Ann. Inst. Statist. Math.
39 (1987), Part A, 457-472

ON NONPARAMETRIC TESTS FOR SYMMETRY

SIGEO AKI

(Received May 28, 1986)

Summary

This paper is concerned with an extension of the problem of testing symmetry about zero of a distribution function. In order to obtain the asymptotic null distribution of test statistics for the problem, a limit theorem is proved, which indeed plays an essential role in the asymptotic theory of testing problem for symmetry.

1. Introduction

Let $X_1, X_2, \cdots, X_n$ be independent random variables with a common distribution function F. F_n denotes the empirical distribution function of the variables $X_1, X_2, \cdots, X_n$. The problem of testing F for symmetry about zero was investigated by several authors. In particular, Butler [3] and Rothman and Woodroofe [5] proposed test statistics based on the empirical distribution function. The statistics are

$$\sqrt{n} \sup_{x \leq 0} |F_n(x)+F_n(-x)-1|$$

and

$$n \int_{-\infty}^{\infty} [F_n(x)+F_n(-x)-1]^2 dF_n(x) ,$$

respectively.

The essential part of deriving the asymptotic null distributions is to show that the stochastic process

$$Q_n(x)=\sqrt{n}(F_n(x)+F_n(-x)-1)$$

converges to a Gaussian process.

In Section 2 we shall prove a limit theorem which enables us to consider a general testing problem for symmetry, which includes the above problem as a particular case.

Key words and phrases: Wiener process, empirical distribution, goodness-of-fit test, test for symmetry, weak convergence.

Let $0<\alpha<1$ be a given constant. In Section 3 we shall study the problem of testing F for the hypothesis,

H_0: There exists a continuous distribution function G which is symmetric about zero and

$$F(x)=\begin{cases} 2\alpha G(x) & \text{if } x\leqq 0\,, \\ \alpha+2(1-\alpha)\left(G(x)-\frac{1}{2}\right) & \text{if } x>0\,, \end{cases}$$

holds.

If we set $\alpha=1/2$ in the hypothesis H_0, then H_0 means that F is symmetric about zero. Therefore we can regard the hypothesis H_0 as a natural extension of that of symmetry. In Section 4 we shall give an interpretation of the value α in a concrete example. Further more general problems will be discussed, which may be called tests for local property of a distribution function rather than those for symmetry.

2. Asymptotic behavior of a stochastic process

Let G_1 be a distribution function on $[0, 1]$. Suppose random variables $Y_1, Y_2, \cdots, Y_n$ are independent and have a common distribution function G_1. Let $\xi_1, \xi_2, \cdots, \xi_n$ be random variables. We assume also that $(Y_1, \cdots, Y_n)$ and $(\xi_1, \cdots, \xi_n)$ are independent. We define a random element of $D[0, 1]$ by

$$u_n(t)=\frac{1}{\sqrt{n}}\sum_{i=1}^{n}\xi_i I_{[0,t]}(Y_i)\,, \qquad 0\leqq t\leqq 1\,,$$

where $I_A(\cdot)$ denotes the indicator function of the set A.

THEOREM 2.1. *Suppose the random variables $\xi_1, \xi_2, \cdots, \xi_n$ are independent and identically distributed with mean 0 and variance 1. Then the random element $u_n(t)$ converges weakly to $W(G_1(t))$ in $D[0, 1]$ as $n\to\infty$, where $W(t)$ is a standard Wiener process.*

PROOF. We first prove the theorem under the assumption that $G_1(t)=t$. For any given numbers $a_1, \cdots, a_m$ and $0\leqq t_1<\cdots<t_m\leqq 1$, the central limit theorem yields that $\sum_{j=1}^{m} a_j u_n(t_j)$ converges in law to $\sum_{j=1}^{m} a_j\cdot W(t_j)$, since

$$\mathrm{E}\,\xi_i\sum_{j=1}^{m} a_j I_{[0,t_j]}(Y_i)=0$$

and

$$\mathrm{E}\left(\xi_i\sum_{j=1}^{m} a_j I_{[0,t_j]}(Y_i)\right)^2=\sum_{j=1}^{m}\sum_{k=1}^{m} a_j a_k (t_j\wedge t_k)\,.$$

Then we can see the convergence of the finite dimensional distributions of u_n by Cramér and Wold's theorem. Let us now show that $\{u_n\}$ is tight. For that, it suffices to prove that

$$\text{(2.1)}\quad E\{|u_n(t)-u_n(t_1)|^2\cdot|u_n(t_2)-u_n(t)|^2\}\leqq(t_2-t_1)^2, \qquad 0\leqq t_1\leqq t\leqq t_2\leqq 1.$$

Since

$$|u_n(t)-u_n(t_1)|^2\cdot|u_n(t_2)-u_n(t)|^2=\frac{1}{n^2}\left(\sum_{i=1}^{n}\xi_i I_{(t_1,t]}(Y_i)\right)^2\left(\sum_{j=1}^{n}\xi_j I_{(t,t_2]}(Y_j)\right)^2$$

and the independence of ξ's and Y's, the left hand side of (2.1) is

$$\frac{n-1}{n}(t-t_1)(t_2-t),$$

and hence (2.1) follows. Then the theorem holds under the assumption that $G_1(t)=t$.

