Journal of Nepal Geological Society, 2002, Vol. 27 (Special Issue), pp. 65-68

Prediction of hazards in coal mines through directional change-point analyses

Ashis Sen Gupta
Applied Statistics division, Indian Statistical Institute, Kolkata, W.B. 700 035, India
(E-mail: ashis@isical.ac.in)

ABSTRACT

Change in the mean trend or preferred direction of cleat measurements have been thought of as good predictors of potentially hazardous mining conditions lying ahead. The detection of such a change is thus of great importance and usefulness. This poses the change-point problem for the mean direction of circular data. Usual (arithmetic) mean for linear data is nonsensical for the circular case. Appropriate statistics and techniques for such circular data need to be developed. First some new graphical tools are presented as quick diagnostic or exploratory data analytic techniques. These are followed up by formal parametric statistical tests. Data from the Wallsend Borehole Colliery is analyzed using these new tools and methods, which seem to be quite promising. DDSTAP- Statistical Analysis Package for Directional Data, recently developed by this author has been utilised for such analyses.

INTRODUCTION

It has been noted in many coal mines that changes in the mean trend or preferred direction of cleat trend measurements are good predictors of potentially hazardous mining conditions induced by faults or dykes lying ahead. An example of this consists of 63 measurements of median trends of samples of five cleat trend measurements taken at 20 m intervals along a tunnel in the Wallsend Borehole Colliery, NSW, Australia, to be henceforth referred to as WB Colliery data. Thus, it is an important practical problem to detect such a change. In its generality, this problem is termed as the change-point problem in statistical inference. However, note that in our situation the observations are not in the linear scale but are rather directional in nature, e.g., the mean or median directions of face-cleat from the colliery.

Change-point problem arises in both the situations with look-back or retrospective and on-line or sequential data. Some preliminary tools for predicting possible hazards in the coal mines have been presented by Shepherd and Fisher (1981, 1982). However, no work on the parametric inference for this problem seems to be available. Here some diagnostic tools as well as formal statistical inference procedures are introduced for analyzing such data. Preliminary discussins on some of the proposed methods may be found in Jammalamadaka and Sen Gupta (2001). DDSTAP- Statistical analysis Package for Directional Data, developed by this author (Sen Gupta 1998) has been utilised for such analysis.

STATISTICAL DIAGNOSTICS

For the look-back as well as the on-line data sets, we propose to use the *Changeogram.* A Changeogram displays pictorially in terms of directed arrow, each of unit length, the direction in terms of the angle as given by the corresponding observation. The marked "eye-ball" deviation between two such successive arrows is indicative of the presence fo a possible change-point. The *Circular Difference (CD)* table is constructed by considering the change of directions between two successive observations as measured by their circular difference. For example if θ_t and θ_{t+1} are two successive observations then their circular differences is $\min(|\theta_{t+1}-\theta_t|, 2\pi-|\theta_{t+1}-\theta_t|)$. The three largest such circular differences may be highlighted in the table to facilitate quick detection of possible change-points. A change-point may be indicated by a "large" CD preceded and succeeded by "small" CDs. The Changeogram and CD table may be used for both the look-back and on-line situations. The Changeogram and the CD table proposed above may be obtained from DDSTAP.

FORMAL STATISTICAL INFERENCE

The underlying parametric probability model for the angular data is taken to be the von Mises or circular normal CN(μ,κ) distribution. Its probability density function is given by

$$f(\theta;\mu,\kappa)=\frac{1}{2\pi I_0(\kappa)}\exp(\kappa\cos(\theta-\mu)), 0\le\theta<2\pi, 0\le\mu<2\pi, \kappa>0.$$

THE LOOK-BACK PROBLEM

Let $\theta_1,\theta_2,........\theta_n$ be independent random variables. The problem is to test $H_0:\theta_1,......,\theta_n$ are distributed as $CN(\kappa,\mu_0)$, against $H_1:\theta_1,......,\theta_r$ are distributed as $CN(\kappa,\mu_0)$ and $\theta_{r+1},.....,\theta_n$ are distributed as $CN(\kappa,\mu_1)$ for some $r, 1\le r\le n-1$. Thus μ_0, μ_1 and κ are the initial mean direction, the possibly changed mean direction and the concentration parameter, respectively, of the underlying distributions. For this testing problem we will appeal to the well known approach of *Likelihood Ratio Test (LRT)*. Exact cut-off points are obtained by simulations.

Let,

$$\upsilon = E[\sin(\theta_i - \frac{\mu_1}{2})] = -\sin\frac{\mu_1}{2}A(\kappa)$$

and

$$\tau^2 = V[\sin(\theta_i - \frac{\mu_1}{2})] = \cos^2\frac{\mu_1}{2} - \frac{I_0''(\kappa)}{I_0(\kappa)}\cos\mu_1 - \sin^2\frac{\mu_1}{2}A^2(\kappa)$$

where,

$A(\kappa) = \frac{I_0'(\kappa)}{I_0(\kappa)}$. Let's denote the time reversed Brownian Motion on $[0,1]$ with drift 0 and diffusion coefficient 1 by $B_0^*(t)$.

