ON SOME STATISTICAL TESTS FOR RANDOMNESS

Yu.S. Kharin, I.V. Snihir

Department of Mathematical Modeling and Data Analysis, Belarus State University, Minsk, Belarus, e-mail: *kharin@bsu.by*

Abstract. This paper presents statistical tests for randomness. We consider the problems of testing whether or not a strictly stationary sequence is an independent sequence and a given permutation is a random permutation. We construct the test based upon an estimator of a "distance to whiteness" introduced by Drouiche (1994) and the test based on the Tracy-Widom distribution of the length of a longest monotonically increasing subsequence of a permutation. Asymptotic distributions of the test statistics are found. Numerical experiments are carried out to emphasize the performance of the tests.

1 Introduction

Statistical testing of randomness is a significant problem in engineering, artificial intelligence, economics, medicine, etc. The paper is organised as follows. In Section 2, we propose a statistical test based on a measure introduced by Drouiche [3] to test for whiteness. Box and Jenkins (1970) introduced the so-called Portmanteau statistic, which has been intensively studied (see, for example, Box and Pierce [2]). The test statistic, used in this paper, is based upon the periodogram of the observed sequence, and the asymptotic distribution of it is derived under the null (white noise) and alternative hypotheses that will be emphasized by some computer simulations.

Section 3 is devoted to the Tracy-Widom distribution, which enable us to construct the test for randomness of the permutation. The test statistic is a length of the longest increasing subsequence in a random permutation. Asymptotic distributions of the length of the longest increasing subsequence are found in 1999 by Baik, Deift and Johansson [1]. The results of numerical experiments are also given.

2 Test for whiteness

Let $\{X_k\}_{k\in Z}$ be a strictly stationary random sequence with a spectral density $\varphi(\omega)$, $\omega\in[0,2\pi]$. The hypothesis H_0 means that the sequence $\{X_k\}_{k\in Z}$ is a sequence of independent identically distributed random variables, the so-called white noise. It is well known that $\{X_k\}_{k\in Z}$ is a white noise sequence if and only if

$$\varphi(\omega)=\frac{\gamma_0}{2\pi}\ \text{for all}\ \omega\in[0,2\pi], \tag{1}$$

where γ_0 is a positive constant. The alternative $H_1=\overline{H}_0$ means that (1) is broken. The problem is in statistical testing H_0, H_1 by N observations $X_1,\dots,X_N$.

We will use the following functional introduced by Drouiche [3]:

$$W(\varphi)=\log\frac{1}{2\pi}\int_{-\pi}^{\pi}\varphi(\omega)d\omega-\frac{1}{2\pi}\int_{-\pi}^{\pi}\log\varphi(\omega)d\omega. \tag{2}$$

It has the following properties:

1) $W(\varphi)\ge 0$;

2) $W(\varphi)=0$ if and only if $\varphi=c$, where c is some positive constant;

3) $W(\lambda\varphi)=W(\varphi)$ for any real $\lambda>0$.

Using the above properties, we can represent H_0, H_1 in the form:

$$H_0 : \varphi = c > 0; \quad H_1 : \varphi \neq c. \tag{3}$$

According to [3] introduce the statistic:

$$\hat{W}_N = \log\frac{1}{2\pi}\int_{-\pi}^{\pi} I_N(\omega)d\omega - \frac{1}{2\pi}\int_{-\pi}^{\pi}\log I_N(\omega)d\omega - \lambda, \tag{4}$$

where $\omega \in [0,2\pi]$, λ denotes the *Euler* constant, $I_N(\omega)$ is a periodogram of the observations:

$I_N(\omega) = \dfrac{1}{2\pi N}\left|\sum_{t=1}^{N} X_t e^{it\omega}\right|^2$ - the estimator of the spectral density $\varphi(\omega)$.

Note, that if $N \to \infty$, then

$$\hat{W}_N\Big|_{H_0} \xrightarrow{a.s.} 0, \quad \hat{W}_N\Big|_{H_1} \xrightarrow{a.s.} W(\varphi) \neq 0.$$

To construct the test, we need to derive the distributions of $\hat{W}_N$ under H_0 and H_1, which also enable us to choose the significance level and to derive the power of the proposed test.

