Ann. Inst. Statist. Math.
Vol. 53, No. 3, 552–566 (2001)

CHANGE–POINT DETECTION IN ANGULAR DATA

IRINA GRABOVSKY AND LAJOS HORVÁTH

Department of Mathematics, University of Utah, 155 South 1440 East, Salt Lake City, UT 84112–0090, U.S.A.

(Received June 14, 1999; revised November 16, 1999)

Abstract. We suggest a modification of the CUSUM procedure to detect changes in angular data. We obtain limit theorems for the test statistics under the no change null hypothesis. We discuss the estimation of the times of changes and show that the binary segmentation provides the times of all changes. Our method is applied to a data set on the activity of a pulsar.

Key words and phrases: Angular data, change–point, pulsar, Brownian bridge, von Mises distribution.

1. Introduction and results

By listening to the radio signals of pulsars, astrophysicists have been searching for pulsars which emit very high energy gamma rays. While the pulsar's radio emissions are very regular, the arrival times of pulsed gamma rays often exhibit stochastic variation and the pulsar signals are also mixed with the radiation background. Astronomers make measurements of the radiation and plot the results on the unit circle. If the pulsar is inactive the plotted points are approximately uniform on the unit circle. If the pulsar is active, the folded distribution will be different from the uniform. Hence astronomers wish to decide if the observations have different distributions and to estimate the periods of pulsar activities. For further discussion and analysis of change–point detection in gamma ray data we refer to Lombard *et al.* (1990) and Lombard (1991). We also note that Lombard (1986) and Csörgő and Horváth (1996) (cf. Csörgő and Horváth (1997), pp. 190–194) used rank–based procedures to detect possible changes in angular data.

In this paper we use the following model: the observation $X_1, X_2, \ldots, X_n$ are independent with distribution functions $F_{(1)}(t), F_{(2)}(t), \ldots, F_{(n)}(t)$. Under the null-hypothesis the observations are identically distributed, i.e.

$$H_0 : F_{(1)}(t) = F_{(2)}(t) = \cdots = F_{(n)}(t) \qquad \text{for all} \quad 0 \le t \le 2\pi.$$

Under the alternative there are R changes in the distribution, i.e.

$$\begin{aligned} H_A :\ & \text{there are integers} \quad 1 < k(1) < k(2) < \cdots < k(R) < n \quad \text{such that} \\ & F_{(1)}(t) = \cdots = F_{(k(1))}(t), \quad F_{(k(1)+1)}(t) = \cdots = F_{(k(2))}(t), \ldots, \\ & F_{(k(R))}(t) = \cdots = F_{(n)}(t) \quad \text{for all} \quad 0 \le t \le 2\pi \quad \text{and} \\ & F_{(k(1))}(t_1) \ne F_{(k(2))}(t_1), \ldots, F_{(k(R-1))}(t_{R-1}) \ne F_{k(R)}(t_{R-1}) \\ & \text{with some} \quad t_1, t_2, \ldots, t_{R-1}. \end{aligned}$$

The assumption

$$k(1) = [n\theta_1], k(2) = [n\theta_2], \ldots, k(R) = [n\theta_R], \quad \text{with some} \quad 0 < \theta_1 < \cdots < \theta_R < 1, \tag{1.1}$$

means that the lengths of the periods where the observations are identically distributed are proportional to the total observation period.

Let $S_1(n) = \sum_{1\le j\le n} \cos X_j$, $S_2(n) = \sum_{1\le j\le n} \sin X_j$ and define the CUSUM process $R_{1,n}(k) = n^{-1/2}\{S_1(k) - \frac{k}{n}S_1(n)\}$ and $R_{2,n}(k) = n^{-1/2}\{S_2(k) - \frac{k}{n}S_1(n)\}$. The procedure is based on $T_n(k) = (R_{1,n}^2(k) + R_{2,n}^2(k))^{1/2}$, $1 \le k \le n$. First we consider the asymptotic properties of $T_n(k)$ under the null hypothesis. Let $\mu_1 = E\cos X_1$, $\mu_2 = E\sin X_1$, $\sigma_1^2 = \operatorname{var}(\cos X_1)$, $\sigma_2^2 = \operatorname{var}(\sin X_1)$ and $\gamma = \operatorname{cov}(\cos X_1, \sin X_1)$.

THEOREM 1.1. *We assume that H_0 holds. Then*

$$\{R_{1,n}(nt), R_{2,n}(nt)\} \xrightarrow{\mathcal{D}^2[0,1]} \{\Gamma_1(t) - t\Gamma_1(1), \Gamma_2(t) - t\Gamma_2(1)\}, \tag{1.2}$$

where $\{\Gamma(t) = (\Gamma_1(t), \Gamma_2(t)), 0 \le t \le 1\}$ is a Gaussian process with $E\Gamma_1(t) = E\Gamma_2(t) = 0$, $E\Gamma_1(t)\Gamma_1(s) = \sigma_1^2 \min(t,s)$, $E\Gamma_2(t)\Gamma_2(s) = \sigma_2^2 \min(t,s)$ and $E\Gamma_1(t)\Gamma_2(s) = \gamma \min(t,s)$. For any sequence $N(n)$ satisfying

$$\frac{N(n)}{n} \xrightarrow{P} \theta \quad \textit{with some} \quad 0 < \theta < 1 \tag{1.3}$$

we have

$$\{R_{1,N(n)}(tN(n)), R_{2,N(n)}(tN(n))\} \xrightarrow{\mathcal{D}^2[0,1]} \{\Gamma_1(t) - t\Gamma_1(1), \Gamma_2(t) - t\Gamma_2(1)\}. \tag{1.4}$$

Theorem 1.1 implies immediately that

$$T_n(nt) \xrightarrow{\mathcal{D}[0,1]} \Delta(t), \tag{1.5}$$

where

$$\Delta(t) = \{(\Gamma_1(t) - t\Gamma_1(1))^2 + (\Gamma_2(t) - t\Gamma_2(1))^2\}^{1/2} \tag{1.6}$$

and

$$T_{N(n)}(tN(n)) \xrightarrow{\mathcal{D}[0,1]} \Delta(t), \tag{1.7}$$

assuming that (1.3) holds. It is easy to see that $\operatorname{var}(R_{1,n}(nt)) \to \sigma_1^2 t(1-t)$ and $\operatorname{var}(R_{2,n}(nt)) \to \sigma_2^2 t(1-t)$ for any $t \in [0,1]$, as $n \to \infty$, which suggest the maximally selected statistic

$$T_n^* = \sup_{1/(n+1)\le t\le n/(n+1)} T_n(nt)/(t(1-t))^{1/2}.$$

Next we show that T_n^* can be approximated with the maximum of χ^2-processes.