Suppose now that $G_1(t)$ is an arbitrary distribution over [0, 1]. Let $\eta_1, \eta_2, \cdots, \eta_n$ be independent uniformly distributed random variables over [0, 1]. We further assume that $\{\eta_i\}$ and $\{\xi_i\}$ are independent. Then the distribution function of $G_1^{-1}(\eta_i)$ is G_1, where $G_1^{-1}(s)=\inf\{t: s\leqq G_1(t)\}$. Since the theorem involves the law of u_n, we may write $Y_i=G_1^{-1}(\eta_i)$. We define

$$V_n(t)=\frac{1}{\sqrt{n}}\sum_{i=1}^{n}\xi_i I_{[0,t]}(\eta_i).$$

Then $V_n(t)$ converges weakly to $W(t)$ in $D[0, 1]$ by the fact we already proved. Note that $u_n(t)=V_n(G_1(t))$. If we regard the time change by G_1 as a function from $D[0, 1]$ to $D[0, 1]$, the function is continuous on $C[0, 1]$ (cf. the proof of Theorem 16.4 of Billingsley [2]). Hence the theorem is now proved by Theorem 5.1 of Billingsley [2] and the fact that Wiener measure has its support on $C[0, 1]$.

In passing, we shall show a more general result about u_n when $Y_1, Y_2, \cdots, Y_n$ are independent and uniformly distributed over [0, 1]. Let us define a random element of $D([0, 1]^2)$ by

$$\text{(2.2)}\qquad u_n(s, t)=\frac{1}{\sqrt{n}}\sum_{i=1}^{[ns]}\xi_i I_{[0,t]}(Y_i), \qquad 0\leqq s,\ t\leqq 1,$$

where we denote, by $[a]$, the largest integer not exceeding a.

THEOREM 2.2. *Suppose $Y_1, Y_2, \cdots, Y_n$ are independent and uniformly distributed over* [0, 1]. *Let $\xi_1, \xi_2, \cdots, \xi_n$ be independent and identically distributed with mean* 0 *and variance* 1. *Suppose further that $(Y_1, \cdots, Y_n)$ and $(\xi_1, \cdots, \xi_n)$ are independent. Then the random element $u_n(s, t)$*

defined by (2.2) *converges weakly to* $B(s, t)$ *in* $D([0,1]^2)$ *as* $n \to \infty$, *where* $B(s, t)$ *is a Brownian sheet, i.e., a zero-mean Gaussian field on* $[0,1]^2$ *with the covariance structure*

$$\mathrm{E}\, B(s_1, t_1)B(s_2, t_2) = (s_1 \wedge s_2)\cdot(t_1 \wedge t_2) .$$

PROOF. We first show the convergence of finite dimensional distributions of $u_n(s, t)$. Suppose we are given m points $(s_1, t_1), \cdots, (s_m, t_m)$ in $[0,1]^2$. Without loss of generality we assume that $s_1 \leqq s_2 \leqq \cdots \leqq s_m$. For any given real constants $a_1, a_2, \cdots, a_m$, we consider the distribution of $\sum_{j=1}^m a_j u_n(s_j, t_j)$. Note that

$$(2.3) \quad \sum_{j=1}^m a_j u_n(s_j, t_j) = \frac{1}{\sqrt{n}} \sum_{i=1}^{[ns_1]} \xi_i (a_1 I_{[0,t_1]}(Y_i) + \cdots + a_m I_{[0,t_m]}(Y_i)) + \frac{1}{\sqrt{n}} \sum_{i=[ns_1]+1}^{[ns_2]} \xi_i (a_2 I_{[0,t_2]}(Y_i) + \cdots + a_m I_{[0,t_m]}(Y_i)) + \cdots + \frac{1}{\sqrt{n}} \sum_{[ns_{m-1}]+1}^{[ns_m]} \xi_i a_m I_{[0,t_m]}(Y_i) .$$

Since all terms of the right hand side of (2.3) are mutually independent, the central limit theorem implies that (2.3) converges in law to the normal distribution with mean 0 and variance

$$(2.4) \quad s_1 \sum_{j=1}^m \sum_{k=1}^m a_j a_k (t_j \wedge t_k) + (s_2 - s_1) \sum_{j=2}^m \sum_{k=2}^m a_j a_k (t_j \wedge t_k) + \cdots + (s_m - s_{m-1}) a_m^2 t_m .$$

It is easily checked that (2.4) is equal to

$$(2.5) \quad s_1 a_1^2 t_1 + 2a_1 s_1 \sum_{k=2}^m a_k (t_1 \wedge t_k) + s_2 a_2^2 t_2 + 2a_2 s_2 \sum_{k=3}^m a_k (t_2 \wedge t_k) + \cdots + s_{m-1} a_{m-1}^2 t_{m-1} + 2a_{m-1} s_{m-1} a_m (t_{m-1} \wedge t_m) + s_m a_m^2 t_m .$$

It is also easily seen that the variance of $\sum_{j=1}^m a_j B(s_j, t_j)$ is equal to (2.5). Therefore finite dimensional distributions of $u_n(s, t)$ converge to those of $B(s, t)$.