Theorem 1. *Let* $\{X_k\}_{k\in Z}$ *be a strictly stationary process with the spectral density* $\varphi(\omega)$, *then as* $N \to \infty$ *under the null hypothesis* H_0 $(\varphi(\omega) = \frac{\gamma_0}{2\pi} > 0)$

$$\sqrt{N}\frac{\hat{W}_N}{\sqrt{2\cdot(\pi^2/6-1)}} \xrightarrow{D} N(0,1); \tag{5}$$

under the alternative hypothesis H_1 $(\varphi(\omega) \neq \frac{\gamma_0}{2\pi})$

$$\sqrt{N}(\hat{W}_N - W) \xrightarrow{D} N\left(0, \left(2\cdot(\pi^2/6-1)\right) + \frac{1}{\pi}\int_{-\pi}^{\pi}\left(1 - 2\pi\varphi(\omega)/\gamma_0\right)^2 d\omega\right). \tag{6}$$

Remark. The asymptotic distributions (5), (6) correct the results proposed in *Theorem 1* [3].

Define the statistic:

$$T_N = \sqrt{N}\frac{|\hat{W}_N|}{\sqrt{2\cdot(\pi^2/6-1)}}. \tag{7}$$

Corollary 1. *The test*

$$H_0 : T_N < \Delta; \quad H_1 : T_N \geq \Delta, \tag{8}$$

where $\Delta = \Phi^{-1}\left(1 - \frac{\varepsilon_0}{2}\right)$, $\Phi(x)$ *is the standard normal distribution function, has the asymptotic* $(N \to \infty)$ *significance level* ε_0.

Corollary 2. *The asymptotic expression for the power of the test* (8) *is*

$$P_{\varepsilon_0} = 1 - \Phi\left(-\sqrt{N}\frac{W}{\sigma_1} + \Delta\frac{\sigma_0}{\sigma_1}\right) \xrightarrow[N\to\infty]{} 1, \tag{9}$$

where $\sigma_0 = \sqrt{2\cdot(\pi^2/6-1)}$, $\sigma_1 = \sqrt{(2\cdot(\pi^2/6-1)) + \frac{1}{\pi}\int_{-\pi}^{\pi}(1-2\pi\varphi(\omega)/\gamma_0)^2\, d\omega}$.

Thus, for any given significance level ε_0 and a sample size *N*, (9) gives us an estimator of the power we could expect from test (8).

Numerical experiments

The alternative situation will be an autoregressive process of order one [AR(1)]. We will consider several values of the unique parameter for this case in order to emphasize the sensibility of the test for whiteness. Consider the case:

$$\begin{cases} H_0 : \{X_t\}_{t\in Z} \text{ is a white noise,} \\ H_1 : \{X_t\}_{t\in Z} \text{ is an AR(1)-process,} \end{cases}$$

where AR(1) means the autoregressive process $\{X_k\}_{k\in Z}$ generated by the following model:

$$X_t - aX_{t-1} = \varepsilon_t,\ |a|<1,\ a\neq 0.$$

According to Theorem 1, under H_1 the statistic $\hat{W}_N$ has the following asymptotic distribution as $N\to\infty$:

$$\sqrt{N}(\hat{W}_N - W) \xrightarrow{D} N\left(0, 2(\pi^2/6-1) + 4\cdot a^2/(1-a^2)\right),\ W = -\log(1-a^2),$$

and thus, the power (9) is given by

$$P_{\varepsilon_0} = 1-\Phi\left(\sqrt{(1-a^2)/2|a|}\left(\sqrt{N}\log(1-a^2) + \Delta\sqrt{2\cdot(\pi^2/6-1)}\right)\right). \quad (10)$$

Figures 1,2 present the theoretical results and the results of 100 simulations of the testing of the procedure (8) for each value of *a* and *N*.

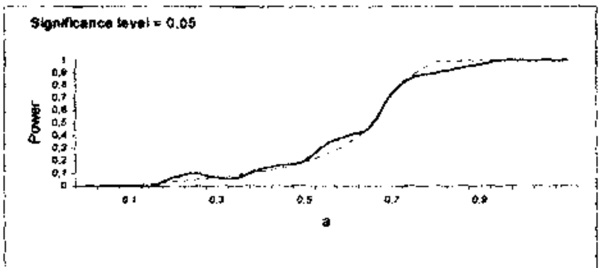

Fig.1. Empirical (solid line) and theoretical (dashed line) power of the test, N=64.