THEOREM 1.2. *We assume that H_0 holds. Then we can define stochastic processes $\{\Delta_n(t), 0 \le t \le 1\}$ such that*

$$\{\Delta_n(t), 0 \le t \le 1\} \stackrel{\mathcal{D}}{=} \{\Delta(t), 0 \le t \le 1\} \quad \textit{for each} \quad n \tag{1.8}$$

and

$$|T_n^* - \Delta_n^*| = o_P(\exp(-\log n)^\epsilon) \tag{1.9}$$

for any $0 < \epsilon < \epsilon_0$ *with some* $\epsilon_0 > 0$, *where*

$$\Delta_n^* = \sup_{1/(n+1) \le t \le n/(n+1)} \Delta_n(t)/(t(1-t))^{1/2}.$$

Unfortunately, nothing is known about the distribution of Δ_n^*. However, it can be easily shown by an exponential transformation, that Δ_n^* in distribution is the maximum of the Euclidean norm of an Ornstein–Uhlenbeck process with dependent components. We refer to Piterbarg ((1996), pp. 115–118) for some results on the tail behavior of Δ_n^*. We can get the asymptotic distribution of T_n^* when $\sigma = \sigma_1 = \sigma_2$ and $\gamma = 0$. Let $a(n) = (2 \log\log n)^{1/2}$ and $b(n) = 2 \log\log n + \log\log\log n$.

THEOREM 1.3. *We assume that* H_0 *holds,* $\gamma = 0$ *and* $\sigma = \sigma_1 = \sigma_2$. *Then*

$$\lim_{n\to\infty} P\left\{a(n)\frac{1}{\sigma}T_n^* \le x + b(n)\right\} = \exp(-2\exp(-x)) \tag{1.10}$$

for all x. *For any sequence* $N(n)$ *satisfying* (1.3) *we have*

$$\lim_{n\to\infty} P\left\{a(N(n))\frac{1}{\sigma}T_{N(n)}^* \le x + b(N(n))\right\} = \exp(-2\exp(-x)) \tag{1.11}$$

for all x.

Next we consider $\sup_{0<t<1} |T_n(nt)|$ and T_n^* under the alternative. Let

$$e_{i,1} = \sum_{1\le j\le i-1} \theta_j(\mu_1(j+1) - \mu_1(j)) + \theta_i(\mu_1(i) - \mu_1^*), \qquad 1 \le i \le R,$$

where $\mu_1^* = \sum_{1\le i\le R+1}(\theta_i - \theta_{i-1})\mu_1(i)$, $\theta_0 = 0$, $\theta_{R+1} = 1$ and $\mu_1(i) = E\cos X_{k(i)}$, $1 \le i \le R+1$. Similarly,

$$e_{i,2} = \sum_{1\le j\le i-1} \theta_j(\mu_2(j+1) - \mu_2(j)) + \theta_i(\mu_2(i) - \mu_2^*), \qquad 1 \le i \le R,$$

where $\mu_2^* = \sum_{1\le i\le R+1}(\theta_i - \theta_{i-1})\mu_2(i)$, with $\mu_2(i) = E\sin X_{k(i)}$, $1 \le i \le R+1$.

THEOREM 1.4. *If* H_A *and* (1.1) *hold and*

$$\max_{1\le i\le R}(|e_{i,1}| + |e_{i,2}|) > 0, \tag{1.12}$$

then there are constants $c^* > 0$ *and* $c^{**} > 0$ *such that*

$$\sup_{0<t<1} T_n(nt) = c^* n^{1/2} + O_P(1) \tag{1.13}$$

and

$$T_n^* = c^{**} n^{1/2} + O_P((\log\log n)^{1/2}). \tag{1.14}$$

We would like to note that Grabovsky (2000) obtained limit results for the distributions of $\sup_t T_n(nt)$ and T_n^* under the alternative.

We can also use $T_n(nt)$ and T_n^* to estimate the time of at least one change. Let $\tilde{k}(n) = \min\{k : T_n(k) = \max_{1\le i\le n} T_n(i)\}$ and $\hat{k}(n) = \min\{k : T_n(k)/(k/n(1-k/n))^{1/2} = T_n^*\}$. We show that both $\tilde{k}(n)$ and $\hat{k}(n)$ can be used to estimate one of the times of changes. This means that $\tilde{k}(n)/n$ will be near θ_{i^*}, where $k(i^*) = n\theta_{i^*}$ is one of times of changes under the alternative. Similarly, $\hat{k}(n)/n$ estimates $\theta_{i^{**}}$.

THEOREM 1.5. *If H_A, (1.1) and (1.12) are satisfied, then there are i^*, $i^{**} \in 1, 2, \ldots, R$ such that*

$$\frac{\tilde{k}(n)}{n} \xrightarrow{P} \theta_{i^*} \tag{1.15}$$

and

$$\frac{\hat{k}(n)}{n} \xrightarrow{P} \theta_{i^{**}}. \tag{1.16}$$

The proof of Theorems 1.1–1.5 will be given in Section 3.. Next we show that Vostrikova's (1981) binary segmentation procedure can be used to divide the data into homogeneous subsets.

Using a test, for example $\sup\{T_n(nt), 0 < t < 1\}$ or T_n^*, we check H_0 using asymptotic critical values. If H_0 is rejected, $\tilde{k}(n)$ (or $\hat{k}(n)$) can be used to estimate one of the times of changes. Then we divide the data into two subsets $X_1, \ldots, X_{\tilde{k}(n)}$ and $X_{\tilde{k}(n)+1}, \ldots, X_n$ and test the "no change" null-hypothesis separately for each of these two subsets. We can use (1.4) or (1.11) to get asymptotic initial values. If the "no change" null-hypothesis is not rejected, the subset $X_1, \ldots, X_{\tilde{k}(n)}$ is homogeneous. If the "no change" is rejected for $X_1, \ldots, X_{\tilde{k}(n)}$, we find another time of change with $\tilde{k}(\tilde{k}(n))$, and we continue the segmentation for $X_1, \ldots, X_{\tilde{k}(\tilde{k}(n))}$ and $X_{\tilde{k}(\tilde{k}(n))+1}, \ldots, X_{\tilde{k}(n)}$. We apply the same procedure to $X_{\tilde{k}(n)}, \ldots, X_n$.