Next we shall prove that $\{u_n(s, t)\}$ is tight. For that, we shall check the moment condition for increments of u_n about all neighboring blocks.

Case 1. For $s_1 \leqq s \leqq s_2$ and $t_1 \leqq t$, we consider the neighboring blocks

$$B = (s_1, s] \times (t_1, t]$$

and

$$C = (s, s_2] \times (t_1, t] .$$

The increments of u_n about B and C are respectively written as

$$u_n(B)=u_n(s_1, t_1)-u_n(s_1, t)-u_n(s, t_1)+u_n(s, t)$$
$$=\frac{1}{\sqrt{n}}\sum_{i=[ns_1]+1}^{[ns]}\xi_i I_{(t_1,t]}(Y_i)\ ,$$

and

$$u_n(C)=\frac{1}{\sqrt{n}}\sum_{i=[ns]+1}^{[ns_2]}\xi_i I_{(t_1,t]}(Y_i)\ .$$

Then we have

$$\text{(2.6)}\qquad \mathrm{E}\,(u_n^2(B)u_n^2(C))=\mathrm{E}\,u_n^2(B)\,\mathrm{E}\,u_n^2(C)$$
$$=\frac{1}{n^2}([ns]-[ns_1])(t-t_1)\cdot([ns_2]-[ns])(t-t_1)$$
$$\leqq\left(\frac{[ns_2]-[ns_1]}{n}\right)^2(t-t_1)^2\ .$$

If $s_2-s_1\geqq 1/n$, then $[ns_2]-[ns_1]\leqq 2n(s_2-s_1)$. If $s_2-s_1<1/n$, then at least either $[ns]=[ns_1]$ or $[ns]=[ns_2]$ holds and hence the left hand side of (2.6) vanishes. Therefore we have

$$\mathrm{E}\,(u_n^2(B)u_n^2(C))\leqq 4(s_2-s_1)^2(t-t_1)^2\ .$$

Case 2. For $s_1\leqq s$ and $t_1\leqq t\leqq t_2$, consider the neighboring blocks

$$B=(s_1, s]\times(t_1, t]$$

and

$$C=(s_1, s]\times(t, t_2]\ .$$

Then the increments of u_n about B and C are respectively written as

$$u_n(B)=\frac{1}{\sqrt{n}}\sum_{i=[ns_1]+1}^{[ns]}\xi_i I_{(t_1,t]}(Y_i)\ ,$$

and

$$u_n(C)=\frac{1}{\sqrt{n}}\sum_{i=[ns_1]+1}^{[ns]}\xi_i I_{(t,t_2]}(Y_i)\ .$$

In this case we have

$$\text{(2.7)}\qquad \mathrm{E}\,(u_n^2(B)u_n^2(C))=\frac{1}{n^2}\sum_{i=[ns_1]+1}^{[ns]}\ \sum_{j=[ns_1]+1}^{[ns]}\ \sum_{k=[ns_1]+1}^{[ns]}\ \sum_{l=[ns_1]+1}^{[ns]}$$
$$\times\mathrm{E}\,\xi_i\xi_j\xi_k\xi_l I_{(t_1,t]}(Y_i)I_{(t_1,t]}(Y_j)I_{(t,t_2]}(Y_k)I_{(t,t_2]}(Y_l)$$
$$=\frac{1}{n^2}\sum_{i\neq j}\mathrm{E}\,\xi_i^2 I_{(t_1,t]}(Y_i)\cdot\mathrm{E}\,\xi_j^2 I_{(t,t_2]}(Y_j)$$

$$=\frac{1}{n^2}([ns]-[ns_1])([ns]-[ns_1]-1)(t-t_1)(t_2-t)\ .$$

If $s-s_1\geqq 1/n$, then $[ns]-[ns_1]\leqq 2n(s-s_1)$ and we have

$$\frac{1}{n^2}([ns]-[ns_1])([ns]-[ns_1]-1)\leqq 4(s-s_1)^2\ . \tag{2.8}$$

If $s-s_1<1/n$, then at least either $[ns]=[ns_1]$ or $[ns]=[ns_1]+1$ holds and the left hand side of (2.8) vanishes. Thus it holds that the left hand side of (2.7) is not greater than

$$4(s-s_1)^2(t-t_1)(t_2-t)\leqq 4((s-s_1)(t_2-t_1))^2\ .$$

Consequently we have, for any pair of neighboring blocks B and C,

$$\mathrm{E}\,((\min\{|u_n(B)|,|u_n(C)|\})^4)\leqq \mathrm{E}\,u_n^2(B)u_n^2(C)\leqq(2\mu(B\cup C))^2\ ,$$

where μ denotes Lebesgue measure on $[0, 1]^2$. This completes the proof by Theorem 3 of Bickel and Wichura [1].

