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.

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

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

$$H_A : \text{there are integers} \quad 1 < k(1) < k(2) < \cdots < k(R) < n \quad \text{such that}$$

$$F_{(1)}(t) = \cdots = F_{(k(1))}(t), \quad F_{(k(1)+1)}(t) = \cdots = F_{(k(2))}(t), \ldots,$$

$$F_{(k(R))}(t) = \cdots = F_{(n)}(t) \quad \text{for all} \quad 0 \leq t \leq 2\pi \quad \text{and}$$

$$F_{(k(1))}(t_1) \neq F_{(k(2))}(t_1), \ldots, F_{(k(R-1))}(t_{R-1}) \neq F_{(k(R))}(t_{R-1})$$

with some $t_1, t_2, \ldots, t_{R-1}$. 

552
The assumption

(1.1) \[ k(1) = [n\theta_1], k(2) = [n\theta_2], \ldots, k(R) = [n\theta_R], \]
with some \( 0 < \theta_1 < \cdots < \theta_R < 1, \)

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

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

**Theorem 1.1.** We assume that \( H_0 \) holds. Then

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

where \( \{ \Gamma(t) = (\Gamma_1(t), \Gamma_2(t)), 0 \leq t \leq 1 \} \) 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

(1.3) \[ \frac{N(n)}{n} \xrightarrow{P} \theta \quad \text{with some} \quad 0 < \theta < 1 \]

we have

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

Theorem 1.1 implies immediately that

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

where

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

and

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

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

\[ T^*_n = \sup_{1/(n+1) \leq t \leq n/(n+1)} T_n(nt)/(t(1-t))^{1/2}. \]

Next we show that \( T^*_n \) can be approximated with the maximum of \( \chi^2 \)-processes.

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

(1.8) \[ \{ \Delta_n(t), 0 \leq t \leq 1 \} \xrightarrow{D} \{ \Delta(t), 0 \leq t \leq 1 \} \quad \text{for each} \quad n \]

and