2. Application to the von Mises distribution

Lombard *et al.* (1990) and Lombard (1991) assume that the observations follow von Mises distribution on the unit circle. Namely, if the density of $F_{(i)}$ is $f_{(i)}$, then

$$f_{(i)}(t) = \frac{\exp(\rho_i \cos t)}{2\pi I_0(\rho_i)}, \qquad 0 \le t < 2\pi, \tag{2.1}$$

where $\rho_i \ge 0$ are unknown parameters and I_0 denotes the modified Bessel function of the first kind and of order zero (cf. Mardia (1972), pp. 57). If $\rho_i = 0$, then the formula in (2.1) means that

$$f_{(i)}(t) = \frac{1}{2\pi}, \qquad 0 \le t < 2\pi \tag{2.2}$$

i.e. we have the uniform distribution on the unit circle. In this case H_0 and H_A mean that

(2.3) $H_0^* : \rho_1 = \cdots = \rho_n$

and

(2.4) H_A^* : there are integers $1 < k(1) < k(2) < \cdots < k(R) < n$ such that $\rho_1 = \cdots = \rho_{k(1)} \neq \rho_{k(1)+1} = \cdots = \rho_{k(2)+1\ldots} \neq \rho_{k(R)+1} = \cdots = \rho_n$.

Lombard *et al.* (1990) argue that the phases would be uniformly distributed in the absence of a signal while the distribution of the phases is different from the uniform when signal is present. Thus the parameters of the homogeneous periods change from zero (the pulsar is inactive) to non-zero (the pulsar is active).

Let us assume that H_0 holds. The common parameter is denoted by ρ. Mardia ((1972), p. 62) showed that $\mu_1(\rho) = \frac{I_1(\rho)}{I_0(\rho)}$, $0 < \rho < \infty$ and $\mu_1(0) = 0$, where $\mu_1(\rho) = E\cos X_1$ and $\mu_2(\rho) = E\sin X_1 = 0$, $\gamma(\rho) = \text{cov}(\cos X_1, \sin X_1) = 0$ for all $0 \le \rho < \infty$. The modified Bessel function of the first kind and of the first order is denoted by $I_1(\rho)$. In this case (1.2) reduces to

$$\{R_{1,n}(nt), R_{2,n}(nt)\} \xrightarrow{\mathcal{D}^2[0,1]} \{\sigma_1 B_1(t), \sigma_2 B_2(t)\}, \tag{2.5}$$

where $\{B_1(t), 0 \le t \le 1\}$ and $\{B_2(t), 0 \le t \le 1\}$ are independent Brownian bridges. Note that the limit still depends on the unknown σ_1 and σ_2. However, these parameters can be easily estimated with

$$\hat{\sigma}_1^2(n) = \frac{1}{n-1} \sum_{1 \le i \le n} \left\{ \cos X_i - \frac{1}{n} \sum_{1 \le j \le n} \cos X_j \right\}^2$$

and

$$\hat{\sigma}_2^2(n) = \frac{1}{n-1} \sum_{1 \le i \le n} \left\{ \sin X_i - \frac{1}{n} \sum_{1 \le j \le n} \sin X_j \right\}^2.$$

The result in (2.5) implies immediately the weak convergence of $\hat{R}_{1,n}(nt) = R_{1,n}(nt)/\hat{\sigma}_1(n)$ and $\hat{R}_{2,n}(nt) = \hat{R}_{2,n}(nt)/\hat{\sigma}_2(n)$.

THEOREM 2.1. *We assume that H_0^* holds. Then*

$$\{\hat{R}_{1,n}(nt), \hat{R}_{2,n}(nt)\} \xrightarrow{\mathcal{D}^2[0,1]} \{B_1(t), B_2(t)\}, \tag{2.6}$$

where $\{B_1(t), 0 \le t \le 1\}$ and $\{B_2(t), 0 \le t \le 1\}$ are independent Brownian bridges. For any sequence $N(n)$ satisfying (1.3) *we have*

$$\{\hat{R}_{1,N(n)}(tN(n)), \hat{R}_{2,N(n)}(tN(n))\} \xrightarrow{\mathcal{D}^2[0,1]} \{B_1(t), B_2(t)\}. \tag{2.7}$$

Theorem 2.1 yields immediately that $\hat{T}_n(nt) = (\hat{R}_{1,n}^2(nt) + \hat{R}_n^2(nt))^{1/2} \xrightarrow{\mathcal{D}[0,1]} \hat{\Delta}(t)$ and $\hat{T}_{N(n)}(tN(n)) \xrightarrow{\mathcal{D}[0,1]} \hat{\Delta}(t)$, where $\hat{\Delta}(t) = (B_1^2(t) + B_2^2(t))^{1/2}$. We note that Kiefer (1959*a*, 1959*b*) obtained formulas for the distribution functions of the supremum and integral functionals of $\hat{\Delta}(t)$.

Let

$$\hat{T}_n^* = \sup_{1/(n+1) \le t \le n/(n+1)} \hat{T}_n(nt)/(t(1-t))^{1/2}.$$

Similarly to Theorem 1.3 we have the following limit result:

THEOREM 2.2. *We assume that H_0^* holds. Then*

$$\lim_{n\to\infty} P\{a(n)\hat{T}_n^* \le x + b(n)\} = \exp(-2\exp(-x)). \tag{2.8}$$

For any sequence $N(n)$ satisfying (1.3) *we have*

$$\lim_{n\to\infty} P\{a(N(n))\hat{T}_{N(n)}^* \le x + b(N(n))\} = \exp(-2\exp(-x)) \tag{2.9}$$

for all x.

Next we consider $\hat{T}_n(nt)$ under H_A^*. The first result yields the consistency of tests based on $\sup\{\hat{T}_n(nt), 0 < t < 1\}$ and $\hat{T}_n^*$.