3. A testing problem for symmetry

Suppose that G_1 is a continuous distribution function which is defined on $(-\infty, \infty)$ and symmetric about zero. Let $0<\alpha<1$ be a given constant. We assume that G_1 satisfies the hypothesis

H_0: There exists a continuous distribution function G which is symmetric about zero and

$$G_1(x)=\begin{cases} 2\alpha G(x) & \text{if}\quad x\leqq 0\ ,\\ \alpha+2(1-\alpha)\left(G(x)-\dfrac{1}{2}\right) & \text{if}\quad x>0\ ,\end{cases}$$

holds.

Let X be a random variable with distribution function G_1 and let H be a continuous and strictly increasing distribution function which is symmetric about zero. If we set $Y=H(X)$, then Y is a random variable on $(0, 1)$ with the distribution function,

$$\mathrm{P}\,(Y\leqq t)=\begin{cases} 2\alpha G(H^{-1}(t)) & \text{if}\quad 0<t\leqq\dfrac{1}{2}\ ,\\ \alpha+2(1-\alpha)\left(G(H^{-1}(t))-\dfrac{1}{2}\right) & \text{if}\quad \dfrac{1}{2}<t<1\ .\end{cases}$$

Now we consider, for a distribution function F defined on $(0, 1)$, the hypothesis

H_1: There exists a continuous distribution function G^* defined on $(0, 1)$ such that

$$F(t)=\begin{cases} \alpha G^*(2t) & \text{if } 0<t\leqq\frac{1}{2}, \\ \alpha+(1-\alpha)(1-G^*(2-2t)) & \text{if } \frac{1}{2}<t<1, \end{cases}$$

holds.

Then the distribution function of Y satisfies the hypothesis H_1. This is easily seen by putting $G^*(t)=2G(H^{-1}(t/2))$.

Conversely, let X be a $(-\infty, \infty)$-valued random variable with a distribution function G_2. Let H be some continuous and strictly increasing distribution function which is defined on $(-\infty, \infty)$ and symmetric about zero. Now we assume that the distribution function of $Y=H(X)$ satisfies H_1. Then we shall show that G_2 satisfies H_0. Note that

$$G_2(x)=\mathrm{P}\,(Y\leqq H(x))\ .$$

Hence we can write, for a continuous distribution function G^* defined on $(0, 1)$,

$$G_2(x)=\begin{cases} \alpha G^*(2H(x)) & \text{if } x\leqq 0, \\ \alpha+(1-\alpha)(1-G^*(2H(-x))) & \text{if } x>0. \end{cases}$$

If we set

$$G^{**}(x)=\begin{cases} \frac{1}{2}G^*(2H(x)) & \text{if } x\leqq 0, \\ 1-\frac{1}{2}G^*(2H(-x)) & \text{if } x>0, \end{cases}$$

then G^{**} is symmetric about zero and

$$G_2(x)=\begin{cases} 2\alpha G^{**}(x) & \text{if } x\leqq 0, \\ \alpha+2(1-\alpha)\left(G^{**}(x)-\frac{1}{2}\right) & \text{if } x>0, \end{cases}$$

holds.

Consequently, without loss of generality, we consider the testing problem whether a distribution function on $(0, 1)$ satisfies the hypothesis H_1.

Let $X_1, X_2, \cdots, X_n$ be independent random variables having a common continuous distribution function F on $(0, 1)$. We define, for $i=1, \cdots, n$,

$$Y_i=\begin{cases} 2X_i & \text{if } X_i\leqq\frac{1}{2}\,, \\ 2(1-X_i) & \text{if } X_i>\frac{1}{2}\,, \end{cases}$$

$$\xi_i=\begin{cases} \sqrt{\frac{1-\alpha}{\alpha}} & \text{if } X_i\leqq\frac{1}{2}\,, \\ -\sqrt{\frac{\alpha}{1-\alpha}} & \text{if } X_i>\frac{1}{2}\,, \end{cases}$$

and

$$u_n(t)=\frac{1}{\sqrt{n}}\sum_{i=1}^{n}\xi_i I_{[0,t]}(Y_i)\,,\qquad 0\leqq t\leqq 1\,. \tag{3.1}$$

Let us consider testing hypothesis H_1 by the test statistic

$$T_n=\sup_{0\leqq t\leqq 1}|u_n(t)|\,. \tag{3.2}$$

THEOREM 3.1. *If F satisfies the hypothesis H_1, then T_n converges weakly to $T=\sup_{0\leqq t\leqq 1}|W(t)|$.*

PROOF. For every i and $0<t<1$, we see

$$\begin{aligned} \mathrm{P}(Y_i\leqq t)&=\mathrm{P}\left(Y_i\leqq t,\, X_i\leqq\frac{1}{2}\right)+\mathrm{P}\left(Y_i\leqq t,\, X_i>\frac{1}{2}\right) \\ &=\mathrm{P}\left(X_i\leqq\frac{t}{2}\right)+\mathrm{P}\left(X_i\geqq 1-\frac{t}{2}\right)=\alpha G^*(t)+(1-\alpha)G^*(t) \\ &=G^*(t)\,, \end{aligned}$$

and

$$\mathrm{P}\left(\xi_i=\sqrt{\frac{1-\alpha}{\alpha}}\right)=\mathrm{P}\left(X_i\leqq\frac{1}{2}\right)=\alpha\,.$$

