable's distribution. Detailed analysis of the symmetry kernel test is made. The properties of the regarded symmetry kernel test are compared with the other symmetry tests using Monte Carlo methods. The symmetry tests are used, as an example, in analysis of the distribution of the Human Development Index (HDI).

Key words: kernel method, symmetry, Li symmetry test, triple test, Gupta symmetry test.

I. INTRODUCTION

In nonparametric hypothesis testing, there is a group of statistical tests, based on wildly accepted measure of global distance between two density functions of the random variable $X : f(x)$ and $g(x)$. One of these measures of closeness is the integrated squared error:

$$I = I(f(x), g(x)) = \int_{-\infty}^{+\infty} (f(x) - g(x))^2 \, dx \tag{1}$$

The expected value of I is the following:

$$M = E(I) = \int_{-\infty}^{+\infty} (f(x) - g(x))^2 \, f(x) dx$$

and can be estimated by:

$$\widehat{M} = \frac{1}{n} \sum_{i=1}^{n} \left(\hat{f}(x_i) - g(x_i) \right)^2 , \tag{2}$$

[*] Ph.D., Chair of Statistical Methods, University of Łódź.

[1] The research was supported by the project number DEC-2011/01/B/HS4/02746 from the National Science Centre.

where:

– $x_1, x_2, ..., x_n$ are realizations of independent, identically distributed random variable X with the unknown density $f(x)$,

– $\hat{f}(x)$ is nonparametric kernel estimator of $f(x)$.

The asymptotic distribution of estimator of I has been analyzed by Bickel, Rosenblat and Hall (see: Pagan A., Ullah A., 1999).

The statistic (2) can be used, for example, in (see: Belaire-Franch J., Contreras D., 2002; Ekstrom M., Jammalamadaka S., 2007; Henze N., Klar B., Meintanis S., 2003):

– testing whether density has a particular form: $H_0 : f(x) = g(x)$, against $H_1 : f(x) \neq g(x)$, where $g(x)$ has, for example, normal distribution $N(\mu, \sigma)$,

– testing the independence between two variables X and Y: $H_0 : f(x,y) = f(x)f(y)$, against $H_1 : f(x,y) \neq f(x)f(y)$,

– testing the symmetry around zero: $H_0 : f(x) = f(-x)$, against $H_1 : f(x) \neq f(-x)$.

II. LI SYMMETRY TEST

Let $f(x)$ denote the continuous density function of a random variable X, and let $x_1, x_2, ..., x_n$ be the observations from $f(x)$. Let $\hat{f}(x)$ denote the kernel density estimator of $f(x)$:

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

where $K(u)$ is kernel function and h is smoothing parameter.

For the testing the symmetry with:

$H_0 : f(x) = f(-x)$ and $H_1 : f(x) \neq f(-x)$,

Li in 1997 (see: Pagan A., Ullah A., 1999) proposed the following form of the integrated squared error:

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

and its estimator:

$$\widetilde{I} = \int_{-\infty}^{+\infty} \left(\widehat{f}(x) - \widehat{f}(-x)\right) d\widehat{F}(x) = \frac{1}{n^2 h} \sum_{i=1}^{n} \sum_{j=1, j\neq i}^{n} \left[K\left(\frac{x_i - x_j}{h}\right) - K\left(\frac{x_i + x_j}{h}\right) \right] + \tag{4}$$

$$+ \frac{1}{n^2 h} \sum_{i=1}^{n} \left[K(0) - K\left(\frac{2x_i}{h}\right) \right] = \widetilde{I}_1 + \widetilde{I}_2$$

where:

$$\widetilde{I}_1 = \frac{1}{n^2 h} \sum_{i=1}^{n} \sum_{j=1, i\neq j}^{n} \left[K\left(\frac{x_i - x_j}{h}\right) - K\left(\frac{x_i + x_j}{h}\right) \right], \tag{5}$$

$$\widetilde{I}_2 = \frac{1}{n^2 h} \sum_{i=1}^{n} \left[K(0) - K\left(\frac{2x_i}{h}\right) \right]. \tag{6}$$