THEOREM 2.3. *If H_A^*,* (1.1) *and* (1.13) *hold, then there are constants $\hat{c}^* > 0$ and $\hat{c}^{**} > 0$ such that*

$$\sup_{0<t<1} \hat{T}_n(nt) = \hat{c}^* n^{1/2} + O_P(1) \tag{2.10}$$

and

$$\hat{T}_n^* = \hat{c}^{**} n^{1/2} + O_P((\log\log n)^{1/2}). \tag{2.11}$$

The estimation of the variances σ_1^2 and σ_2^2 has little effect on the estimation procedure discussed in Section 1. Similarly to $\tilde{k}(n)$ and $\hat{k}(n)$ we define

$$\tilde{k}^*(n) = \min\{k : \hat{T}_n(k) = \max_{1\le i\le n} \hat{T}_n(i)\}$$

and

$$\hat{k}^*(n) = \min\{k : \hat{T}_n(k)/(k/n(1-k/n))^{1/2} = \hat{T}_n^*\}.$$

Our last result is the consistency of $\tilde{k}^*(n)/n$ and $\hat{k}^*(n)/n$.

THEOREM 2.4. *If H_A^*,* (1.1) *and* (1.12) *are satisfied, then there are j^* and $j^{**} \in 1, 2, \dots, R$ such that*

$$\frac{\tilde{k}^*(n)}{n} \xrightarrow{P} \theta_{j^*} \tag{2.12}$$

and

$$\frac{\hat{k}^*(n)}{n} \xrightarrow{P} \theta_{j^{**}}. \tag{2.13}$$

In the von Mises model we can compute the constants $\mu_1(i)$ and $\mu_2(i)$ appearing in (1.12). By Mardia ((1972), p. 57) $\mu_1(i) = I_1(\rho_{k(i)})/I_0(\rho_{k(i)})$, $1 \le i \le R+1$ ($\rho_{k(R+1)} = \rho_n$), where $I_1(0)/I_0(0) = 0$ and $\mu_2(i) = 0$, $1 \le i \le R$. Hence $e_{i,2} = 0$ for all $1 \le i \le R+1$ and condition (1.13) holds if and only if $\max\{|e_{i,1}|, 1 \le i \le R\} > 0$. In the model introduced by Lombard *et al.* (1990), the active phases correspond to $\rho > 0$ (and therefore $\mu_1 \ne 0$). Consequently, the condition $\max\{|e_{i,1}|, 1 \le i \le R\} > 0$ is satisfied.

We applied our results to the data of astrophysical observations provided to us by Lombard. The data consist of 1,555 observations. We assumed the observations follow

Table 1. Binary segmentation scheme.

interval	$\sup_t \hat{T}_n(nt)$	$\hat{T}_n^*$	decision	least sq. fit	
1-1555	$p = 0.0725$	$p = 0.1163$	change		
	$\tilde{k}^* = 1060$	$\hat{k}^* = 1460$	$\hat{k} = 1060$		
1-1060	$p = 0.007$	$p = 0.0823$	change		
	$\tilde{k}^* = 678$	$\hat{k}^* = 678$	$\hat{k} = 678$		
1-678	$p = 0.8318$	$p = 0.2574$	no change	$\rho = 0.04$	error=0.001
678-1060	$p = 0.53$	$p = 0.11$	no change	$\rho = 0.3$	error=0.06
1060-1555	$p = 0.496$	$p = 0.5378$	no change	$\rho = 0$	error= 0.001

Table 2. Asymptotic critical values for $\sup_t \hat{T}_n(nt)$.

p	0.9	0.95	0.99
asymp. Value	1.45	1.58	1.84

Table 3. Asymptotic critical values for $\hat{T}_n^*$.

p	0.9	0.95	0.99
asymp. Value	2.94	3.66	5.29
$n = 50$	2.21	2.64	3.50
$n = 100$	2.23	2.69	3.56
$n = 500$	2.29	2.75	3.67

von Mises distributions. We used $\sup\{\hat{T}_n(nt), 0 < t < 1\}$ and $\hat{T}_n^*$ for testing and the corresponding arg max estimations to get estimates for the times of changes. Table 1 summarizes the outcome of the binary segmentation scheme. The last column of Table 1 shows the estimated value of ρ and the least squares fit to the data. It is clear from the table that the first 678 observations correspond to the period of the pulsar's inactivity. They are followed by the next group of 382 observations indicating the pulsar's activity. The last group of 495 observations shows again inactivity. Our conclusion coincides almost exactly with the conclusion reached by Lombard (1991) who used a different approach.

It has been observed that the rate of convergence of maximally selected statistics to the extreme value limit distribution can be very slow. We refer to Horváth and Gombay (1996) for a discussion on the rate of convergence of maximally selected likelihood ratios. We will see in the proof of Theorem 2.2 that two approximations are used to derive (2.8). First $\hat{T}_n^*$ is replaced with a χ^2–process and then we obtain the limit distribution of the χ^2–process. We show that the distribution of $\hat{T}_n^*$ and $\sup\{\hat{\Delta}(t)/(t(1-t))^{1/2}, h \le t \le 1-\ell\}$ are close if $h = \ell = (\log n)^3/n$. (We note that $\sup\{\hat{\Delta}(t)/(t(1-t))^{1/2}, h \le t \le 1-\ell\}$ has an extreme value limit distribution with norming and centering sequences $a(n)$ and $b(n)$). According to Vostrikova (1981)

$$P\left\{\sup_{h \le t \le 1-\ell} \frac{\hat{\Delta}(t)}{(t(1-t))^{1/2}} \le x\right\}$$

Table 4. Simulated critical values for $\sup_t \hat{T}_n(nt)$ and $\hat{T}_n^*$.