On the other hand, it holds that

$$\mathrm{P}\left(\xi_i=\sqrt{\frac{1-\alpha}{\alpha}},\, Y_i\leqq t\right)=\mathrm{P}\left(Y_i\leqq t,\, X_i\leqq\frac{1}{2}\right)=\mathrm{P}\left(X_i\leqq\frac{t}{2}\right)=\alpha G^*(t)\,.$$

Then we can see that ξ_i and Y_i are independent and of course we have that $(\xi_1,\cdots,\xi_n)$ and $(Y_1,\cdots,Y_n)$ are independent. Further it is easy to see that $\xi_1,\xi_2,\cdots,\xi_n$ are independent and identically distributed with mean 0 and variance 1. Thus ξ's and Y's satisfy the assumptions of Theorem 2.1. Therefore we have by the theorem that the process $u_n(t)$ defined by (3.1) converges to the process $W(G^*(t))$ in $D[0,1]$. Then the theorem follows since G^* is a continuous distribution function on $(0,1)$.

The asymptotic null distribution is the same as that of Butler's statistic. The distribution function of T is given by

$$\frac{4}{\pi}\sum_{k=0}^{\infty}\frac{(-1)^k}{2k+1}\exp\left\{-\frac{\pi^2(2k+1)^2}{8u^2}\right\},$$

(cf. e.g., Feller [4]).

For a level α_0, we adopt the set (c_0, ∞) as a critical region, where c_0 is determined by $\mathrm{P}(T>c_0)=\alpha_0$. Then the test is consistent, that is, the next result holds.

THEOREM 3.2. *If F does not satisfy the hypothesis H_1, $\mathrm{P}(T_n>c)$ converges to 1 for every $c>0$.*

PROOF. First, we show under the assumption that $F(1/2)=\alpha'\neq\alpha$. Note that

$$u_n(1)=\frac{1}{\sqrt{n}}\sum_{i=1}^{n}\xi_i$$

and

$$\mathrm{E}\,\xi_i=\frac{1}{\sqrt{\alpha(1-\alpha)}}(\alpha'-\alpha)\neq 0\,.$$

Then

$$\mathrm{P}\,(|u_n(1)|>c)\rightarrow 1$$

follows by the law of large numbers.

Next, we prove under the assumption that $F(1/2)=\alpha$. Since F does not satisfy the hypothesis H_1, there exist two continuous distribution functions G^* and G^{**} on $(0, 1)$ such that

$$G^*\neq G^{**}$$

and

$$F(t)=\begin{cases}\alpha G^*(2t) & \text{if } 0\leqq t\leqq \frac{1}{2}\\ \alpha+(1-\alpha)(1-G^{**}(2-2t)) & \text{if } \frac{1}{2}<t\leqq 1\end{cases}$$

holds. We define

$$F_n^1(t)=\frac{1}{\alpha n}\sum_{i=1}^{n}I_{[0,t]}(Y_i)\cdot I_{[0,1/2]}(X_i)$$

and

$$F_n^2(t)=\frac{1}{(1-\alpha)n}\sum_{i=1}^{n} I_{[0,t]}(Y_i)\cdot I_{(1/2,1]}(X_i)\ .$$

Then we can write

$$\frac{1}{\sqrt{\alpha}\ \sqrt{1-\alpha}}u_n(t)=\sqrt{n}\,(F_n^1(t)-F_n^2(t))\ .$$

Noting that

$$F_n^1(t)-F_n^2(t)=(F_n^1(t)-G^*(t))-(F_n^2(t)-G^{**}(t))+(G^*(t)-G^{**}(t))\ ,$$

it suffices to prove

$$\sup_{0\leq t\leq 1}|F_n^1(t)-G^*(t)|\to 0 \tag{3.3}$$

in probability and

$$\sup_{0\leq t\leq 1}|F_n^2(t)-G^{**}(t)|\to 0 \tag{3.4}$$

in probability. We denote by F_n the empirical distribution function of the variables $X_1, X_2, \cdots, X_n$. We define

$$G_n(t)=\begin{cases}\alpha F_n^1(2t) & \text{if}\quad 0\leqq t\leqq\dfrac{1}{2}\ ,\\[2ex] \alpha F_n^1(1)+(1-\alpha)(1-F_n^2(2-2t)) & \text{if}\quad \dfrac{1}{2}<t\leqq 1\ .\end{cases}$$

Then we can easily see that

$$F_n(t)=\begin{cases}G_n(t) & \text{if}\quad 0\leqq t\leqq\dfrac{1}{2}\ ,\\[2ex] G_n(t+)-\alpha(F_n^1(1)-1) & \text{if}\quad \dfrac{1}{2}<t\leqq 1\ .\end{cases} \tag{3.5}$$

