Journal of Statistical Planning and Inference 140 (2010) 1788–1804

Contents lists available at ScienceDirect

Journal of Statistical Planning and Inference

journal homepage: www.elsevier.com/locate/jspi

Asymptotic normality of two symmetry test statistics based on the L_1-error

Noureddine Berrahou, Djamal Louani *

L.S.T.A., Université de Paris 6, 175, rue du Chevaleret, 8ème étage, bâtiment A, 75013 Paris, France

A R T I C L E I N F O

Article history:
Received 16 June 2008
Received in revised form
7 October 2009
Accepted 11 January 2010
Available online 20 January 2010

MSC:
60F10
62G07
62F05
62H15

Keywords:
Density estimation
Histogram
Kernel
Asymptotic normality
Symmetry test

A B S T R A C T

In this paper, we propose two new tests to test the symmetry of a distribution. These tests are built up on the asymptotic normality of the L_1-distance to the symmetry of the Kernel and histogram density estimates. A simulation study is carried out to evaluate performances of the kernel based test.

1. Introduction

Let $X_1,X_2,\ldots,X_n$ be a sequence of i.i.d. real random variable defined on a probability space $(\Omega,\mathcal{F},P)$ with distribution function F and density function f with respect to the Lebesgue measure μ. Several methods have been introduced in the literature to estimate density functions. Only the histogram and kernel methods are considered in the study dealt with in this paper. In order to estimate the density f with the histogram method, let $\mathcal{P}_n$ be a partition of $\mathbb{R}$ into intervals $A_{n,j}, j \in \mathbb{Z}$, of size $h_n>0$, where $\mathbb{Z}$ is the set of relative integers. The histogram estimator of the density f is defined, for any $x \in \mathbb{R}$, by

$$f_{n,1}(x)=\frac{\Lambda_n(A_{n,j})}{h_n} \quad \text{whenever } x \in A_{n,j},$$

where Λ_n is the empirical probability measure associated to the sample $X_1,X_2,\ldots,X_n$.

The kernel density estimator was introduced for the first time by Rosenblatt (1956) and have been extensively studied in the literature since then. The kernel estimator of the density f is defined, for any $x \in \mathbb{R}$, by

$$f_{n,2}(x)=\frac{1}{nh_n}\sum_{i=1}^{n}K\left(\frac{x-X_i}{h_n}\right),$$

where K is a function satisfying some regularity conditions (often K is a density itself) and (h_n) is a sequence of bandwidths.

* Corresponding author.
E-mail addresses: noureddine.berrahou@upmc.fr (N. Berrahou), djamal.louani@upmc.fr (D. Louani).

0378-3758/$ - see front matter © 2010 Elsevier B.V. All rights reserved.
doi:10.1016/j.jspi.2010.01.006

The L_1-distance to the symmetry of a density f is defined by

$$V=\int |f(x)-f(-x)|\,dx.$$

On the basis of observations $X_1,X_2,\ldots,X_n$, V may be estimated by the following quantities:

$$V_{n,1}=\int |f_{n,1}(x)-f_{n,1}(-x)|\,dx$$

and

$$V_{n,2}=\int |f_{n,2}(x)-f_{n,2}(-x)|\,dx.$$

Our aim in this paper is to study the asymptotic normality of the statistics $V_{n,1}$ and $V_{n,2}$ and to build up testing procedures to test the null hypothesis

$$H_0: \text{“ } V=0\text{” versus the alternative } H_1: \text{“ } V>0\text{” }.$$

The symmetry hypothesis relative to a distribution may be expressed in various forms. Either the case involving the density and the L_1-distance that we study here, the symmetry about a known median v, using the distribution function F, may be defined by the following hypothesis:

$$H_0': \text{“ } 1-F(v-x)-F(v+x)=0, \quad x\geq 0\text{” }.$$

A number of statistics have been proposed to test the hypothesis H_0' or any other related hypothesis against the general alternative that the distribution is not symmetric or narrower classes of alternatives. Note that the paper of Berrahou and Louani (2006) gives an overview relative to the most studied ones. Their list may be completed with the paper by Ahmad and Li (1997), Mizushima and Nagao (1998) and Dette et al. (2002) where the L_2-distance is considered when testing the symmetry of a density and a regression function.

2. Results

Throughout the paper, assume that the density f is a symmetric function.

2.1. Histogram method

Theorem 1. *Assume that for any $j\in\mathbb{Z}$, $A_{n,j}=-A_{n,-j}$ and $\sqrt{n}h_n\to\infty$, $a_n\to 0$ as $n\to\infty$. Then,*

$$\frac{\sqrt{n}}{\sigma_1}(V_{n,1}-\mathbb{E}(V_{n,1}))\xrightarrow{\mathcal{D}}\mathrm{N}(0,1)\quad \text{as } n\to\infty$$

and

$$\lim_{n\to\infty} n\,Var(V_{n,1}-\mathbb{E}(V_{n,1}))=\sigma_1^2,$$

where $\sigma_1^2=4(1-2/\pi)$ and $\xrightarrow{\mathcal{D}}$ denotes the convergence in distribution.

2.2. Kernel method

Consider the following hypothesis upon the kernel K that will be used below.

($\mathbf{H}_1$) K is a symmetric probability density having support contained in the closed ball of radius $\frac{1}{2}$ centered at zero and is bounded by a constant κ.

Let Z=(Z_1,Z_2) be a bidimensional random vector distributed according to the Gaussian law $N_2(0,I)$ where I is the identity matrix. In the sequel, the L_2-norm of the kernel K is denoted by $\|K\|_2$. For any $t\in\mathbb{R}$, define the function

$$\rho(t):=\frac{\int_{\mathbb{R}}K(u)K(u+t)\,du}{\|K\|_2^2}.$$

Set

$$\sigma_2^2=4\|K\|_2^2\int_{-1}^{1}Cov(|\sqrt{1-\rho^2(t)}Z_1+\rho(t)Z_2|,|Z_2|)\,dt.$$

Theorem 2. *Assume that the kernel K satisfies the hypothesis* ($\mathbf{H}_1$). *If $\sqrt{n}h_n\to\infty$, $h_n\to 0$ as $n\to\infty$, then, we have*

$$\frac{\sqrt{n}}{\sigma_2}(V_{n,2}-\mathbb{E}(V_{n,2}))\xrightarrow{\mathcal{D}}\mathrm{N}(0,1)\quad \text{as } n\to\infty$$