Under the assumption of $h \to 0$ and $nh \to \infty$, it is possible to show that under H_0:

$$T = n\sqrt{h} \frac{\left(\widetilde{I} - \frac{K(0)}{nh}\right)}{\widehat{\sigma}_1} \sim N(0,1) \tag{7}$$

and

$$T_1 = n\sqrt{h} \frac{\widetilde{I}_1}{\widehat{\sigma}_1} \sim N(0,1), \tag{8}$$

where

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

III. CLASSICAL NONPARAMETRIC SYMMETRY TESTS

Li symmetry test is the nonparametric one, assuming that the density function of population is unknown. The group of classical nonparametric includes also, for example: Gupta symmetry test and triple test.

3.1. Gupta symmetry test

Let $X_1, \ldots, X_n$ be independent, identically distributed random variable X with the unknown continuous density $f(x)$, and $x_1, x_2, \ldots, x_n$ are realizations of random variable X. Let θ denote unknown median of this distribution. The null hypothesis that the population is symmetric about θ is of the form:

$$H_0 : P(X \le \theta + b) + P(X \le \theta - b) = 1 \text{ for all } b.$$

It means that $H_0 : P(0 < X - \theta < b) = P(-b < X - \theta < 0)$ for all $b > 0$.

The test statistic (see: Hollander M., Wolfe D., 1976):

$$J = \frac{8(A_1 - A_2) - 1}{4n(n-1)} \cdot \left\{ \frac{1 + 3\left(1 - \frac{2t}{A_3}\right)^2}{12n} \right\}^{-\frac{1}{2}} \sim N(0,1), \tag{9}$$

where:

– $A_1 = \sum_{j=2}^{n} \sum_{i=1}^{j-1} \delta_{ij}$,

– $A_2 = \frac{\left[\frac{(n-2)}{2}\right]\left[\frac{n}{2}\right]}{2}$,

– $A_3 = \frac{\max\left\{1, \sum_{i=1}^{n} a_i\right\}}{2n^{\frac{4}{5}}}$,

– $[x]$ denotes the largest integer less than or equal to x,

– $\delta_{ij} = \begin{cases} 1 & \text{for} \quad X_i + X_j > 2Me \\ 0 & \text{for} \quad X_i + X_j \le 2Me \end{cases}$, for $j = 2, \ldots, n \quad i = 1, \ldots, j-1$.

– Me is the sample median,

– $a_i = \begin{cases} 1 & \text{for} \quad Me - n^{-\frac{1}{5}} \le X_i \le Me + n^{-\frac{1}{5}} \\ 0 & \text{otherwise} \end{cases}$ for $i = 1, \ldots, n$.

For the alternative hypothesis $H_1 : P(0 < X - \theta < b) \le P(-b < X - \theta < 0)$ for all $b > 0$, we reject H_0 if $J \ge u_\alpha$, accept H_0 if $J < u_\alpha$.

For the alternative hypothesis $H_1 : P(0 < X - \theta < b) \ge P(-b < X - \theta < 0)$ for all $b > 0$, we reject H_0 if $J \ge -u_\alpha$, accept H_0 if $J > -u_\alpha$.

For the alternative hypothesis $H_1 : P(0 < X - \theta < b) \ne P(-b < X - \theta < 0)$ for at least one positive b, we reject H_0 if $J \ge -u_{\alpha_1}$ or $J \le -u_{\alpha_2}$, accept H_0 if $-u_{\alpha_1} < J < -u_{\alpha_2}$, where $\alpha = \alpha_1 + \alpha_2$.

3.2. Triple test

Let $X_1, \ldots, X_n$ be independent, identically distributed random variable X with the unknown continuous density $f(x)$, and $x_1, x_2, ..., x_n$ are realizations of random variable X.