From the Glivenko-Cantelli theorem, it holds that

$$\sup_{0\leq t\leq 1/2}|F_n(t)-F(t)| \tag{3.6}$$

converges to zero in probability. By (3.5), (3.6) can be written as

$$\alpha\cdot\sup_{0\leq t\leq 1}|F_n^1(t)-G^*(t)|\ ,$$

and hence (3.3) follows. If $1/2<t\leqq 1$, we have from (3.5)

$$F_n(t)-F(t)=(1-\alpha)(G^{**}(2-2t)-F_n^2(2-2t+))+\alpha(F_n^1(1)-1)\ .$$

Then we have

$$\sup_{0\leq t\leq 1}|F_n^2(t)-G^{**}(t)|=\sup_{1/2<t\leq 1}|G^{**}(2-2t)-F_n^2(2-2t)|$$

$$\leqq \frac{1}{1-\alpha} \sup_{1/2<t\leqq 1} |F_n(t)-F(t)| + \frac{\alpha}{1-\alpha}(F_n^1(1)-1) .$$

Therefore (3.4) follows from the Glivenko-Cantelli theorem and the law of large numbers. This completes the proof.

We have been considering testing the hypothesis H_1 with the statistic T_n defined by (3.2). However, we can take another statistic for the problem. For example, we can take

$$T_n^1 = \int_0^1 (u_n(t))^2 dH_n(t) ,$$

where u_n is the same as (3.1) and $H_n(t)$ is the empirical distribution function of the variables $Y_1, Y_2, \cdots, Y_n$. The asymptotic null distribution of T_n^1 coincides with that of the statistic proposed by Rothman and Woodroofe [5]. It can be proved in almost the same way as Theorem 2 of Rothman and Woodroofe [5] by using Theorem 2.1.

4. Remarks

First we shall give an interpretation of α in the problem of Section 3. Suppose that $Z_1, Z_2, \cdots, Z_m, \cdots$ is a sequence of independent random variables with common continuous distribution function G_2 defined on $(0, 1)$, which is symmetric about 1/2. We assume that we observe Z_i with probability β for overlooking when $Z_i>1/2$. To be precise, we assume that there exists a sequence of independent and identically distributed random variables $I_1, I_2, \cdots, I_m, \cdots$, where

$$I_i = \begin{cases} 1 & \text{with probability} \quad \beta , \\ 0 & \text{with probability} \quad 1-\beta . \end{cases}$$

We also assume that Z's and I's are independent. Suppose that we can observe Z_i if and only if $(Z_i \leqq 1/2)$ or $(Z_i>1/2$ and $I_i=0)$. Let $X_1, X_2, \cdots, X_n, \cdots$ be a sequence of observations in the above situation. Then $X_1, X_2, \cdots, X_n$ are independent identically distributed random variables with the following distribution function F. If $0\leqq x\leqq 1/2$, then we have

$$\begin{aligned} (4.1) \qquad F(x) = \mathrm{P}\,(X_1\leqq x) = \mathrm{P}\,(Z_1\leqq x) & \\ + \mathrm{P}\left(Z_1>\frac{1}{2}, I_1=1, Z_2\leqq x\right) & \\ + \mathrm{P}\left(Z_1>\frac{1}{2}, I_1=1, Z_2>\frac{1}{2}, I_2=1, Z_3\leqq x\right) & \\ + \cdots & \end{aligned}$$

$$=G_2(x)\sum_{i=0}^{\infty}\left(\frac{\beta}{2}\right)^i=\frac{2}{2-\beta}G_2(x)\ .$$

If $1/2<x\leqq 1$, then

$$\begin{aligned}\mathrm{P}\left(\frac{1}{2}<X_1\leqq x\right)&=\mathrm{P}\left(Z_1>\frac{1}{2},\ I_1=0,\ Z_1\leqq x\right)\\&\quad+\mathrm{P}\left(Z_1>\frac{1}{2},\ I_1=1,\ Z_2>\frac{1}{2},\ I_2=0,\ Z_2\leqq x\right)\\&\quad+\cdots\\&=(1-\beta)\cdot\left(\sum_{i=0}^{\infty}\left(\frac{\beta}{2}\right)^i\right)\cdot\left(G_2(x)-\frac{1}{2}\right)\ .\end{aligned}$$

Hence, if $1/2<x\leqq 1$, then

$$F(x)=\frac{1}{2-\beta}+2\cdot\frac{1-\beta}{2-\beta}\left(G_2(x)-\frac{1}{2}\right)\ . \tag{4.2}$$

From the symmetry of G_2, the right hand side of (4.2) can be written as

$$\frac{1}{2-\beta}+2\cdot\frac{1-\beta}{2-\beta}\left(\frac{1}{2}-G_2(1-x)\right)\ . \tag{4.3}$$

If we put $\alpha=1/(2-\beta)$ and $G^*(t)=2G_2(t/2)$, then F satisfies the hypothesis H_1 from (4.1) and (4.3).

