Qüestiió - V. 10, n.º 1 (març 1986) pp. 1-6

OUTLIERS IN CIRCULAR DATA: A BAYESIAN APPROACH

G. ARNAIZ TOVAR AND C. RUIZ RIVAS
UNIVERSIDAD AUTÓNOMA DE MADRID

The problem of outliers in circular data is studied from a Bayesian point of view. Surprising observations are identified by means of a predictive measure. On the basis of Box-Tiao methodology, the mean-shift model and some aspects of the contamination of the concentration parameter for a Von Mises distribution are analyzed. Intuitive aspects of the resultant weights and their applications in some classical examples are included.

Keywords: BAYESIAN DIRECTIONAL DATA ANALYSIS. REJECTION OF OUTLIERS. VON MISES DISTRIBUTION.

1. INTRODUCTION.

The statistical analysis of circular data creates specific problems that differ from the usual ones on the line. In particular, Lewis (/8/) notes the specific problem of outliers in this field. Barnett and Lewis (/1/) and Beckman and Cook (/2/) in their extensive work on outliers, include, in the circular data section, the results from -- Collet (/5/) who analyzed the performance of 4 statistics (the likelihood ratio, the one suggested by Mardia (/8/) and two others based on intuitive considerations) for the detection of a single outlier, characterized by the maximum angular deviation from the sample mean direction. Mardia and El-Atoum (/9/) provide some aspects of the Bayesian inference for the Von Mises-Fisher distribution.

The most usual distributions on the circle are the Von Mises and the wrapped Normal which have a close relationship, as analized by Stephens (/11/). The first, statistically the most manageable, has the density function

$$f(\theta/\mu,k)=\frac{1}{2\pi I_o(k)}\exp\{k\cos(\theta-\mu)\},\quad 0\leq\theta<2\pi$$

where $0\leq\mu<2\pi$ (mean direction), $k>0$ (concentration) and I_p is the modified Bessel function of the first kind and order p. Notating $A(k) = I_1(k)/I_o(k)$, $A'(k) = 1-A^2(k)-A(k)/k$ and following Bernardo (/3/) for the reference prior distributions, we shall distin-- guish the cases a) k, a nuisance parameter "a priori" independent of μ $(f(\mu,k)\propto(A'(k))^{1/2})$, b) k a parameter of interest $(f(\mu,k)\propto(kA(k)A'(k))^{1/2}$; Mardia and El-Atoum (/9/)).

2. IDENTIFICATION OF SURPRISING OBSERVATIONS.

One of the greatest practical problems of the Bayesian models for outliers is the computational explosion in the calculus of the posterior distribution, even for small sample sizes. Pettit and Smith (/10/), through considerations on the prior distribution of the contamination parameter and by relating the concepts of "contaminant" and "surprising -- observation", reduce the calculus to the observations identified as "surprising" using a predicative measure as a measure of "surprise".

In our context, given a sample $\underset{\sim}{\theta}=(\theta_1,\ldots,\theta_n)$, we shall notate R, $\bar{\theta}$ the resultant and mean direction of the data and for each $M\subset\{1,2,\ldots,n\} = N$, card. $M = m$, $L = N-M$,

- G. Arnaiz Tovar y C. Ruiz Rivas - Universidad Autónoma de Madrid - Depto. de Estadística - Fac. de Ciencias 28049 MADRID

- Article rebut el novembre de 1985.

$\ell=n-m$, $\theta(M)=\{\theta_i, i\epsilon M\}$ and R_M, $\bar{\theta}_M$ the resultant and mean direction of the data $\theta(M)$.

Let us suppose that the standard model which generates the correct observations is a Von Mises distribution, $M(\mu,k)$, and let us consider a prior density of the form $g(k)$ for (μ,k).

Small values of the predictive measure

$$h(\theta(M)/\theta(L)) = \int f(\theta(M)/\mu,k)\, f(\mu,k/\theta(L))\, d\mu dk \propto$$

$$\propto \frac{1}{E_k\left[\dfrac{I_o(kR_L)}{(I_o(k))^{\ell}}\right]}$$

where E_k is the expectation with respect to the prior distribution of k, indicate that the observations $\theta(M)$ are "surprising" in relation to $\theta(L)$ (and the prior specification for (μ,k)).