Let, $S_i^k(\eta,\alpha) = \sum_{i=1}^{k}\sin(\eta_i - \alpha)$ and

$$C_i^k(\eta,\alpha) = \sum_{i=1}^{k}\cos(\eta_i - \alpha)$$

Theorem

(a) Let all the parameters μ_0, μ_1 and κ be known.

In testing H_0 against H_1, the LRT statistics is equivalent to:

$$\widetilde{\wedge} = \max_r S_{r+1}^{n}\left[\theta, \frac{\mu_1}{2}\right] \quad \text{(Sen Gupta and Laha 2001).}$$

The asymptotic null distribution of

$$\max_r \frac{S_{r+1}^{n}(\theta, \frac{\mu_1}{2}) - (\eta - r)v}{n^{1/2}\tau}$$

is the same as that of $\sup_{0\le t\le 1} B_0^*(t)$

(b) Let both μ_0 and μ_1 be unknown. Also, let, ${}_0\hat{\mu}_0$, ${}_1\hat{\mu}_0$ denote the estimate of m_0 under H_0 and H_1, respectively, and ${}_1\hat{\mu}_1$ denote the estimate of μ_1 under H_1.

In testing H_0 against H_1 the LRT statistics is

$$\Lambda = \min_r \Lambda_r$$

Where,

$$\wedge_r = \exp\{\kappa[C_1^n(\theta,{}_0\hat{\mu}_0) - C_1^r(\theta,{}_1\hat{\mu}_0) - C_{r+1}^n(\theta,{}_1\hat{\mu}_1)]\}$$

GENERALIZATIONS

Sometimes the data may be suspected of having more than one change-points, say *m* change-points, τ_i, $i=1,\ldots,m$. The above procedure for a single change-point can be easily generalised. One simple approach would be to partition the data set into blocks of observations $\theta_j, j = \tau+1,\ldots,\tau_{i+1}, i = 0,\ldots,m+1, \tau_0 \equiv 1, \tau_{m+1} \equiv n$. One would then apply the above single change-point test taking two successive blocks and may even merge the preceding blocks if non significance, i.e., conclusion of no change-point, has been arrived at for these blocks on invoking the LRT.

Another important practical aspect is that of the case of unknown k. If the data set is quite large, the *Maximum Likelihood Estimate (MLE)* of κ under H_0 may be assumed to be its true (unknown) value and can replace it (approximately) in the above theorem.

THE ON-LINE PROBLEM

When observations appear sequentially, either on the temporal or on the spatial scale, one is more interested to draw inference on the possible onset of change with the introduction of each data point. While such an onset should be detected as early as possible the consequence of false alarm should also be borne in mind. It is of interest then to control both the False Alarm Rate (FAR) as well as the probability of early detection.

Consider now independent random variables θ_i, $i=1,2,\ldots$ being observed sequentially, where the first k observations come from $CN(\mu_1,\kappa)$ and the remaining observations come from $CN(\mu_2,\kappa)$. We refer to κ as the change-point which is unknown. The sequential change-point testing problem can then be formally stated as the one of obtaining a sequential test for the null hypothesis $H_{0r}: \kappa \ge r$ against $H_{1r}: \kappa \le r-1$, at the r-th stage of sampling. One would stop sampling as soon as the null hypothesis is rejected in favor of the occurrence of the change in the mean direction. For the two statistics to be defined below, this occurs when the statistics first crosses a constant boundary, i.e., a threshold value. The computation of these exact threshold values is non-trivial, though some "asymptotic" approximations are available for certain families of distributions.

Two sequential test statistics for the change-point problem are defined now. The well-known Roberts-Shirayev (RS) statistic is given by,

$$R(\theta;n) = \sum_{k=1}^{n}\prod_{i=k}^{n} f_1(\theta_i)/f_0(\theta_i)$$

While the Page-CUSUM (PC) type statistic is given by,

$$P(\theta;n) = \max_{1\le k\le n}\prod_{i=k}^{n} f_1(\theta_i)/f_0(\theta_i)$$

The stopping rule for these test statistics is of the general from,

$$N(A) = \min\{n : T_n \ge A\}$$

where T_n is the RS or PC statistic and the constant A is to be so determined as to control the FAR by satisfying a preassigned value of $P[N(A)>n|\text{No change}, n \text{ specified}]$. Let the onset of the change be at time $n=M$. If $N(A)<M$, then it is a false alarm. Further, the mean delay in detection is given by $E[N(A)-M|N(A)\geq M]$.

Consider now the case of the CN distribution. Let, $f_{k,n}=\exp[\kappa\sum_{i=k}^{n}(\cos(\theta_i-\mu_2)-\cos(\theta_i-\mu_1))]$.