Consequently, we may consider in this situation that testing the hypothesis H_1 means testing whether the distribution function of the hidden variable Z_i is symmetric about 1/2. In the case, α is determined by β, and G^* is determined by the distribution function of the hidden variable Z_i.

Next we remark on an alternative approach to the problem in Section 3. It is easily seen that, if a distribution function F satisfies the hypothesis H_0, then

$$(1-\alpha)F(x)+\alpha F(-x)=\alpha\ ,\qquad \text{for}\quad x\leqq 0\ ,$$

holds. In fact, the statistic T_n defined by (3.2), when the observations are supposed to be transformed to (0, 1) by a continuous and strictly increasing distribution function which is symmetric about zero, can be written as

$$\sqrt{n}\sup_{x\leqq 0}\left|\frac{1}{\sqrt{\alpha(1-\alpha)}}((1-\alpha)F_n(x)+\alpha F_n(-x)-\alpha)\right|\ . \tag{4.4}$$

Therefore, of course, as in the proof of the theorem of Butler [3], we can derive the asymptotic null distribution of (4.4) directly using the fact that $\sqrt{n}\,(F_n(t)-F(t))$ converges weakly to $\beta(F(t))$, where $\beta(t)$ is a

Brownian bridge. It is indeed not so hard to see that the asymptotic null distribution can be represented as

$$\frac{1}{\sqrt{\alpha}} \sup_{0\leq t\leq \alpha} \left| \sqrt{1-\alpha}\,\beta(t)+\frac{\alpha}{\sqrt{1-\alpha}}\beta\left(1-\frac{1-\alpha}{\alpha}t\right)\right|$$

and that the process $\sqrt{1-\alpha}\beta(t)+(\alpha/\sqrt{1-\alpha})\beta(1-((1-\alpha)/\alpha)t)$, $0\leqq t\leqq \alpha$ is a standard Wiener process. We, however, think that our approach in Section 3 is not only much simpler than the direct calculation but also giving an intuitive explanation why the asymptotic distribution of the statistic can be represented as a function of a standard Wiener process.

We lastly give three typical examples as applications of Theorem 2.1. They may be called testing problems for local property of a distribution function rather than those for symmetry.

Example 1. Let $X_1, X_2, \cdots, X_n$ be independent random variables having a common continuous distribution function F on $(0, 1)$. Let $0<\alpha<1$ be a given constant. Suppose that φ is a strictly increasing, continuous mapping of $[1/2, 1]$ onto itself. We consider the statistical hypothesis

H_φ: There exists a continuous distribution function G^* on $(0, 1)$ such that

$$F(t)=\begin{cases} \alpha G^*(2t) & \text{if } 0<t\leqq \frac{1}{2}, \\ \alpha+(1-\alpha)(1-G^*(2-2\varphi(t))) & \text{if } \frac{1}{2}<t<1, \end{cases}$$

holds.

We define, for $i=1,\cdots,n$,

$$Y_i=\begin{cases} 2X_i & \text{if } X_i\leqq \frac{1}{2}, \\ 2(1-\varphi(X_i)) & \text{if } X_i>\frac{1}{2}, \end{cases}$$

and

$$\xi_i=\begin{cases} \sqrt{\frac{1-\alpha}{\alpha}} & \text{if } X_i\leqq \frac{1}{2}, \\ -\sqrt{\frac{\alpha}{1-\alpha}} & \text{if } X_i>\frac{1}{2}. \end{cases}$$

Similarly as in the proof of Theorem 3.1, we can see, if F satisfies the hypothesis H_φ, that the distribution function of Y_i is G^* and that ξ's and Y's are independent. Then Theorem 2.1 is applicable and we obtain the asymptotic null distribution of T_n defined by (3.2).

Example 2. Let X's be the same as in Example 1. For simplicity we further assume that F is absolutely continuous with respect to Lebesgue measure. Let f be the density of F. Let $0<c_1<c_2<1-c_1<1$ be given numbers. We consider the statistical hypothesis

H_3: f satisfies the relations

$$\begin{cases} f(t)=f(1-t) & \text{if } 0\leqq t\leqq c_1\,, \\ f(t)=f(c_1+c_2-t) & \text{if } c_1\leqq t\leqq \dfrac{c_1+c_2}{2}\,, \\ f(t)=f(1-c_1+c_2-t) & \text{if } c_2\leqq t\leqq \dfrac{c_2-c_1+1}{2}\,. \end{cases}$$