As I_o is strictly increasing, the bigger the concentration of the data with labels in L the smaller will result the predictive measure. In otherwords, among all the subsets of m observations, the most surprising in relation to the others will be those which do not belong to the set of the most concentrated ℓ observations. This procedure, in - accord with common sense, implies a successive search for the outliers in the extremes of the maximum arcs between consecutive observations.

3. MEAN-SHIFT MODEL"

Let us suppose that with probability $1-\alpha$ an observation comes from the standard model $M(\mu,k)$ and with probability α from the contamination model $M(\mu+a,k)$, $a\epsilon(0,2\pi)$.

Let B_M be the event "observations $\theta(M)$ are from the contamination model and the $\theta(L)$ from the standard model".

The likelihood may be stated as:

$$f(\underset{\sim}{\theta}/\mu,k,a,B_M)=[2\pi I_o(k)]^{-n}\exp\{kR\gamma_M\cos(\bar{\delta}_M-\mu)\} \qquad (3.1)$$

where $R\gamma_M=(R_L^2+R_M^2+2R_LR_M\cos(\bar{\theta}_M-\bar{\theta}_L-a))^{1/2}$ and $\bar{\gamma}_M$ are the resultant and the mean direction of the transformed data:

$$\gamma_i = \begin{cases} \theta_i, & i \epsilon L \\ \theta_i-a, & i \epsilon M \end{cases} \qquad i = 1,2,\dots,n$$

Following the Box and Tiao (/4/) methodology and taking k,a as nuisance parameters, the posterior density for μ will be expressed as a weighted average of 2^n distributions:

$$f(\mu/\underset{\sim}{\theta}) = \sum_M W_M f(\mu/\underset{\sim}{\theta},B_M)$$

$$W_M = P(B_M/\underset{\sim}{\theta}) = C\left(\frac{\alpha}{1-\alpha}\right)^m \frac{f(\underset{\sim}{\theta}/B_M)}{f(\underset{\sim}{\theta}/B_\phi)}$$

C (normalization constant) such that $\sum_M W_M = 1$. The weights W_M will weight up the outlier quality of each subset, $\theta(M)$, of the data.

Given the likelihood (3.1), the results depend on the prior specification for (μ,k,a). Let us analyze some cases:

3.1. K KNOWN.

Let us suppose a independent of μ and for μ the reference prior $U(0,2\pi)$. Now:

$$w_M \propto \left(\frac{\alpha}{1-\alpha}\right)^m E_a[I_o(kR\gamma_M)]$$

Particularly, if a is known, the factor $I_o(kR\gamma_M)$ has an obvious intuitive sense: in that it will increase with increase in the concentration of the transformed data γ_i. Its maximum value will clearly be obtained for a set $\theta(M)$ displaced approximately an angle "a" from the data mass. In this case, the posterior density for μ will be a weighted sum of densities $M(\bar{\gamma}_M, kR\gamma_M)$.

On the other hand, taking a prior $U(0,2\pi)$ - for a, we have

$$w_M \propto \left(\frac{\alpha}{1-\alpha}\right)^m I_o(kR_M)I_o(kR_L)$$

which will increase with the simultaneous increase in concentration of the sets $\theta(M)$ and $\theta(L)$. In this case:

$$f(\mu/\underset{\sim}{\theta},B_M) \sim M(\bar{\theta}_L,kR_L).$$

Finally, let us suppose that we have a prior knowledge of a expressed by means of a density $M(a_o,k_o)$ with a_o,k_o known, now:

$E_a[I_o(kR_M)] = 2\pi I_o(kR_M)I_o(kR_L)f(0)$

where the factor f(0) represents the value at the 0^o angle of the density function of the random variable $\beta = \beta_1 - \beta_2 - a$ (mod. 2π), $\beta_1 \sim M(\bar{\theta}_M, kR_M)$, $\beta_2 \sim M(\bar{\theta}_L, kR_L)$. If we consider that the convolution of Von Mises distributions may be taken, approximately, as Von Mises (Mardia, /7/), f is approximately $M(\bar{\theta}_M - \bar{\theta}_L - a_o, k_1)$ where k_1 is such that $A(k_1) = A(kR_M)A(kR_L)A(k_o)$.

