INFORMATICA, 1991, Vol.2, No.1, 066–076

ROBUST ESTIMATION OF A CHANGE POINT IN THE PROPERTIES OF AUTOREGRESSIVE SEQUENCES

Joana LIPEIKIENĖ

Institute of Mathematics and Informatics,
Lithuanian Academy of Sciences,
232600, Vilnius, Akademijos St.4, Lithuania

Abstract. The problem of determination of a change point in the properties of autoregressive sequences with unknown distribution is analysed. Two robust algorithms for the estimation of a change point when the distribution is symmetric and asymmetric are presented.

Key words: autoregressive sequence, change point, robust estimation.

1. Introduction. Measurements of operational regimes in technical and biological systems in practical situations are usually characterized by the presence of outliers, i.e., measurements with large errors. It is characteristic not only of independent but also correlated data. The values abruptly differing among the observed ones influence the accuracy of statistical inferences. The problem of robust estimation of autoregressive parameters is reflected in a great many of papers (e.g., Martin, 1979, 1982; Heathcote, 1983). Our aim was to determine the robust estimates of change points in the properties of autoregressive sequences with symmetric and asymmetric distribution and to compare their exactness with the maximum likelihood estimates of the change points.

2. Statement of the problem. We have a sequence $\{X_t\}$ expressed by the autoregressive equation

$$X_t + \sum_{j=1}^{p} a_j^{(i)} X_{t-j} = b^{(i)} V_t, \tag{1}$$

where $\{V_t\}$ denotes the sequence of mutually independent identically distributed variables with an unknown (symmetric or asymmetric) distribution $P(z)$. The parameters of the sequence change at an unknown instant of time u as follows:

$$\begin{aligned} A_i &= (a_j^{(i)}, (j = 1, \dots, p), b^{(i)}) = \\ &= \begin{cases} A_1, & t = \dots, 1, 2, \dots, u \\ A_2, & t = u+1, \dots, N, \dots \end{cases} \end{aligned} \tag{2}$$

Given the realization $\{x_t\}$ $(t = 1, \dots, N)$ of the random sequence $\{X_t\}$, it is required to find the estimate $\widehat{u}$ of a change point u.

3. Solution of the problem. Telksnys (1973), Kligienė (1973), Lipeika (1979) assumed the distribution of the sequence $\{V_t\}$ to be normal, the problem was to find the estimate of u yielding minimal mean risk: that is, the maximum point of the likelihood function (Kligienė, 1973; Lipeika, 1979) or same of the aposteriori probability density (Telksnys, 1973).

In case the distribution $P(z)$ is not known, the realizations $\{x_t\}$ of the sequence $\{X_t\}$ may yield outliers. Hence it is natural to determine the estimate of a change point so as to make the influence of abruptly differing values insignificant. Here we must separate the cases of symmetric and asymmetric distribution of the sequence $\{V_t\}$. In the case of symmetric distribution we may use an idea of M-estimates determined for symmetric distribution (Martin, 1982).

M-estimate $\widehat{u}$ of change point u. (The case of symmetric distribution). By analogy with (Martin, 1982) we suggest to determine M-estimates of a change point u, using a stabilizing loss function. By an M-estimate $\widehat{u}$ of a change point u we mean the estimate which minimizes a general function of prediction errors

$$S(u) = \sum_{t=p+1}^{N} \rho(V_t^{(i)}), \tag{3}$$

where

$$V_t^{(i)} = \frac{X_t + \sum\limits_{j=1}^{p} a_j^{(i)} X_{t-j}}{b^{(i)}}, \quad i = \begin{cases} 1, & t = p+1, \dots, u, \\ 2, & t = u+1, \dots, N, \end{cases}$$

$\rho(\cdot)$ denotes the stabilizing function of the prediction errors $V_t^{(i)}$. The flexible choice of this function enables one to reduce the influence of outliers while computing the time when properties change. Thus the estimate $\widehat{u}$ is defined as the minimal point of the function $S(u)$ over all feasible times of change u.

By definition (3),

$$S(u) = \sum_{t=p+1}^{N} \rho(V_t^{(i)}) = \sum_{t=p+1}^{u} \rho(V_t^{(1)}) + \sum_{t=u+1}^{N} \rho(V_t^{(2)}),$$

$$S(u+1) = \sum_{t=p+1}^{u} \rho(V_t^{(1)}) + \rho(V_{u+1}^{(1)}) + \sum_{t=u+1}^{N} \rho(V_t^{(2)}) - \rho(V_{u+1}^{(2)}),$$

$$S(u+1) = S(u) + \rho(V_{u+1}^{(1)}) - \rho(V_{u+1}^{(2)}). \tag{4}$$

Therefore, the computation of $S(u)$ for all feasible times of change $u = p+1, \dots, N$ can be recursive. To determine

$S(u+1)$, it is sufficient to compute $\rho(V_{u+1}^{(1)})$ and $\rho(V_{u+1}^{(2)})$ and to use formula (4).