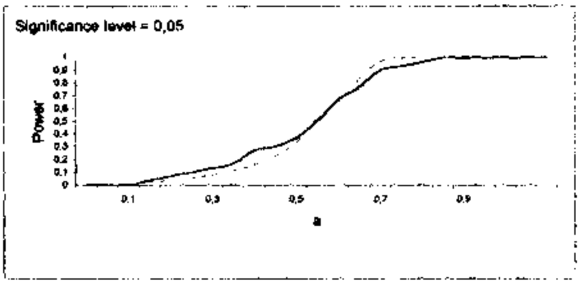

Fig.2. Empirical (solid line) and theoretical (dashed line) power of the test, N=128.

3 Test for randomness of a permutation

Consider the permutation group S_n on *{1,2,...,n}*. We say that $1\le i_1<\ldots<i_k\le n$ is an increasing subsequence of length *k* of $s=(s_1,\ldots,s_n)\in S_n$ iff $s(i_1)<\ldots<s(i_k)$. The length of the longest increasing subsequence of a permutation $s\in S_n$ will be denoted by $L_n := L_n(s)$. Let us define the null hypothesis $H_0 : s=(s_1,\ldots,s_n)$ is a random permutation, and the alternative one $H_1 = \overline{H_0}$; it means that under H_0 all *n!* permutations are equiprobable. The goal is to test H_0, H_1 by observations.

Note [4,5], that the asymptotic approximations of the expected value and variance of L_n as $n\to\infty$ are $E_{H_0}\{L_n\}\approx 2\sqrt{n}$, $D_{H_0}\{L_n\}\approx n^{1/3}$. Thus, let us introduce the statistic:

$$X_n = \frac{L_n - 2\sqrt{n}}{n^{1/6}}. \quad (11)$$

It is proven [1], that X_n converges in distribution as $n \to \infty$ to the Tracy-Widom distribution, introduced by Tracy and Widom in [6]. This distribution can be defined as a solution of the second order differential equation $u_{xx} = 2u^3 + xu$. It is known that

$$u(x) \approx -\frac{e^{-(2/3)x^{3/2}}}{2\sqrt{\pi}x^{1/4}},\ x \to \infty.$$

Then the Tracy-Widom distribution has the distribution function

$$F(t) = \exp\left(-\int_t^\infty (x-t)u^2(x)dx\right), \quad (12)$$

$$1 - F(t) \approx \frac{e^{-(4/3)t^{3/2}}}{16\pi t^{3/2}} \text{ as } t \to \infty.$$

Using completely different mathematical methods, it was possible to observe the similarity between the results of the following asymptotics (regimes) [5]:

$$\lim_{t\to\infty}\lim_{n\to\infty}\frac{\log P\{L_n > (2+tn^{-1/3})\sqrt{n}\}}{t^{3/2}} = -\frac{4}{3};\ \lim_{t\to 0}\lim_{n\to\infty}\frac{\log P\{L_n > (2+t)\sqrt{n}\}}{t^{3/2}n^{1/2}} = -\frac{4}{3};$$

$$\lim_{n\to\infty}\frac{\log P\{L_n > (2+tn^{-\eta})\sqrt{n}\}}{n^{(1-3\eta)/2}t^{3/2}} = -\frac{4}{3},\ 0 < \eta < \frac{1}{3},\ t > 0.$$

Let us define the statistic, on the basis of which we will construct a new test

$$T_n = \frac{L_n - 2\sqrt{n}}{n^{\frac{1}{2}-\eta}}, \quad (13)$$

where η determines the type of asymptotics (regime).