The posterior density for μ will, in this case, be the weighted sum of the densities:

$$f(\mu/\underset{\sim}{\theta}, B_M) \propto I_o[(k^2R_M^2 + k_o^2 + 2kk_oR_M\cos(\bar{\theta}_M - \mu - a_o)^{1/2}]\exp\{kR_L\cos(\bar{\theta}_L - \mu)\}$$

which are the product of a density $M(\bar{\theta}_L, kR_L)$ and the density of the difference (mod. 2π) of $M(\bar{\theta}_M, kR_M)$ and $M(a_o, k_o)$.

EXAMPLE 1: Estimation of an axis.

The estimation of an axis, though formally different from the outlier problem, may be undertaken by means of the previous model, taking $a=\pi$, $\alpha=1/2$. Let us consider the following example: (Wagner's data quoted by Mardia (/7/, p. 168)).

The vanishing amgles of 13 pigeons released one by one in the Toggenburg Valley were 135, 145, 165, 170, 200, 300, 325, 335, 350, 350, 350, 355, 20. The hypothesis was that the pigeons would fly in the direction of the axis of the valley.

Applying the previous model with $a=180^o$, $\alpha=1/2$, $k=7.13$ (1) (See table 1).

The weights for m=1 provide the identififation of the first five data as clearly distinct from the rest. The weight w_M for $M = \{1,2,3,4,5\}$ is of the order of 10^{27}, the next largest being of the order of 3000 times smaller. In this situation, if the weights are normalized so that $\sum_M w_M = 1$, we will have:

$$f(\mu/\underset{\sim}{\theta}) \simeq f(\mu/\underset{\sim}{\theta}, B_M) \sim M(\bar{\gamma}_M, kR\gamma_M)$$

where $\bar{\gamma}_M = 343.12^0$, $R\gamma_M = 12.05$.

The 0.95 HPD interval for μ is [330.5, 355.8]. The 0.95 confidence interval is [329.2, 357.6] (Mardia (/7/)).

EXAMPLE 2. Data from Ferguson, Landreth and Mckeown (1967) cited in Collet (/5/).

The aim was to investigate the homing ability of the northern cricket frog, Acris crepitans. 14 of them were released after 30 hours enclosure within a dark environmental chamber located to the north west of the collection point. Taking 0^o to be due North, the orientations of the frogs were 104, 110, 117, 121, 127, 130, 136, 145, 152, 178, 184, 192, 200, 316.

Taking k = 2.1737, and different options -- for a, we found the following weights (2) for m = 1: (table 2)

which clearly identifies 316^o as an outlier; in fact, for a = 60 we can identify the observations 10, 11, 12, 13 as a group taking a deviated direction.

3.2. K UNKNOWN.

In this case, assuming "a priori" that k is large we can obtain some approximations for the weights and the posterior distributions.

Let us consider k and a, independent nuisance parameters, "a priori" independent of μ. Given the reference prior distribution for -- (μ,k), $(A'(k))^{1/2}$ which for large k can be taken proportional to 1/k, and in a situation of prior ignorance of a ($a \sim U(0,2\pi)$), we obtain approximately

$$f(\underset{\sim}{\theta}/B_M) \propto \frac{1}{(R_L R_M)^{1/2}} (n-R_M-R_L)^{1-(n/2)}$$

$$w_M \propto \left(\frac{\alpha}{1-\alpha}\right)^m \left(\frac{R(n-R)}{R_M\ R_L}\right)^{1/2} \left(\frac{n-R}{n-R_M-R_L}\right)^{(n/2)-1}$$

the factor $(n-R)/(n-R_M-R_L)$ which obviously has the greatest influence on the weight, may be seen as a generalization, for m outliers, of the statistic suggested by Mardia (/8/) and analyzed by Colled (/5/) to test the discordance of a single deviated observation, θ_k. This statistic is $(n-1-R_{N-\{k\}})/(n-R)$ where k is such that $R_{N-\{k\}}$ =

$\max_i R_{N-\{i\}}$, and agrees, for m = 1, with the largest value of the factor (n-R)/ /(n-1-R_L).