The choice of estimation function $\rho(\cdot)$**.** As it was mentioned above, the function $\rho(\cdot)$ must reduce the influence of outliers but generally the choice is free. Martin (1979) assumed $\rho(\cdot)$ to be nonnegative, symmetric and twice differentiable for the estimates of autoregressive parameters were consistent and asymptotically normal. We tried to use a few most popular functions and considered the influence of estimation function $\rho(\cdot)$ on the exactness of estimates $\widehat{u}$. For this purpose we have used the following functions:
Huber (1964)

$$\rho_{Hu}(V_k) = \begin{cases} \frac{V_k^2}{2}, & |V_k| \leqslant l, \\ l\,|V_k| - \frac{l}{2}, & |V_k| > l, \end{cases} \tag{5}$$

Beaton–Tukey (1974)

$$\rho_{B-T}(V_k) = \begin{cases} \frac{V_k^2}{2}(1 - \frac{V_k^2}{a^2} + \frac{V_k^4}{a^4}), & |V_k| \leqslant a, \\ 0, & |V_k| > a, \end{cases} \tag{6}$$

Andrews (1972)

$$\rho_A(V_k) = \begin{cases} -\frac{1}{b}\cos(V_k|b), & |V_k| \geqslant b\pi, \\ 0, & |V_k| > b\pi, \end{cases} \tag{7}$$

and the square function

$$\rho(V_k) = \frac{V_k^2}{2}, \tag{8}$$

where the constants a, b, l may be chosen.

Lipeikienė (1986) presented the investigation results on the influence of estimation function $\rho(\cdot)$ on the exactness of a change point M-estimates $\widehat{u}$. There were simulated realizations of autoregressive sequences with ε-contaminated distribution $P(z) = (1-\varepsilon)N(0,1) + \varepsilon N(0,\sigma^2)$ and abruptly changing parameters $A_i(2)$. Mean square devitation Δ of the estimate $\widehat{u}$

$$\Delta = \sqrt{\frac{1}{n}\sum_{k=1}^{n}(u-\widehat{u}^{(k)})^2}, \tag{9}$$

where n is the number of generated realization, u is the change point, $\widehat{u}^{(k)}$ is the change-point M-estimate in the k-th realization, was used as a measure of exactness. We considered the situations, when a change of parameters was small, more significant and large. In all these situations the best estimates of a change point were obtained when Huber estimation function ρ_{Hu} (5) was used. It must be noted that in the tables of Δ (Lipeikienė, 1986) there is no great difference between the least square estimates and M-estimates of a change-point. We can see some gain of using M-estimates when contamination is greater (e.g., when $\varepsilon = 0.2, 0.5$). For the case of asymmetric distribution the results are different.

Functional least square estimate of a change point. (The case of asymmetric distribution). We use an idea of functional least square estimates of the autoregressive parameters (Heathcote, 1983), which are introduced for random sequences with asymmetric distribution. Thus we define the functional least square estimate u^* of a change point u as a minimum point of the functional

$$L_N(u,s) = -s^2 \log\left|(N-p)^{-1}\sum_{t=p+1}^{N}\exp(isV_t^{(i)})\right|^2, \tag{10}$$

where $u = p+1, \ldots, N$ and s-real parameter from an interval S, which does not include the origin $s = 0$.

Thus we must minimize the functional

$$L_N(u,s) = -s^{-2}\log\{U_N^2(u,s) + V_N^2(u,s)\}, \qquad (11)$$

where

$$U_N(u,s) = \frac{1}{N-p}\Big[\sum_{j=p+1}^{u} \cos sV_j^{(1)} + \sum_{j=u+1}^{N} \cos sV_j^{(2)}\Big], \quad (12)$$

$$V_N(u,s) = \frac{1}{N-p}\Big[\sum_{j=p+1}^{u} \sin sV_j^{(1)} + \sum_{j=u+1}^{N} \sin sV_j^{(2)}\Big], \quad (13)$$

and compute $V_j^{(1)}$ from the equation (1):

$$V_j^{(i)} = \frac{X_t - \sum\limits_{j=1}^{p} a_j^{(i)} X_{t-j}}{b^{(i)}},$$

using the parameters A_1 before the admitted change point u and the parameters A_2 after point u. Thus, computing $L_N(u,s)$ over all feasible change points u, we get a set $\{u^*(s), s \in S\} = \operatorname*{arg\,min}\limits_{u=p+1,\dots,N} L_N(u,s)$ and it remains to choose an optimal meaning s_1. We shall discuss this question further.