We let

$$A_1=[0,\, c_1)$$

$$A_2=\left[c_1,\, \frac{c_1+c_2}{2}\right)$$

$$A_3=\left[\frac{c_1+c_2}{2},\, c_2\right)$$

$$A_4=\left[c_2,\, \frac{1-c_1+c_2}{2}\right)$$

$$A_5=\left[\frac{1-c_1+c_2}{2},\, 1-c_1\right)$$

and

$$A_6=[1-c_1,\, 1]\,.$$

If we define, for $i=1, 2,\cdots, n$,

$$(4.5)\qquad Y_i=\begin{cases} \dfrac{1}{c_1}X_i & \text{if } X_i\in A_1\,, \\ \dfrac{2}{c_2-c_1}(X_i-c_1) & \text{if } X_i\in A_2\,, \\ \dfrac{2}{c_2-c_1}(c_2-X_i) & \text{if } X_i\in A_3\,, \\ \dfrac{2}{1-c_1-c_2}(X_i-c_2) & \text{if } X_i\in A_4\,, \end{cases}$$

$$\begin{cases} \dfrac{2}{1-c_1-c_2}(1-c_1-X_i) & \text{if} \quad X_i \in A_5 , \\ \dfrac{1}{c_1}(1-X_i) & \text{if} \quad X_i \in A_6 , \end{cases}$$

and

$$\xi_i = \begin{cases} 1 & \text{if} \quad X_i \in A_1 \cup A_2 \cup A_4 , \\ -1 & \text{if} \quad X_i \in A_3 \cup A_5 \cup A_6 , \end{cases}$$

then Y's and ξ's satisfy the assumptions of Theorem 2.1 and hence we can obtain the similar result as Theorem 3.1 on the statistic T_n.

Example 3. Let X's, c_1 and c_2 be the same as in Example 2. We also assume that F is absolutely continuous. f denotes the density of F. Let $\alpha_1, \alpha_2, \cdots, \alpha_6$ be positive constants such that $\alpha_1+\alpha_2+\cdots+\alpha_6=1$. We consider the statistical hypothesis

H_4: f satisfies the relations

$$\begin{cases} \alpha_6 f(t)=\alpha_1 f(1-t) & \text{if} \quad 0 \leqq t \leqq c_1 , \\ \alpha_3 f(t)=\alpha_2 f(c_1+c_2-t) & \text{if} \quad c_1 \leqq t \leqq \dfrac{c_1+c_2}{2} , \\ \alpha_5 f(t)=\alpha_4 f(1-c_1+c_2-t) & \text{if} \quad c_2 \leqq t \leqq \dfrac{c_2-c_1+1}{2} , \end{cases}$$

$$\begin{cases} \alpha_1=F(c_1) , & \alpha_2=F\left(\dfrac{c_1+c_2}{2}\right)-F(c_1) , \\ \alpha_3=F(c_2)-F\left(\dfrac{c_1+c_2}{2}\right) , & \alpha_4=F\left(\dfrac{c_2-c_1+1}{2}\right)-F(c_2) , \\ \alpha_5=F(1-c_1)-F\left(\dfrac{c_2-c_1+1}{2}\right) , & \alpha_6=1-F(1-c_1) , \end{cases}$$

and further for every $0<t<1$,

$$G_1^*(t)=G_2^*(t)=G_3^*(t)$$

hold, where

$$G_1^*(t)=\frac{1}{\alpha_1}F(c_1 t) ,$$

$$G_2^*(t)=\frac{1}{\alpha_2}\left(F\left(c_1+\frac{c_2-c_1}{2}t\right)-F(c_1)\right)$$

and

$$G_3^*(t)=\frac{1}{\alpha_4}\left(F\left(c_2+\frac{1-c_1-c_2}{2}t\right)-F(c_2)\right) .$$

We let $A_1, A_2, \cdots, A_6$ be the same as those of Example 2. We define Y_i by (4.5) and

$$\xi_i=\begin{cases} c\cdot\frac{1}{\alpha_1} & \text{if } X_i\in A_1 , \\ c\cdot\frac{1}{\alpha_2} & \text{if } X_i\in A_2 , \\ -c\cdot\frac{1}{\alpha_3} & \text{if } X_i\in A_3 , \\ c\cdot\frac{1}{\alpha_4} & \text{if } X_i\in A_4 , \\ -c\cdot\frac{1}{\alpha_5} & \text{if } X_i\in A_5 , \\ -c\cdot\frac{1}{\alpha_6} & \text{if } X_i\in A_6 , \end{cases}$$

where

$$c=\frac{1}{\sqrt{\frac{1}{\alpha_1}+\cdots+\frac{1}{\alpha_6}}} .$$

Then it is easy to see that Y's and ξ's are independent and that, for every $i=1,\cdots,n$, ξ_i has mean 0 and variance 1. Hence we can obtain the asymptotic null distribution of T_n by using Theorem 2.1.

THE INSTITUTE OF STATISTICAL MATHEMATICS

REFERENCES

[1] Bickel, P. J. and Wichura, M. J. (1971). Convergence criteria for multiparameter stochastic process and some applications, *Ann. Math. Statist.*, **42**, 1656-1670.
[2] Billingsley, P. (1968). *Convergence of Probability Measures*, Wiley, New York.
[3] Butler, C. (1969). A test for symmetry using the sample distribution function, *Ann. Math. Statist.*, **40**, 2209-2210.
[4] Feller, W. (1966). *An Introduction to Probability Theory and Its Applications*, Vol. II, Wiley, New York.
[5] Rothman, E. D. and Woodroofe, M. (1972). A Cramér-von Mises type statistic for testing symmetry, *Ann. Math. Statist.*, **43**, 2035-2038.