The posterior density for μ is a weighted sum of the densities:

$$f(\mu/\underset{\sim}{\theta},B_M)\propto[n-R_M-R_L\cos(\bar{\theta}_L-\mu)]^{(1-n)/2}$$

Analogously, if a is known

$$w_M\propto\left(\frac{\alpha}{1-\alpha}\right)^m\left(\frac{R}{R\gamma_M}\right)^{1/2}\left(\frac{n-R}{n-R\gamma_M}\right)^{(n-1)/2}$$

$$f(\mu/\underset{\sim}{\theta},B_M)\propto[n-R\gamma_M\cos(\bar{\gamma}_M-\mu)]^{-(n/2)}$$

Let us analyse, in this case, the data of the previous examples:

EXAMPLE 1: a=180°, α=1/2, n=13, R=2.8

Label	Weights (m=1)
1	2.58
2	2.72
3	2.83
4	2.82
5	2.47
6	0.74
7	0.63
8	0.63
9	0.64
10	0.64
11	0.64
12	0.63
13	0.69

M={1 2 3 4 5} , $w_M=7.45\times10^5$, $\bar{\gamma}_M=343.12$,

$R\gamma_M=12.05$

With the next biggest weight of the order of 3000 times smaller, when normalizing the weights, we have

$$f(\mu/\underset{\sim}{\theta})\simeq f(\mu/\underset{\sim}{\theta},B_M)\propto[13-12.05\cos(\mu-343.12)]^{-6.5}$$

By integrating this density numerically, - the 0.95 HPD interval for μ is found to be [328.72, 357.52].

EXAMPLE 2: a∼U(0,2π), n=14, R=10.1527

Label	Weights (m=1)
1	3.20
2	2.86
3	2.55
4	2.42
5	2.26
6	2.20
7	2.12
8	2.07
9	2.08
10	2.67
11	2.97
12	3.49
13	4.23
14	143.16

which again identifies the observation 14 (316°) as an outlier.

4. CONTAMINATION OF THE CONCENTRATION PARAMETER.

Given the standard model, M(μ,k), let us now suppose that, whith probability α, an observation mauy be generated by the contamination model M(μ,ak), 0≤a<1 (for a=0 the contaminants come from a U(0,2π) distribution).

Now, the likelihood is:

$$f(\underset{\sim}{\theta}/\mu,k,a,B_M) =$$

$$=\frac{1}{(2\pi)^n I_o(k)^1 I_o(ka)^m}\exp\{kR_L\cos(\bar{\theta}_L-\mu)+akR_M\cos(\bar{\theta}_M-\mu)\}$$

Let us consider the situation: a known, a>0 and the reference prior density $(kA(k)A'(k))^{1/2}$ for (μ,k) which for large k, may be taken approximately proportional to $1/(k)^{1/2}$ .Assuming "a priori" that k is large, we obtain

$$w_M\propto\left(\frac{\alpha}{1-\alpha}\right)^m a^{m/2}\left(\frac{R}{T}\right)^{1/2}\left(\frac{n-R}{1+am-T}\right)^{n/2}$$

where $T^2 = R_L^2 + a^2R_M^2 + 2aR_MR_L\cos(\bar{\theta}_M-\bar{\theta}_L)$.

This posterior density for k is a weighted sum of Gamma distributions, $\Gamma(\frac{n}{2},\frac{1}{1+am-T})$, and the posterior density for μ is a weighted sum of densities

$$f(\mu/\underset{\sim}{\theta},B_M)\propto\frac{(1+am-T)^{n/2}}{[1+am-T\cos(\beta-\mu)]^{(n+1)/2}}$$

where $T\cos(\beta-\mu) = R_L\cos(\bar{\theta}_L-\mu) + aR_M\cos(\bar{\theta}_M-\mu)$.

For a = 0 we have, analogously:

$$w_M \propto \left(\frac{\alpha}{1-\alpha}\right)^m \left(\frac{R}{R_L}\right)^{1/2} \frac{(n-R)^{n/2}}{(1-R_L)^{1/2}}$$

$$f(k/\underset{\sim}{\theta}, B_M) \sim \Gamma\left(\frac{1}{2}, \frac{1}{1-R_L}\right)$$

$$f(\mu/\underset{\sim}{\theta}, B_M) \propto \frac{(R_L)^{1/2}(1-R_L)^{1/2}}{[1-R_L\cos(\bar{\theta}_L-\mu)]^{(1+1)/2}}$$

5. SUMMARY.

Following the Bos-Tiao methodology, we have analized the mean-shift model and some aproximations for the contamination in the concentration parameter of a Von Mises distribution.