The recurrent formulas reduce the amount of computation what follows from the definition of $L_N(u,s)$. Let us define

$$U_N(u,s) = \frac{1}{N-p} A_N(u,s),$$
$$V_N(u,s) = \frac{1}{N-p} B_N(u,s),$$

i.e.,

$$A_N(u,s) = \sum_{j=p+1}^{u} \cos\, sV_j^{(1)} + \sum_{j=u+1}^{N} \cos\, sV_j^{(2)}, \qquad (14)$$

$$B_N(u,s) = \sum_{j=p+1}^{u} \sin\, sV_j^{(1)} + \sum_{j=u+1}^{N} \sin\, sV_j^{(2)}. \qquad (15)$$

Thus

$$A_N(u+1,s) = A_N(u,s) + \cos sV^{(1)}_{u+1} - \cos sV^{(2)}_{u+1}. \quad (16)$$

By analogy

$$B_N(u+1,s) = B_N(u,s) + \sin sV^{(1)}_{u+1} - \sin sV^{(2)}_{u+1}. \quad (17)$$

Computing $L_N(u,s)$ over all feasible $u = p+1,\ldots,N$ and using formulas (16), (17), we reduce significantly the computation amount.

Recall that for computing $A_N(u,s)$ and $B_N(u,s)$ beginning from $u = p+1$ we must use real initial conditions

$$A_N(p,s) = \sum_{t=p+1}^{N} \cos sV^{(2)}_t, \quad (18)$$

$$B_N(p,s) = \sum_{t=p+1}^{N} \sin sV^{(2)}_t \quad (19)$$

for the true computation of $L_N(u,s)$.

The choice of parameter s. After determination of the set $\{U^*, s \in S\}$ we must choose s_1 from a set S. It is natural to choose s_1 optimal in a sense of mean square deviation of the estimate $\widetilde{u} = u^*(s_1)$. To get an explicit expression of the mean square deviation of the estimate $\widetilde{u}$, we have used the method of mathematical simulation. Namely, we simulated n realizations of random sequences with parameters changing at any point u_1 (parameters were assumed to be known). For every realization we determine $u^*(s)$ for every $s \in S$ and, as u_1 is known, we determine the optimal s_1, i.e., such s which minimizes the deviation

$$\Delta = \sqrt{\sum_{i=1}^{n}(u_i(s) - u_1)^2}. \quad (20)$$

Then we use s_1 to determine $\widehat{u}$ in the realization investigated.

4. Simulation results. In order to investigate the performance of both methods for the detection of a change point (M-estimate and functional least square estimate), a simulation study (100 replications) was performed for samples of the size $N = 1000$. The samples were generated by the autoregressive model (1) with different symmetric and asymmetric distributions. The exactness of estimates was investigated by the calculation of the mean square deviation (9) of three estimates: M-estimates, functional least square estimates and ordinary least square estimates. It is interesting to see the behaviour of M-estimates in the cases of asymmetric distribution. Let us consider two typical examples.

Example 1. The samples were generated by the fourth-order autoregressive model with the parameters $A_1 = (a_1^{(1)} = 1,\ a_2^{(1)} = 1.06,\ a_3^{(1)} = 0.42,\ a_4^{(1)} = 0.2,\ b^{(1)} = 1)$ before a change moment $u = 500$, and with the parameters $A_2 = (a_1^{(2)} = 0.7,\ a_2^{(2)} = 1.02,\ a_3^{(2)} = 0.2,\ a_4^{(2)} = 0.32,\ b^{(2)} = 1)$ after a change point. We simulated realizations with the following distributions: normal χ_k^2 $(k = 2,3,4)$, ε-contaminated $0.9N(0,1) + 0.1N(10,10)$ and lognormal. In Table 1 you will find our result on the meanings of deviation Δ of the three estimates for all these distributions. M-estimates were computed using Huber function (5). The meanings of Δ in Table 1 show functional least square estimates to be much better than least square and M-estimates, though the results for normal distribution are opposite.