Theorem 2. *If* $n \to \infty$, $0 \le \eta \le \frac{1}{3}$ *and the null hypothesis* H_0 *is true, then*

$$P_{H_0}\{T_n \le x\} \to F_*(x),$$

where $F_*(x) = 1 - \exp\left\{-\frac{4}{3}n^{(1-3\eta)/2}t^{3/2}\right\}$ *is the Tracy-Widom distribution function.*

Corollary. *At* $n \to \infty$ *the asymptotic size of the test*

$$H_0 : T_n < \Delta;\ \ H_1 : T_n \ge \Delta, \quad (14)$$

where $\Delta = \sqrt[3]{\frac{9}{16}n^{(1-3\eta)}}\log^{2/3}\varepsilon_0$, *coincides with the significance level* ε_0.

Numerical experiments

1) The numerical simulations were carried out when under H_0 was considered a random permutation $s = (s_1, \ldots, s_n)$, generated according to [7, §3.2] with the uniform distribution on S_n. We analysed two regimes where the statistic (13) was calculated at $\eta = 1/3$ and at $\eta = 0$, ε_0=0.05, n={8, 16, 32, 64, 128, 256}. The results of the experiments are the values of the statistic based on $L_n := L_n(s)$ and the frequency ν of the acceptance of the decision in favour of H_1. Table 1 presents the results of 100 simulations of the testing of

the procedure (14) for each value of η and n.

η	n					
	8	**16**	**32**	**64**	**128**	**256**
0	0.00	0.00	0.03	0.05	0.11	0.07
1/3	0.00	0.02	0.06	0.07	0.08	0.10

Table 1. Frequency ν of the acceptance of hypothesis H_1.

Analyzing the fact, that $\nu \to 0.05$ for any chosen regime and for all n, is possible to make a conclusion, that the examined permutation is "good" and the result, received by a numerical way, is coordinated with theoretical properties and confirms that the examined permutation is a random one.

2) Under H_1 we will define a permutation $s \in S_n$ as

$$\begin{pmatrix} 1 & 2 & 3 & \dots & a-1 & a & a+1 & \dots & n \\ a & a+1 & a+2 & \dots & 2a-2 & 2a-2 & 2a & \dots & a-2 \end{pmatrix},$$

it means that $s = (s_1, \dots, s_n)$ is a permutation received by a cyclic shift of the sequence $(1, \dots, n)$ on a positions to the left, $1 \le a \le n-1$. We analyzed two regimes where the statistic (13) was calculated at $\eta = 1/3$ and at $\eta = 0$ as well, ε_0=0.05, n={8, 16, 32, 64, 128, 256}, a={3, 5, 7}. Table 2 presents the results of the test (14).

η / a	n																	
	8			**16**			**32**			**64**			**128**			**256**		
	3	**5**	**7**	**3**	**5**	**7**	**3**	**5**	**7**	**3**	**5**	**7**	**3**	**5**	**7**	**3**	**5**	**7**
0	H_0	H_0	H_0	H_1	H_0	H_0	H_1	H_1	H_0	H_1	H_1	H_1	H_1	H_1	H_1	H_1	H_1	H_1
1/3	H_0	H_0	H_0	H_1	H_0	H_0	H_1	H_1	H_0	H_1	H_1	H_1	H_1	H_1	H_1	H_1	H_1	H_1

Table 2. The acceptance of hypothesis H_1 or H_0 for each value of η, n, a.

References

[1] J. Baik, P. Deift, K. Johansson. On the distribution of the length of the longest increasing subsequence of random permutations. *J. Amer. Math. Soc. 12 (1999), no.4, 1119-1178.*

[2] G. Box, D. Pierce. Distribution of residual autocorrelations in autoreggressive integrated moving average time series models. *J. Amer. Stat. Ass. 65 (1970), 1509-1526.*

[3] K. Drouiche. A new test for whiteness. *Transactions on signal processing 48 (2000), no.7, 1864-1871.*

[4] B. Logan, L. Shepp. A variational problem for random Young tableaux. *Advances in Mathematics 26 (1977), 206-222.*

[5] M. Lowe, F. Merkl. Moderate deviations for longest increasing subsequences: the upper tail. *Communications on Pure and Applied Mathematics (2001), vol. LIV, 1488-1520.*

[6] C.A. Tracy, H. Widom. Level-spacing distributions and the Airy kernel. *Comm. Math. Phys. 159 (1994), no.1, 151-174.*

[7] Yu. Kharin, S. Agievich. Computer practicum on mathematical methods of information processing. *Minsk, BSU, 2001(in Russian).*