In different situations, expresions for the weights (w_M) which measures the anomaly of the subset M of the data, interpreting thus intuitively the results, have been obtained.

It should be emphasized in the k-unknown case, the appearance in the expression of the weights of a factor interpreted such as a m-outliers generalization of the statistic proposed by Mardia /8/.

Finally, the analysis of some classical examples shows how, from the practical point of view, the calculus in the first place of the individual weights (for different values of the contamination parameter) can be useful for the identification of different groups of outliers. We must also point out that the problem of estimating an axis can be aborded through a particular case from these models.

6. REFERENCES.

/1/ BARNET, V. AND LEWIS, T.: "Outliers in Statistical Data". 2 nd. Edition. Wiley (1984).

/2/ BECKMAN, R.J. AND COOK, R.D.: "Outlier..s" Technometrics, 25, 2, 119-163. (1983).

/3/ BERNARDO, J.M.: "Reference Posterior Distributions for Bayesian Inference". J.R. Stat. Soc., B,41, 2, 113-147 (1979).

/4/ BOX, G.E.P. AND TIAO, G.C.: "A Bayesian Approach to some Outlier Problems". Biometrika, 55, 1, 119-129. (1968).

/5/ COLLED, D.: "Outliers in Circular Data". Appl. Statis. 29, 1, 50-57 (1980).

/6/ DOBSON, A.J.: "Simple Approximations for the Von Mises Concentration Statistic". Appl. Statis., 27, 3, 345-347. (1978).

/7/ MARDIA, K.V.: "Statistics of Directional Data". Academic Press. (1972).

/8/ MARDIA, K.V.: "Statistics of Directional Data", J.R. Stat. Soc., B, 37, 349-393. (1975).

/9/ MARDIA K.V. AND EL-ATOUM, S.A.M.: "Bayesian inference for the Von Mises-Fisher Distribution". Biometrika, 63 1, 203-206 (1976).

/10/ PETTIT, L.I. AND SMITH, A.F.M.: "Ourliers and influential observations in Linear Models". Second Valencia International Meeting on Bayesian Statistics.(1983).

/11/ STEPHENS, M.A.: "Random walk on a circle" Biometrika, 50, 3-4, 385-390. (1963).

/12/ WATSON, G.S.: "Statistics on spheres". Wiley. (1983).

6.NOTES.

(2) Non-normalized weights. The constant $\frac{\alpha}{1-\alpha}$ is not included.

(1) We consider k known for illustrative purposes.The value 7.13 is obtained from from the maximum likelihood estimator of the concentration with the first five observations shifted 180° (calculated following Dobson, /6/)).

TABLE 1

Label	Data	Weights (m=1) Non normalized weights
1	135	0.41 E+06
2	145	0.75 E+06
3	165	0.12 E+07
4	170	0.11 E+07
5	200	0.24 E+06
6	300	0.29 E-02
7	325	0.10 E-04
8	335	0.21 E-05
9	350	0.15 E-05
10	350	0.15 E-05
11	350	0.15 E-05
12	355	0.23 E-05
13	20	0.37 E-03

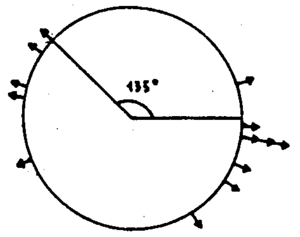

TABLE 2

Label	$a \sim U(0,2\pi)$	$a = 60^{\circ}$	$a = 180^{\circ}$
1	0.5601	0.1345	0.0531
2	0.4817	0.1468	0.0384
3	0.4141	0.1673	0.0276
4	0.3848	0.1827	0.0235
5	0.3512	0.2118	0.0192
6	0.3385	0.2296	0.0177
7	0.3204	0.2731	0.0157
8	0.3093	0.3636	0.0145
9	0.3133	0.4621	0.0149
10	0.4408	1.1960	0.0316
11	0.5061	1.4882	0.0427
12	0.6263	1.9748	0.0675
13	0.7972	2.5830	0.1125
14	20.9982	4.1505	66.8600