1. Introduction

The problem of goodness-of-fit testing is one of the central themes of statistical theory and practice because it is always important to verify the correspondence between the mathematical model and the observed real data. The classical approaches to this hypothesis testing problem can be found in many books (see, e.g., Durbin [4], Greenwood and Nikulin [6] or Lehmann and Romano [10]). The most known tests are: Pearson’s goodness-of-fit Chi-Squared test, Kolmogorov-Smirnov and Cramér-von Mises tests. We recall the construction of the Cramér-von Mises test. Suppose that we observe \( n \) independent identically distributed random variables \( (X_1, \ldots, X_n) = X^n \) with continuous distribution function \( F(x) \) and the basic hypothesis is simple: \( F(x) = F_0(x) \). Then we can introduce the Cramér-von Mises statistic,
where \( \hat{F}_n(x) \) is the empirical distribution function, and after the transformation \( t = F_0(x) \), we get the standard form of the statistic \( W^2_n \) and its limit distribution \( W^2 \) (under \( \mathcal{H}_0 \))

\[
W^2_n = n \int_{-\infty}^{\infty} [\hat{F}_n(x) - F_0(x)]^2 \, dF_0(x), \quad \hat{F}_n(x) = \frac{1}{n} \sum_{i=1}^{n} 1_{\{X_i < x\}},
\]

where \( \hat{F}_n(x) \) is the empirical distribution function, and after the transformation \( t = F_0(x) \), we get the standard form of the statistic \( W^2_n \) and its limit distribution \( W^2 \) (under \( \mathcal{H}_0 \))

\[
W^2_n = n \int_{0}^{1} (\hat{F}_n(t) - t)^2 \, dt = \int_{0}^{1} \beta^2_n(t) \, dt \quad \implies \quad W^2 = \frac{1}{2} \int_{0}^{1} \beta^2(t) \, dt.
\]

Here \( \beta(t) \) is Brownian bridge (Gaussian process with \( \mathbb{E}\beta(t) = 0 \) and \( K(t, s) = \mathbb{E}\beta(t)\beta(s) = t \land s - ts \)).

The hypothesis \( \mathcal{H}_0 \) is accepted, if \( W^2_n < C_{n, \alpha} \) and rejected, if \( W^2_n \geq C_{n, \alpha} \), where the constant \( C_{n, \alpha} \) is solution of the equation \( \mathbb{P}(W^2 \geq C_{n, \alpha}) = \alpha \). The constant \( C_{n, \alpha} \) can be approximated by its limit \( (n \to \infty) \) value \( c_{\alpha} \), solution of the equation \( \mathbb{P}(W^2 \geq c_{\alpha}) = \alpha \). It is possible to expand \( \beta(\cdot) \) into the Karhunen–Loève series (see Ash and Gardner [1], pp. 38–43)

\[
\beta(t) = \sum_{k=1}^{\infty} \frac{\sqrt{7}}{\pi k} \xi_k \sin(\pi kt), \quad t \in [0, 1], \quad W^2 \overset{d}{=} \sum_{k=1}^{\infty} \frac{\xi_k^2}{(\pi k)^2}, \quad \xi_k \overset{\text{i.i.d.}}{\sim} \mathcal{N}(0, 1).
\]

These relations are used to approximate the value \( c_{\alpha} \). More about Karhunen–Loève theory and the use of Karhunen–Loève expansion in the problems of goodness-of-fit testing can be found in Martynov [11] and Deheuvels and Martynov [3].

In this Note, we study the similar problem of goodness-of-fit testing, when the basic model is an ergodic diffusion process and the corresponding statistic is of Cramér–von Mises type.