Taking all possible triples from the sample ($\binom{n}{3}$ combinations), it is possible to say that a triple of observations is skewed to the right if the middle observation is closer to the smaller observation than it is to the larger. The null hypothesis is $H_0 : \eta = 0$.

The triple test statistic, which asymptotic distribution is standard normal, is given by:

$$T = \frac{\hat{\eta}}{\sqrt{\frac{\hat{\sigma}_\eta^2}{n}}} \sim N(0,1), \tag{10}$$

where:

- $\hat{\eta} = \frac{1}{\binom{n}{3}} \sum_{i,j,k} f^*(X_i, X_j, X_k)$,
- $f^*(X_i, X_j, X_k) = \frac{1}{3}\left[sign(X_i + X_j - 2X_k) + sign(X_i + X_k - 2X_j) + sign(X_j + X_k - 2X_i)\right]$,
- $sign(a) = \begin{cases} 1 & \text{for} \quad a > 0 \\ -1 & \text{for} \quad a < 0 \\ 0 & \text{for} \quad a = 0 \end{cases}$,

– $\widehat{\sigma}_{\hat{\eta}}^2 = \frac{1}{\binom{n}{3}} \sum_{c=1}^{3} \binom{3}{c} \binom{n-3}{3-c} \widehat{\varsigma}_c$,

– $\hat{\varsigma}_1 = \frac{1}{n} \sum_{i=1}^{n} (f_1^*(X_i) - \hat{\eta})^2$,

– $\hat{\varsigma}_2 = \frac{1}{\binom{n}{2}} \sum\sum_{j<k} \left(f_2^*(X_j, X_k) - \hat{\eta}\right)^2$,

– $\hat{\varsigma}_3 = \frac{1}{9} - \hat{\eta}^2$,

– $f_1^*(X_i) = \frac{1}{\binom{n-1}{2}} \sum\sum_{\substack{j<k \\ i \neq k, j \neq k}} f^*(X_i, X_j, X_k)$,

– $f_2^*(X_i, X_k) = \frac{1}{n-2} \sum\sum_{\substack{i=1 \\ i \neq j \neq k}} f^*(X_i, X_j, X_k)$.

Outstanding advantage of the triple test is its insensitivity to outliers in the sample but its drawback is the assumption of the independence of the data.

IV. COMPARISON OF THE SYMMETRY TESTS

A study was conducted to compare three, regarded above, nonparametric symmetric tests. The analysis was done using five variants of populations, from which the samples were drawn.

The variants are the following:

A: normally standardized distributed population – $N(0,1)$, symmetric distribution,

B: gamma distributed population – $G(0,5;2)$, asymmetric distribution, J-shaped, chi-squared distribution with 1 degree of freedom,

C: gamma distributed population – $G(15;2)$, asymmetric distribution, chi-squared distribution with 30 degrees of freedom – moderate asymmetry,

D: gamma distributed population – $G(3;1)$, asymmetric distribution – strong asymmetry,

E: gamma distributed population – $G(100;2)$, asymmetric distribution, chi-squared distribution with 200 degrees of freedom – weak asymmetry.

From these populations, samples are drawn (n = 10, 30, 50, 100). Each study was repeated 1000 times. On the base of the samples, test statistics were computed. The cases where the null hypothesis (of the symmetry of the distribution) was rejected were calculated using α=0.1 and 0.05.

In the kernel test were used:

– the Gaussian kernel: $K(u) = \frac{1}{\sqrt{2\pi}} \exp\left(-\frac{1}{2} u^2 \right)$,

$h = 1.06 \hat{\sigma} n^{-\frac{1}{5}}$ – practical rule of choosing the smoothing parameter:

where

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

The results of the conducted study are presented in the tables.