	$\sup_t \hat{T}_n(nt)$			$\hat{T}_n^*$		
p	0.9	0.95	0.99	0.99	0.95	0.99
$p = 0$						
$n = 50$	1.37	1.48	1.78	1.97	2.33	3.02
$n = 100$	1.40	1.52	1.83	2.05	2.46	3.12
$n = 500$	1.45	1.59	1.80	2.16	2.54	3.31
$p = 0.1$						
$n = 50$	1.35	1.46	1.72	1.93	2.27	3.03
$n = 100$	1.39	1.53	1.78	1.94	2.38	3.26
$n = 500$	1.45	1.58	1.79	2.27	2.65	3.59
$p = 0.2$						
$n = 50$	1.35	1.49	1.65	2.00	2.25	2.81
$n = 100$	1.39	1.52	1.76	1.99	2.37	3.36
$n = 500$	1.42	1.56	1.82	2.19	2.55	3.34
$p = 0.3$						
$n = 50$	1.36	1.49	1.67	1.98	2.31	2.87
$n = 100$	1.41	1.57	1.78	2.05	2.54	3.13
$n = 500$	1.42	1.56	1.81	2.18	2.57	3.51
$p = 0.4$						
$n = 50$	1.39	1.52	1.75	2.04	2.33	3.02
$n = 100$	1.40	1.52	1.84	2.07	2.46	3.32
$n = 500$	1.40	1.53	1.70	2.09	2.54	3.50
$p = 0.5$						
$n = 50$	1.34	1.47	1.68	1.94	2.34	2.99
$n = 100$	1.39	1.50	1.77	1.96	2.39	3.25
$n = 500$	1.44	1.56	1.79	2.24	2.77	3.59
$p = 0.6$						
$n = 50$	1.34	1.47	1.70	2.00	2.39	3.24
$n = 100$	1.40	1.53	1.89	2.13	2.54	3.56
$n = 500$	1.40	1.54	1.76	2.17	2.65	3.34
$p = 0.7$						
$n = 50$	1.40	1.50	1.73	2.08	2.46	3.15
$n = 100$	1.40	1.50	1.78	2.03	2.48	3.22
$n = 500$	1.42	1.58	1.84	2.12	2.56	3.37
$p = 0.8$						
$n = 50$	1.38	1.50	1.84	2.03	2.58	3.52
$n = 100$	1.42	1.56	1.81	2.29	2.70	3.46
$n = 500$	1.45	1.60	1.83	2.32	2.71	3.58
$p = 0.9$						
$n = 50$	1.36	1.48	1.75	2.13	2.55	3.39
$n = 100$	1.38	1.51	1.75	2.12	2.59	3.57
$n = 500$	1.42	1.52	1.86	2.22	2.66	3.71

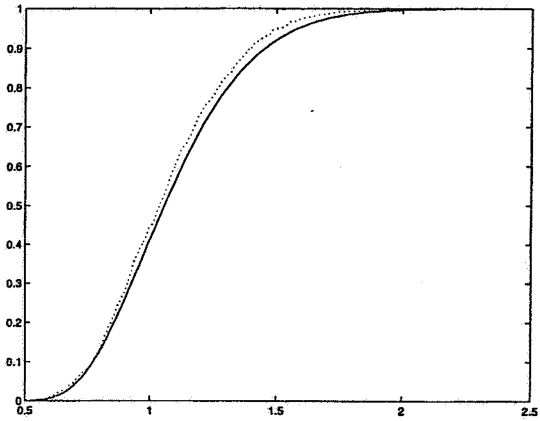

Fig. 1. The limit and the simulated distribution functions of $\sup_t \hat{T}_n(nt)$.

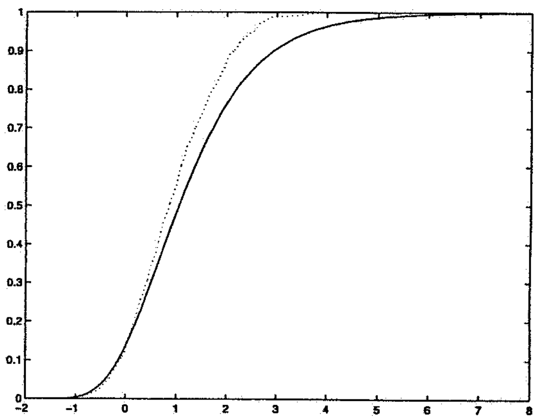

Fig. 2. The limit and the simulated distribution functions of $\hat{T}_n^*$.

$$= \frac{x^2}{2}e^{-x^2/2}\left\{\log\frac{(1-h)(1-\ell)}{h\ell} - \frac{2}{x^2}\log\frac{(1-h)(1-\ell)}{h\ell} + \frac{4}{x^2} + O(\frac{1}{x^4})\right\} \quad \text{as} \quad x \to \infty$$

for any $0 < h, \ell < 1$.

Tables 2 and 3 contain the critical values for $\sup_t \hat{T}_n(nt)$ and $\hat{T}_n^*$ using the limit distributions and Vostrikova's (1981) formula with $h = \ell = (\log n)^3/n$. We performed Monte Carlo simulations to study the accuracy of the approximations. We used (2.1) with $\rho = 0, .1, .2, \ldots, .9$ as the density of the observations under the null hypothesis. The sample sizes were $n = 50, 100$ and 500 and each samples were repeated 5,000 times. The results are given in Table 4. Figures 1 and 2 give the graphs of the limit and the simulated distributions in case of $n = 500$ and $\rho = .6$. It is clear from the simulations that the limit results for $\sup_t \hat{T}_n(nt)$ are acceptable even in case of small sample sizes. The limit result overestimates the critical values for $\hat{T}^*$. Using Vostrikova's tail approximation for the supremum of the χ^2–process we get more suitable critical values.

3. Proofs of Theorems 1.1 –1.5

PROOF OF THEOREM 1.1. By Einmahl (1989) we can find a Gaussian process $\mathbf{\Gamma}(t) = (\Gamma_1(t), \Gamma_2(t))$ with covariance structure specified in Theorem 1.1 such that

$$\|(S_1(T)_1, S_2(T)) - \mathbf{\Gamma}(T)\| \overset{a.s.}{=} O(\log T), \qquad \text{as} \quad T \to \infty, \tag{3.1}$$

where $\|\cdot\|$ denotes the maximum norm of vectors. Now (1.2) follows immediately from (3.1).

To prove (1.4) first we note

$$\begin{aligned}\sup_{0\le t\le 1} |S_1(tN(n)) - \Gamma_1(n\theta t)| &\le \sup_{0\le t\le N(n)} |S_1(t) - \Gamma_1(t)| + \sup_{0\le t\le 1} |\Gamma_1(n\theta t) - \Gamma_1(tN(n))| \\ &= A_1(n) + A_2(n).\end{aligned}$$

Using (3.1) and condition (1.3) we get $A_1(n) \overset{a.s.}{=} O(\log n)$. By (1.3) it is enough to consider $\sup_{0\le t\le n\theta} \sup_{0\le s\le \epsilon n} |\Gamma_1(t+s) - \Gamma_1(t)|$ where $\epsilon > 0$ can be as small as we wish. Since $\Gamma_1(t)/\sigma_1$ is a Wiener process (Brownian motion) we obtain for any n that

$$\sup_{0\le t\le n\theta} \sup_{0\le s\le \epsilon n} |\Gamma_1(t+s) - \Gamma_1(t)| \overset{\mathcal{D}}{=} n^{1/2} \sup_{0\le t\le \theta} \sup_{0\le s\le \epsilon} |\Gamma_1(t+s) - \Gamma_1(t)|.$$