Example 2. The samples were generated by the second-order autoregressive model with the parameters $A_1 = (a_1^{(1)} = 0.9,\ a_2^{(1)} = 0.7,\ b^{(1)} = 1)$ before a change moment $u = 500$, and with the parameters $A_2 = (a_1^{(2)} = 0.7,\ a_2^{(2)} = 0.7, b^{(2)} = 1)$ after a change point u. The simulation study was the same.

Table 1. Meanings of deviation Δ

Distribution	s_1	Δ of M-estimates	Δ of least square estimates	Δ of functional last square estimates
Normal	1	28	22	25
Normal	0.5	28	22	23
Normal	0.05	28	22	22
χ_2^2	0.5	22	29	17
χ_3^2	0.5	32	23	20
χ_4^2	0.2	40	24	21
ε-contaminated $0.9N(0,1)+$ $0.1N(10,10)$	1	15	31	7
Lognormal	1	20	31	11

The results are presented in Table 2. According to them, we can make similar conclusions: in the presence of asymmetric error distribution a functional least square estimate performs much better than an ordinary least square estimate or an M-estimate.

s_1 was chosen according to the simulation method described in §3. The interval S from which we choose s_1 was $(0.05, 3)$. The simulation investigation showed that the best interval S for searching the optimal s_1 is $S = [0.05, 1.5]$ because Δ considerably increases if s_1 takes meanings outside this interval.

Similar results were obtained in all the investigated situations.

5. Conclusions. The introduced method of determining a change point in the properties of autoregressive sequen-

Table 2. Meanings of deviation Δ

Distribution	S_1	Δ of M-estimates	Δ of least square estimates	Δ of functional last square estimates
Normal	1	50	43	57
Normal	0.5	50	43	57
χ_2^2	1	43	39	31
χ_3^2	0.75	79	65	42
χ_4^2	0.35	43	39	31
ε-contaminated $0.9N(0,1)+$ $0.1N(10,10)$	1	18	59	12
Lognormal	1	32	48	18

ces extend the opportunities of the autoregressive model for the solution of practical problems. The simulation results showed least square estimates to be not very sensitive to a small number of outliers when the distribution of a sequence is symmetric. However, the use of robust functional least square estimates for the determination of a change point in the properties of autoregressive sequences with asymmetric distribution is evident.

REFERENCES

Andrews, D.D., P.J.Bickel, F.R.Hampel, P.J.Huber, W.H.Rodgers and J.W.Turkey (1972). *Robust Estimates of Location.* Survey and Advances. University Press, Princeton.

Beaton, A.E., and J.W.Tukey (1974). The fitting of power series, meaning polynomials, illustrated on bandspectroscopic data. *Technometrics,* **16**(2), 147–185.

Heathcote, C.R., and A.H.Welsh (1983). The robust estimation of autoregressive processes by functional least squares. *J. Appl. Prob.,* **20**, 737–753.

Huber, P.J. (1964). Robust estimation of a location parameter. *Ann. Math. Stat.,* **35**, 73–101.

Kligienė, N. (1973). *Estimation of a Change Point in Parameters of Random Sequence Distribution.* Cand. Sci. theses. Vilnius University, Vilnius (in Russian).

Lipeika, A. (1979). On the determination of changes in the properties of an autoregressive sequence. In A.Telksnys (Ed.), *Statisticheskie Problemy Upravleniya,* Vol.39. Inst. Math. Cybern. Lithuanian Acad. Sci., Vilnius. pp. 9–23 (in Russian).

Lipeikienė, J. (1986). The influence of evaluation functions of the exactness of M-estimates. In *Statisticheskie Problemy Upravleniya,* Vol.77. pp. 35–46 (in Russian).

Martin, R.D. (1979). Robust estimation for time series autoregressions. In R.L.Launer and G.N.Wilkinson (Eds.), *Robustness in Statistics,* Academic Press, New York. pp. 147–176.

Martin, R.D. (1982). The Cramer-Rao bound and robust M-estimates for autoregressions. *Biometrika,* **69**(2), 437-442.

Telksnys, L. (1973). Determination of changes in the properties of random processes by the Bayes method. In *Proceedings of the Third IFAC Symposium,* The Hague Delf, Netherlands.

Received April 1990

J.Lipeikienė is a senior researcher at the Recognition Processes Department in the Institute of Mathematics and Informatics, Lith. Acad. Sci. She received the degree of Candidate of Technical Sciences from the Kaunas Polytechnic Institute, Kaunas, Lithuania, in 1980.