Let \( X^T = [X_t, 0 \leq t \leq T] \) be an observed trajectory of continuous time ergodic diffusion process

\[
dX_t = S(X_t) \, dt + dW_t, \quad X_0, \quad 0 \leq t \leq T,
\]

where \( X_0 \) is its initial value, independent of the Wiener process \( \{W_t, t \geq 0\} \). The trend coefficient \( S(\cdot) \) is supposed to be unknown to the observer and he has to verify if \( S(x) = S_0(x) \), where \( S_0(x) \) is some known function. The Cramér–von Mises type statistics, based on the observation \( X^T \) solution of (1) can be:

\[
V^2_T = T \int_{-\infty}^{\infty} (\hat{F}_T(x) - F_{S_0}(x))^2 \, dx, \quad W^2_T = T \int_{-\infty}^{\infty} (\hat{f}_T(x) - f_{S_0}(x))^2 \, dx,
\]

where \( \hat{F}_T(x) \) and \( \hat{f}_T(x) \) are some estimates of the distribution function \( F_{S_0}(x) \) of the invariant density \( f_{S_0}(x) \) respectively. These statistics converge in distribution (under the null hypothesis) to quadratic functionals of Gaussian processes (see Kutoyants [9]). However, due to the structure of the covariance of these processes, the choice of the thresholds \( c_{\alpha} \) for the tests \( S_T = 1_{\{V^2_T \geq C_{n, \alpha}\}}, \rho_T = 1_{\{W^2_T \geq C_{n, \alpha}\}} \) is much more complicated (see Dachian and Kutoyants [2]).

In the present work we are interested by the case when the trend coefficient \( S_0(\cdot) \) of the process (1) under hypothesis takes just two values \(+1\) and \(-1\) (simple switching \( S_0(x) = -\text{sgn}(x) \)), i.e., is discontinuous function (see Kutoyants [8], for the further details). We study the limit distribution of the corresponding Cramér–von Mises type statistic based on the empirical density estimator (local time estimator) of the invariant density. The main tool is the K-L expansion of the limit Gaussian process. Finally, using this expansion and numerical simulation we calculate the threshold of this test.

2. Simple switching diffusion

Let us introduce the simple switching diffusion process

\[
dX_t = -\text{sgn}(X_t) \, dt + dW_t, \quad X_0, \quad 0 \leq t \leq T,
\]

i.e., \( S_0(x) = -\text{sgn}(x) \). It is easy to see that the conditions of the existence of the solution and ergodicity (see, e.g., conditions \( \mathcal{E}S \) and \( \mathcal{A}_0 \) in Kutoyants in [9]), are fulfilled and (2) is an ergodic diffusion process with the stationary density \( f_{S_0}(x) = e^{-2|x|} \), for \( x \in \mathbb{R} \).

We observe a trajectory \( X^T \) of (1) and we have to test the following hypotheses:

\( \mathcal{H}_0: \quad S(\cdot) = S_0(\cdot), \quad \mathcal{H}_1: \quad S(\cdot) \neq S_0(\cdot). \)

To test this null hypothesis we use the following Cramér–von Mises type statistic:

\[
W^2_T = \int_{-\infty}^{\infty} \beta^2_T(x) \, dx = T \int_{-\infty}^{\infty} \left( f^2_T(x) - f_{S_0}(x) \right)^2 \, dx, \quad f^2_T(x) = \frac{2\rho_T(x)}{T},
\]
where $\mathcal{L}(x)$ is the local time of the diffusion process and $\tilde{f}_n(x)$ is the local time estimator of the invariant density (see, Kutoyants in [9]).

Fix a number $\alpha \in (0, 1)$ and define the class $\mathcal{K}_\alpha$ of tests of asymptotic level $1 - \alpha$ as follows:

$$\mathcal{K}_\alpha = \left\{ \phi_T : \limsup_{T \to \infty} E_{\mathcal{H}_0} \phi_T (X^T) \leq \alpha \right\}.$$ 

Let us introduce the Gaussian process $(\zeta_f(x), x \in \mathbb{R})$, with zero mean and covariance function

$$R_f(x, y) = 2 \left( 1_{\{x, y < 0\}} e^{2(x+y)} + 1_{\{x, y > 0\}} e^{-2(x+y)} \right) - (2(|x| + |y|) + \text{sgn}(xy))e^{-2(|x|+|y|)}.$$ 