By the continuity of $\Gamma_1(t)$ we have

$$\lim_{\epsilon\downarrow 0} \sup_{0\le t\le \theta} \sup_{0\le s\le \epsilon} |\Gamma_1(t+s) - \Gamma_1(t)| = 0 \qquad a.s.,$$

and therefore $A_2(n) = o_P(n^{1/2})$. Thus we have

$$\sup_{0\le t\le 1} |S_1(tN(n)) - \Gamma_1(n\theta t)| = o_P(n^{1/2}) \tag{3.2}$$

and similarly

$$\sup_{0\le t\le 1} |S_2(tN(n)) - \Gamma_2(n\theta t)| = o_P(n^{1/2}). \tag{3.3}$$

Computing the covariance functions one can easily verify that

$$\{(n\theta)^{-1/2}(\Gamma_1(n\theta t), \Gamma_2(n\theta t)), 0 \le t \le 1\} \overset{\mathcal{D}}{=} \{(\Gamma_1(t), \Gamma_2(t)), 0 \le t \le 1\}. \tag{3.4}$$

By (1.3), (3.2)–(3.4) we have

$$\sup_{0\le t\le 1} \|N^{-1/2}(n)(S_1(tN(n)), S_2(tN(n))) - (n\theta)^{-1/2}\mathbf{\Gamma}(n\theta t)\| = o_P(1),$$

which completes the proof of (1.4).

PROOF OF THEOREM 1.2. Since $\{S_1(nt), S_2(nt), 0 \le t \le 1/2\}$ and $\{S_1(n) - S_1(nt), S_2(n) - S_2(nt), 1/2 < t \le 1\}$ are independent, by Einmahl (1989) for each n we can define two independent Gaussian processes $\mathbf{\Gamma}_n^{(1)}$ and $\mathbf{\Gamma}_n^{(2)}$ such that

$$\sup_{2/n\le t\le 1/2} \|(S_1(nt), S_2(nt)) - \mathbf{\Gamma}_n^{(1)}(nt)\| / \log(nt) = O_P(1) \tag{3.5}$$

and

$$(3.6)\qquad \sup_{1/2\le t\le 1-2/n} \|(S_1(n)-S_1(nt),S_2(n)-S_2(nt))-\mathbf{\Gamma}_n^{(2)}(n-nt)\|/\log(n-nt) = O_P(1).$$

Next we define

$$\hat{\mathbf{B}}_n(t)=\begin{cases} n^{-1/2}(\mathbf{\Gamma}_n^{(1)}(nt)-t(\mathbf{\Gamma}_n^{(1)}(n/2)+\mathbf{\Gamma}_n^{(2)}(n/2))), & 0\le t\le 1/2\\ n^{-1/2}(-\mathbf{\Gamma}_n^{(2)}(n-nt)+(1-t)(\mathbf{\Gamma}_n^{(1)}(n/2)+\mathbf{\Gamma}_n^{(2)}(n/2))), & 1/2\le t\le 1.\end{cases}$$

Using standard arguments (cf. for example, Csörgő and Horváth ((1997), 17–19)) one can derive from (3.5) and (3.6) that

$$(3.7)\qquad \sup_{0\le t\le 1}\|(R_{1,n}(nt),R_{2,n}(nt))-\hat{\mathbf{B}}_n(t)\| = O_P(n^{-1/2}\log n)$$

and

$$(3.8)\qquad n^{\nu}\sup_{\nu/n\le t\le 1-\nu/n}\|(R_{1,n}(nt),R_{2,n}(nt))-\hat{\mathbf{B}}_n(t)\|/(t(1-t))^{1/2-\nu} = O_P(1)$$

for any $\nu>0$ and $0\le\nu<1/2$. Next we note that

$$(3.9)\qquad \{\hat{\mathbf{B}}_n(t),0\le t\le 1\}\stackrel{\mathcal{D}}{=}\{\Gamma_1(t)-t\Gamma_1(1),\Gamma_2(t)-t\Gamma_2(1),0\le t\le 1\}.$$

Let $\Delta_n(t)=(\hat{B}_{n,1}^2(t)+\hat{B}_{n,2}^2(t))^{1/2}/(t(1-t))^{1/2}$. It follows from (3.9) that (1.8) holds. Next we define $c(n)=\exp((\log n)^{\alpha})/n$ with some $0<\alpha<1$. We can assume that $\sigma_1\le\sigma_2$. We observe that for any $0<c_1<c_2<1$

$$(3.10)\qquad \begin{aligned}\sup_{c_1\le t\le c_2}|\hat{B}_{n,2}(t)|/(t(1-t))^{1/2} &\le \sup_{c_1\le t\le c_2}|\Delta_n(t)|\\ &\le \sup_{c_1\le t\le c_2}|\hat{B}_{n,1}(t)|/(t(1-t))^{1/2}\\ &\quad+\sup_{c_1\le t\le c_2}|\hat{B}_{n,1}(t)|/(t(1-t))^{1/2}.\end{aligned}$$

By the Darling and Erdős (1956) law (cf. also Csörgő and Horváth ((1997), pp. 363–372)) we have

$$(3.11)\qquad (2\log\log n)^{-1/2}\sup_{1/n\le t\le 1-1/n}|\hat{B}_{n,2}(t)|/(t(1-t))^{1/2}\xrightarrow{P}\sigma_2,$$

$$(3.12)\qquad (2\log\log n)^{-1/2}\sup_{1/n\le t\le c(n)}|\hat{B}_{n,i}(t)|/(t(1-t))^{1/2}\xrightarrow{P}\alpha^{1/2}\sigma_i,\qquad i=1,2$$

and

$$(3.13)\qquad (2\log\log n)^{-1/2}\sup_{1-c(n)\le t\le 1-1/n}|\hat{B}_{n,i}(t)|/(t(1-t))^{1/2}\xrightarrow{P}\alpha^{1/2}\sigma_i,\qquad i=1,2.$$

Choosing $0<\alpha<(\sigma_2/(\sigma_1+\sigma_2))^2$, by (3.10) – (3.13) we conclude

$$(3.14)\qquad \lim_{n\to\infty}P\left\{\sup_{1/n\le t\le 1/n}\Delta_n(t)=\sup_{c(n)\le t\le 1-c(n)}\Delta_n(t)\right\}=1.$$