Table 1. Number of decisions of rejection of null hypothesis of symmetry (α = 0.1) for 1000 repetition

Variant	Size of sample	Kernel test	Gupta test	Triple test
A	10	62	44	61
	30	85	84	102
	50	75	80	108
	100	94	84	117
B	10	520	323	657
	30	997	774	997
	50	1000	903	1000
	100	1000	990	1000
C	10	1000	35	81
	30	1000	127	216
	50	1000	177	289
	100	1000	205	302
D	10	1000	88	198
	30	1000	266	639
	50	1000	241	831
	100	1000	483	857
E	10	1000	0	86
	30	1000	85	115
	50	1000	115	138
	100	1000	139	200

Source: own calculations.

Table 2. Number of decisions of rejection of null hypothesis of symmetry (α=0.05) for 1000 repetition

Variant	Size of sample	Kernel test	Gupta test	Triple test
A	10	46	17	27
	30	64	46	59
	50	58	38	53
	100	71	38	61
B	10	416	161	657
	30	995	672	997
	50	1000	845	1000
	100	1000	979	1000
C	10	1000	15	40
	30	1000	67	135
	50	1000	108	197
	100	1000	136	213
D	10	1000	88	117
	30	1000	266	501
	50	1000	241	747
	100	1000	385	804
E	10	1000	0	39
	30	1000	40	69
	50	1000	66	69
	100	1000	78	119

Source: own calculations.

For two regarded values of α , the best results, in most cases, are for the kernel test. Even for small sample's sizes the number of rejections of the null hypothesis when it is not true is very big. It means that the kernel test is characterized by very good properties. In comparison with classical nonparametric test (Gupta test and triple test) the number of proper decisions was, nearly always, bigger in the case of kernel test, especially for big sample size. In the case of variant A, where the population is symmetric the number of wrong decisions is the smallest for kernel test.

Additionally, in simulation study, one more variant was regarded, where population consists of 187 values of Human Development Index (in 2011year). From this population some samples were drawn (with 1000 repetitions), the test's statistics were computed and the number of rejections of null hypothesis were calculated. The results are the following:

Table 3. Number of decisions of rejection of null hypothesis of symmetry (α=0.05) for 1000 repetition

Size of sample	Kernel test	Gupta test	Triple test
10	1000	0	129
30	1000	1	690
50	1000	20	872
100	1000	395	967
187	reject H_0	reject H_0	reject H_0

Source: own calculations.

It appeared that only in the case of kernel test the number of rejections were the same for different sample sizes. The similar results are in the case of Gupta test but only for very big sample (n = 100). When all the population was taking into account, the results for three tests in detecting asymmetry were the same. The distribution of HDI is not symmetric.

REFERENCES

Belaire-Franch J., Contreras D., (2002), A Pearson's Test for Symmetry with an Application to the Spanish business Cycle, Spanish Economic Review, 4, 221–238

Ekstrom M., Jammalamadaka S., (2007), An Asymptotically Distribution-free Test of Symmetry, Journal of Statistical Planning and Inference, 137, 799–810

Henze N., Klar B., Meintanis S., (2003), Invariant Tests for Symmetry about an Unspecified Point Based on the Empirical Characteristic Function, Journal of Multivariate Analysis, 87, 275–297

Hollander M., Wolfe D., (1976), *Nonparametric Statistical Methods*, Wiley Series in Probability and Statistics

Pagan A., Ullah A., (1999), Nonparametric Econometrics, Cambridge University Press

http://www.economicdynamics.org/codes/razzak.prg

Aleksandra Baszczyńska

UWAGI O JĄDROWYM TEŚCIE SYMETRYCZNOŚCI

W pracy przedstawiono wybrane statystyczne testy wykorzystywane w weryfikacji hipotezy o symetryczności rozkładu zmiennej losowej. Szczegółowej analizie poddano test symetryczności oparty o metodę jądrową. Porównano własności zaprezentowanych testów symetryczności oraz zastosowano je go analizy rozkładu wskaźnika rozwoju społecznego (HDI).