Denote by $d_\alpha$ the value defined by the equation

$$P \left( \int_{-\infty}^{\infty} \zeta_f^2(x) \, dx > d_\alpha \right) = P(W^2 > d_\alpha) = \alpha.$$ 

**Theorem 2.1.** The Cramér–von Mises type test $\Phi_T(X^T) = 1_{[W^2 > d_\alpha]}$ belongs to $\mathcal{K}_\alpha$.

**Proof.** The main idea of the proof is that under $\mathcal{H}_0$ the distribution of $W^2_2$ converge to the distribution of $W^2$. In fact by using Theorem A22, in Ibragimov–Khasminskii [7], we show that for any constant $L > 0$, the distribution of $\int_{|x| \leq L} \zeta_f^2(x) \, dx$ convergence to the distribution of $\int_{|x| > L} \zeta_f^2(x) \, dx$. Then we show that $\int_{|x| > L} \zeta_f^2(x) \, dx$ tends to zero as $L \to +\infty$. For more details see [5]. □

Let $J_n(\cdot)$ and $Y_n(\cdot)$ be the Bessel functions of first and second kind of order $n$, respectively, with $n = 0, 1, 2, \ldots$.

**Proposition 2.2.** The Gaussian process $(\zeta_f(x), x \in \mathbb{R})$ has a Karhunen–Loève expansion given by

$$\zeta_f(x) = \sum_{n=1}^{\infty} 2 \xi_n \frac{\bar{\alpha}(v_{2,n})e^{-|x|} J_1(v_{2,n}e^{-|x|}) - J_2(v_{2,n})(e^{-|x|}Y_1(v_{2,n}e^{-|x|}) + \frac{2}{\pi v_{2,n}})}{J_1(v_{2,n})\bar{\alpha}(v_{2,n}) - (v_{2,n})J_2(v_{2,n})} - \sum_{n=1}^{\infty} \frac{2 \xi_n^2 \text{sgn}(x)e^{-|x|} J_1(v_{1,n}e^{-|x|})}{J_0(v_{1,n})},$$

where $\xi_n, n \geq 1$, $\xi_n^2, n \geq 1$ denote two independent sequences of i.i.d. $\mathcal{N}(0, 1)$ random variables and $v_{1,n}, v_{2,n}, n = 1, 2, \ldots$, are respectively the positive zeros of $J_1(v_{1,n})$ and $J_1(v_{2,n})$ defined by the equation $f(v_n) = f_2(v_n) - f_1(v_n)$, with $f_1(v_n) = Y_0(v_n) - \frac{1}{\pi} \ln(v_n/2) + \gamma$, $f_2(v_n) = Y_2(v_n) + \frac{4}{\pi v_n^2}$, and $\gamma = 0.577215 \ldots$ the Euler’s constant.

<table>
<thead>
<tr>
<th>alpha</th>
<th>threshold</th>
</tr>
</thead>
<tbody>
<tr>
<td>0.01</td>
<td>1.5008</td>
</tr>
<tr>
<td>0.05</td>
<td>2.4202</td>
</tr>
<tr>
<td>0.025</td>
<td>3.4332</td>
</tr>
<tr>
<td>0.010</td>
<td>4.8678</td>
</tr>
<tr>
<td>0.005</td>
<td>6.0036</td>
</tr>
</tbody>
</table>

[Fig. 1. Choice of threshold.]
Proof. The detailed proof is given in [5].

Note that the random variable $W^2$ is a weighted sum of independent $\chi^2_1$ components. We make the truncation of this sum keeping $10^7$ terms and using the numerical simulation of the corresponding Gaussian random variables we obtain values for the threshold $d_\alpha$ as shown in Fig. 1. □

Acknowledgements

The author is very grateful to Professor Marina Kleptsyna for essential help in correcting the first version of this result, and to Professor Yury Kutoyants for suggesting this problem, and fruitful discussions.

References