Now (3.8) with $\nu = 0$ and (3.14) imply that

$$\lim_{n\to\infty} P\left\{T_n^* = \sup_{c(n)\le t\le 1-c(n)} T_n(nt)/(t(1-t))^{1/2}\right\} = 1. \tag{3.15}$$

The equations (3.14) and (3.15) mean that it is enough to approximate $T_n(nt)/(t(1-t))^{1/2}$ on $[c(n), 1-c(n)]$. By (3.7) we have that

$$\begin{aligned}
&\sup_{c(n)\le t\le 1-c(n)} |(R_{1,n}^2(t) + R_{2,n}^2(t))^{1/2} - (\hat{B}_{n,1}^2(t) + \hat{B}_{n,2}^2(t))^{1/2}|/(t(1-t))^{1/2} \\
&\quad \le \sup_{0\le t\le 1} |(R_{1,n}^2(t) + R_{2,n}^2(t))^{1/2} - (\hat{B}_{n,1}^2(t) + \hat{B}_{n,2}^2(t))^{1/2} \sup_{c(n)\le s\le 1-c(n)} (s(1-s))^{-1/2}| \\
&\quad = O_P(\log n \exp(-(\log n)^\alpha/2)),
\end{aligned}$$

which completes the proof of (1.9).

PROOF OF THEOREM 1.3. If $\gamma = 0$ and $\sigma_1 = \sigma_2 = \sigma$, then

$$\left\{\frac{1}{\sigma}\Delta_n(t), 0 < t < 1\right\} \stackrel{\mathcal{D}}{=} \{(B_1^2(t) + B_2^2(t))^{1/2}/(t(1-t))^{1/2}, 0 < t < 1\},$$

where $\{B_1(t), 0 \le t \le 1\}$ and $\{B_2(t), 0 \le t \le 1\}$ are independent Brownian bridges. Hence (1.9) and Horváth (1993) (cf. also Csörgő and Horváth ((1997), pp. 363–372)) imply (1.10).

Let $\xi(n) = a(n)T_n^*/\sigma - b(n)$. By the Skorohod–Dudley–Wichura representation theorem (cf., for example, Shorack and Wellner ((1986), p. 47)) we can write (1.10) as

$$\|(\xi^*(n), N^*(n)/n) - (\xi^*, \theta)\| \stackrel{a.s.}{\to} 0, \tag{3.16}$$

where ξ^* is a random variable with distribution function $\exp(-2\exp(-x))$ and

$$\{\xi^*(n), N^*(n)\} \stackrel{\mathcal{D}}{=} \{\xi(n), N(n)\} \quad \text{for each} \quad n. \tag{3.17}$$

Let $\delta, \epsilon > 0$. By (3.16) there is K such that $P\{\sup_{K\le n<\infty} |\xi^*(n) - \xi^*| > \epsilon\} \le \delta$. Since $N^*(n) \to \infty$ a.s. , there is n_0 such that $P\{N^*(n) \le K\} \le \delta$, if $n \ge n_0$, and therefore $P\{|\xi^*(N^*(n)) - \xi^*| > \epsilon\} \le 2\delta$, if $n \ge n_0$. Now (1.11) follows from (3.17).

PROOF OF THEOREM 1.4. Let

$$l_i(k) = \begin{cases} k(\mu_i(1) - \mu_i^*), & \text{if} \quad 1 \le k \le k(1) \\ (\mu_i(1) - \mu_i^*)k(1) + k(\mu_i(2) - \mu_i^*), & \text{if} \quad k(1) \le k \le k(2) \\ \vdots & \\ (\mu_i(1) - \mu_i^*)k(1) + \cdots + (\mu_i(R) - \mu_R^*)k(R) + k(\mu_i(R_1) - \mu_i^*), & \\ \qquad\qquad \text{if} \quad k(R) < k \le k(R+1) & \end{cases}$$

for each $i = 1, 2$. By the weak convergence of partial sums we have

$$n^{-1/2} \max_{1\le k\le n} |S_i(k) - \frac{k}{n} S_i(n) - \ell_i(k)| = O_P(1), \qquad i = 1, 2,$$

and therefore

$$\max_{1\le k\le n} |T_n(k) - n^{-1/2}(\ell_1^2(k)+\ell_2^2(k))^{1/2}| = O_P(1). \tag{3.18}$$

Observing that

$$\max_{1\le k\le n} \frac{1}{n}(\ell_1^2(k)+\ell_2^2(k))^{1/2} \to \max_{1\le i\le R}(e_{i,1}^2+e_{i,2}^2)^{1/2} > 0,$$

the proof of (1.13) is complete.

Using the Darling and Erdős (1956) law on $[1, k(1)]$, $[k(R), n)$ and the weak convergence of partial sums on $[k(1), k(R)]$, similarly to (3.18) we have

$$\begin{aligned}\max_{1\le k<n} |T_n(k) - n^{-1/2}(\ell_1^2(k)+\ell_2^2(k))^{1/2}| \Big/ \left(\frac{k}{n}\left(1-\frac{k}{n}\right)\right)^{1/2} \\ = O_P((\log\log n)^{1/2}),\end{aligned} \tag{3.19}$$

which gives immediately (1.14).

Proof of Theorem 1.5. Let i^* the smallest integer satisfying

$$\max_{1\le i\le k}(e_{i,1}^2+e_{i,2}^2)^{1/2} = (e_{i^*,1}^2+e_{i^*,2}^2)^{1/2}.$$

Now (3.18) implies (1.15). Similarly, (1.16) follows from (3.19).

4. Proofs of Theorems 2.1– 2.4

By the central limit theorem we have

$$|\hat\sigma_1^2(n) - \sigma_1^2| = O_P(n^{-1/2}) \text{ and } |\hat\sigma_2^2(n) - \sigma_2^2| = O_P(n^{-1/2}). \tag{4.1}$$

Proof of Theorem 2.1. Using (2.5) we get that

$$\max_{1\le k\le n}\left|\frac{R_{i,n}(k)}{\hat\sigma_i(n)} - \frac{R_{i,n}(k)}{\sigma_i}\right| = O_P(n^{-1/2}), \qquad i=1,2,$$

and therefore (2.6) follows from (2.5).