Then in the general form of the RS and PC statistics, $f_{k,n}$ results from the part $\prod_{i=k}^{n} f_1(\theta_i)/f_0(\theta_i)$. In fact the PC statistic simplifies even further to the equivalent form where $f_{k,n}$ is replaced by only the part within the exponent.

For dealing with unknown nuisance parameters, it is suggested that these can be replaced by their MLEs to yield a generalization of RS statistic.

Example

In this section the WB Colliery data is analyzed using the above techniques. The Changeogram (Fig. 1) and the Circular Difference table (Table 1) for this data indicate two suspected change-points quite evidently.

The formal test (LRT) resulted in the value of the test statistic corresponding to the change-point candidate as the 39th observation. This was found significant indicating the possible change in the mean of the distribution with the change-point as the 39th observation.

For the on-line analysis, the Sample Path plot corresponding to the RS statistic is given in Fig. 2. It is quite impressive how quickly this plot detects the change-point-virtually with no delay. Subsequent to the collection of this data, the presence of *Igneous Dyke* has been discovered (Fisher 1993, p. 171) in the region of the 39th observation taken in the tunnel.

Table 1: Circular Difference table for WB Colliery data.

t	cd	t	cd	t	cd	t	cd	t	cd
2	10	15	11	28	9	41	7	54	5
3	10	16	11	29	15	42	3	55	1
4	4	17	10	30	5	43	7	56	11
5	11	18	4	31	0	44	6	57	49
6	16	19	11	32	5	45	0	58	2
7	6	20	42	33	14	46	2	59	12
8	5	21	42	34	12	47	2	60	9
9	4	22	12	35	6	48	43	61	42
10	6	23	6	36	17	49	5	62	10
11	12	24	7	37	12	50	6	63	15
12	12	25	38	38	5	51	6		
13	9	26	35	39	22	52	0		
14	9	27	9	40	56	53	4		

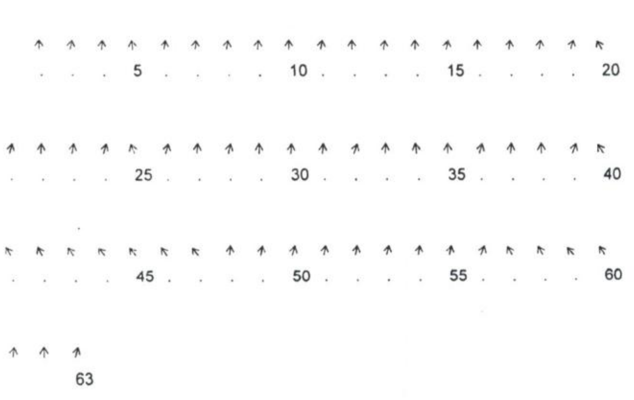

Fig. 1: Changeogram for WB Colliery data.

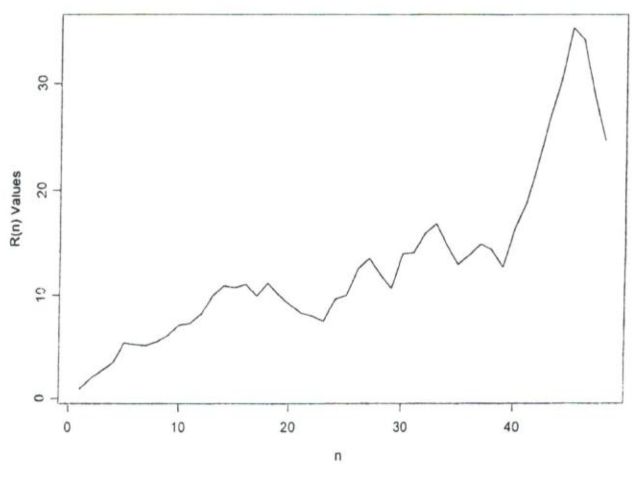

Fig. 2: Sample path for WB Colliery data.

ACKNOWLEDGEMENT

The author is thankful to the editors and referee for several suggestions which have improved the presentation of the paper.

REFERENCES

Fisher, N. I., 1993, *Statistics of circular Data*. Cambridge University Press, Cambridge, 277 p.

Jammalamadaka, S. R. and Sen Gupta, A., 2001, *Topics in Circular Statistics*. World Scientific Publishing Co., New Jersey, 322 p.

Sen Gupta, A., 1998, DDSTAP-*Statistical Analysis Package for Directional Data*. Indian Statistical Institute, Kolkata.

Sen Gupta, A. and Laha, A., 2001, Change-point problems for the mean direction of the circular normal distribution. Under revision for publication.

Shepherd, J. and Fisher, N. I., 1981, A rapid method of mapping fracture trends in collieries. Aust. Coal Miner, August, pp. 24-33.

Shepherd, J. and Fisher, N. I., 1982, Rapid method of mapping fracture trends in collieries. Trans, Soc. Min. Eng. AIME, v. 270, pp. 1931-1932.