Arguing as in the proof of Theorem 1.1 one can easily show that (2.6) implies (2.7).

Proof of Theorem 2.2. By the Darling and Erdős (1956) law and the central limit theorem we have

$$\max_{1\le k\le \log n} |R_{i,n}(k)/(k/n)^{1/2}| = O_P((\log\log\log n)^{1/2}), \qquad i=1,2. \tag{4.2}$$

and

$$\max_{n-\log n\le k<n} |R_{i,n}(k)/((n-k)/n)^{1/2}| = O_P((\log\log\log n)^{1/2}), \qquad i=1,2. \tag{4.3}$$

Thus (1.10) implies that

$$
\begin{aligned}
(4.4)\qquad & \lim_{n\to\infty} P\{a(n) \max_{\log n \le k \le n-\log n} (R_{1,n}^2(k)/\sigma_1^2 \\
& \qquad + R_{2,n}^2(k)/\sigma_2^2)^{1/2} / \left((k/n)(1-(k/n))\right)^{1/2} \le x + b(n)\} \\
& = \exp(-2\exp(-x))
\end{aligned}
$$

for all x. Putting together (4.1)–(4.3) we get

$$
\max_{1\le k\le \log n} |\hat{R}_{i,n}(k)|/(k/n)^{1/2} = O_P((\log\log\log n)^{1/2}), \qquad i = 1, 2
$$

and

$$
\max_{n-\log n\le k<n} |\hat{R}_{i,n}(k)| \Big/ \left(1-\frac{k}{n}\right)^{1/2} = O_P((\log\log\log n)^{1/2}), \qquad i = 1, 2.
$$

On the other hand, (4.1) and (4.4) yield

$$
\begin{aligned}
& \max_{\log n\le k\le n-\log n} \left| \left(\frac{R_{1,n}^2(k)}{\sigma_1^2} + \frac{R_{2,n}^2(k)}{\sigma_2^2} \right)^{1/2} - \left(\frac{R_{1,n}^2(k)}{\hat{\sigma}_1^2} + \frac{R_{2,n}^2(k)}{\hat{\sigma}_2^2} \right)^{1/2} \right| \Big/ \left(\frac{k}{n}\left(1-\frac{k}{n}\right)\right)^{1/2} \\
& = O_P\left(\left(\frac{\log\log n}{\log n} \right)^{1/2} \right),
\end{aligned}
$$

and therefore (2.8) follows from (4.4).

The result in (2.9) can be derived from (2.8) in the same way as (1.11) was derived from (1.10) using the Skorohod–Dudley–Wichura representation.

PROOF OF THEOREM 2.3. Under H_A^* we can find two positive constants such that

$$
|\hat{\sigma}_1^2(n) - \gamma_1^2| = O_P(n^{-1/2}) \qquad \text{and} \qquad |\hat{\sigma}_2^2(n) - \gamma_2^2| = O_P(n^{-1/2}).
$$

So instead of (3.18) we write

$$
\max_{1\le k\le n} |\hat{T}_n(k) - n^{-1/2}((\ell_1(k)/\gamma_1)^2 + (\ell_2(k)/\gamma_2)^2)^{1/2}| = O_P(1)
$$

and

$$
\begin{aligned}
& \frac{1}{n} \max_{1\le k\le n} ((\ell_1(k)/\gamma_1)^2 + (\ell_2(k)/\gamma_2)^2)^{1/2} \\
& \qquad \to \max_{1\le i\le R} ((e_{i,1}/\gamma_1)^2 + (e_{i,2}/\gamma_2)^2)^{1/2} > 0,
\end{aligned}
$$

completing the proof (2.10). Similar arguments give (2.11).

PROOF OF THEOREM 2.4. Following the proof of Theorem 1.5 one can derive (2.12) and (2.13) from (2.10) and (2.11).

Acknowledgements

We are grateful to Professor Fred Lombard for providing the data on the pulsar.

References

Csörgő, M. and Horváth, L. (1996). A note on the change-point problem for angular data, *Statist. Probab. Lett.*, **27**, 61–65.

Csörgő, M. and Horváth, L. (1997). *Limit Theorems in Change–Point Analysis*, Wiley, Chichester.

Darling, D. A. and Erdős, P. (1956) A limit theorem for the maximum of normalized sums of independent random variables, *Duke Math. J.*, **23**, 143–155.

Einmahl, U. (1989). Extensions of results of Komlós, Major and Tusnády to the multivariate case, *J. Multivariate Anal.*, **28**, 20–68.

Grabovsky, I. (2000). Asymptotic Analysis in Change–Point Problems, PhD Thesis, Department of Mathematics, University of Utah, Salt Lake City, Utah.

Horváth, L. (1993). The maximum likelihood method for testing changes in the parameters of normal observations, *Ann. Statist.*, **21**, 671–680.

Horváth, L. and Gombay, E. (1996). On the rate of approximations for the maximum likelihood test in change–point models, *J. Multivariate Anal.*, **59**, 120–152.

Kiefer, J. (1959*a*). A functional equation technique for obtaining Wiener process probabilities associated with theorems of Kolmogorov–Smirnov type, *Math. Proc. Cambridge Philos. Soc.*, **55**, 328–332.

Kiefer, J. (1959*b*). K-sample analogues of the Kolmogorov-Smirnov and Cramér–von Mises tests, *Ann. Math. Statist.*, **30**, 420–447.

Lombard, F. (1986). The change–point problem for angular data: A nonparametric approach, *Technometrics*, **28**, 391–397.

Lombard, F. (1991). A change–point analysis of some data arising in gamma-ray astronomy, *South African Statist. J.*, **25**, 83–98.

Lombard, F., de Jager, O. C. and Schultz, D. M. (1990). The detection of a change point in periodic gamma ray data, *Nuclear Phys. B Proc. Suppl.*, **14A**, 285–290.

Mardia, K. V. (1972). *Statistics of Directional Data*, Academic Press, New York.

Piterbarg, V. I. (1996). Asymptotic methods in the theory of Gaussian processes and fields, *Trans. Math. Monogr.*, **148**, Amer. Math. Soc. Providence, Rhode Island.

Shorack, G. R. and Wellner, J. A. (1986). *Empirical Processes with Application to Statistics*, Wiley, New York.

Vostrikova, L. Ju. (1981). Detecting “disorder” in multidimensional random processes, *Soviet. Math. Dokl.*, **24**, 55–59.