Ann. Inst. Statist. Math.
34 (1982), Part B, 389-398

A QUASI BAYESIAN APPROACH TO OUTLIER DETECTION

GENSHIRO KITAGAWA AND HIROTUGU AKAIKE

(Received Aug. 1, 1981; revised Dec. 14, 1981)

Summary

A quasi Bayesian procedure is developed for the detection of outliers. A particular Gaussian distribution with ordered means is assumed as the basic model of the data distribution. By introducing a definition of the likelihood of a model whose parameters are determined by the method of maximum likelihood, the posterior probability of the model is obtained for a particular choice of the prior probability distribution. Numerical examples are given to illustrate the practical utility of the procedure.

1. Introduction

The problem of outlier detection attracted much attention because of its practical and conceptual importance. Numerous papers treated this problem from the traditional testing point of view. However, in the detection of unknown number of multiple outliers, a severe difficulty is caused by the so-called masking effect (Tietjen and Moore [15]) and none of the solutions hitherto proposed is considered to be entirely satisfactory (Hawkins [12]). The work by Box and Tiao [7] demonstrated the importance of formulating the problem explicitly in terms of a Bayesian modeling. Recently Freeman [8] reviewed three Bayesian models of outliers in data from the linear model discussed by Box and Tiao [7], Abraham and Box [1] and Guttman, Dutter and Freeman [11]. In the paper Freeman noticed the difficulty of handling improper prior distributions of parameters when the numbers of the parameters are different among the models. Box and Tiao [7] and Abraham and Box [1] avoided this difficulty by considering only the models with a fixed number of parameters. In Guttman, Dutter and Freeman [11], an ad hoc procedure is developed to get information on the number of outliers.

Although Freeman [8] tried to avoid the difficulty by using proper priors throughout, the difficulty of choosing an appropriate set of priors is clearly demonstrated by his numerical examples. In the discussion

of his paper it is mentioned that it might be that attempts like the AIC criterion to produce a standard way of answering a wide variety of questions regardless of their different contexts are doomed to failure. This comment sounds rather out of place as it is the importance of proper modeling and objective criterion for model selection that has been stressed by the introduction of AIC.

The purpose of the present paper is to show that by developing an appropriate family of models and using the concept of the likelihood of a model obtained by properly extending the basic idea of the AIC criterion (Akaike [4]) we can in fact develop a procedure which practically avoids the difficulty discussed by Freeman. The predictive log likelihood of a model determined by specifying the set of assumed outliers and applying the method of maximum likelihood, is defined as an approximately unbiased estimate of the expected log likelihood of the model. A quasi Bayesian procedure is then realized by using the predictive likelihoods and appropriately chosen prior probabilities of the models. Numerical results are given to show that the procedure produces reasonable results for some of the examples frequently discussed in the literature.

2. The model and its predictive likelihood

Let $x^t=(x_1,\cdots,x_n)$ be a vector of n observations. J is an ordered set of k integers $\{i_1,\cdots,i_k\}$ chosen from the integers $\{1,2,\cdots,n\}$. When J is specified, we consider that k observations $x_{i_1},\cdots,x_{i_k}$ are outliers, whereas the rests, x_j $(j\notin J)$, are normal observations drawn from a Gaussian distribution with unknown mean μ_0 and variance σ^2. The outliers are assumed to be obtained from Gaussian distributions with ordered means, $\mu_1\leqq\mu_2\leqq\cdots\leqq\mu_k$, and common variance σ^2. Thus the component model given J is specified by the data distribution

$$f(x\,|\,J,\theta)=\prod_{j\notin J}\frac{1}{\sigma}\phi\left[\frac{x_j-\mu_0}{\sigma}\right]\prod_{j=1}^{k}\frac{1}{\sigma}\phi\left[\frac{x_{i_j}-\mu_j}{\sigma}\right]$$

where $\theta=(\mu_0,\mu_1,\cdots,\mu_k,\sigma^2)$ and $\phi(x)$ denotes the standard Gaussian density function.

Given the observation, $x_1,\cdots,x_n$, the log likelihood of the component model is given by

$$\log f(x\,|\,\theta,J)=-\frac{n}{2}\log 2\pi\sigma^2-\frac{1}{2\sigma^2}\left\{\sum_{j\notin J}(x_j-\mu_0)^2+\sum_{j=1}^{k}(x_{i_j}-\mu_j)^2\right\}.$$

Thus the maximum likelihood estimate of the mean of normal observations, μ_0, is obtained by $\hat{\mu}_0=(1/(n-k))\sum_{j\notin J}x_j$. While those of the mean

values of outliers, $\hat{\mu}_i$ $(i=1,\cdots,k)$, are obtained by maximizing the log likelihood function under order restrictions. This is equivalent to finding the solution to the problem of quadratic programming:

$$\text{minimize} \qquad F(\mu_1,\cdots,\mu_k)=\sum_{j=1}^{k}(\mu_j-x_{i_j})^2$$

$$\text{subject to} \qquad \mu_1\leqq\mu_2\leqq\cdots\leqq\mu_k\ .$$

They are easily obtained numerically by the pool-adjacent-violators algorithm (Barlow et al. [5]) which will be briefly described in Section 4. For the model specified by $J=\{i_1,\cdots,i_k\}$ which satisfies the natural ordering condition, $x_{i_1}\leqq x_{i_2}\leqq\cdots\leqq x_{i_k}$, they are simply given by $\hat{\mu}_j=x_{i_j}$. The maximum likelihood estimate of the variance σ^2 is then obtained by

$$\hat{\sigma}^2=\frac{1}{n}\left\{\sum_{j\notin J}(x_j-\hat{\mu}_0)^2+\sum_{j=1}^{k}(x_{i_j}-\hat{\mu}_j)^2\right\}\ .$$

According to the entropy maximization principle [3], we evaluate the goodness of the model specified by these maximum likelihood estimates by its expected log likelihood

$$(1)\qquad \mathrm{E}_y\log f(y\,|\,J,\hat{\theta})=-\frac{n}{2}\log 2\pi\hat{\sigma}^2 -\frac{1}{2\hat{\sigma}^2}\left\{n\sigma^2+(n-k)(\mu_0-\hat{\mu}_0)^2+\sum_{j\in J}(\mu_j-\hat{\mu}_j)^2\right\}\ ,$$

where E_y denotes the expectation under the assumed distribution of y, $f(y\,|\,J,\theta)$. In a practical situation, the true parameter $\theta=(\mu_0,\mu_1,\cdots,\mu_k,\sigma^2)$ is unknown and thus the present form of the expected log likelihood is useless. Following the idea underlying the definition of AIC ([2], [3]) we try to correct the bias of the maximized log likelihood,

$$(2)\qquad \log f(x\,|\,J,\hat{\theta})=-\frac{n}{2}\log 2\pi\hat{\sigma}^2-\frac{n}{2}\ ,$$

as an estimate of the expected log likelihood. From (1) and (2), the average increase of the maximum log likelihood is obtained by

$$\begin{aligned}C_{k,n}&=\mathrm{E}_x\{\log f(x\,|\,J,\hat{\theta})-\mathrm{E}_y\log f(y\,|\,J,\hat{\theta})\}\\&=\mathrm{E}_x\left[\frac{1}{2\hat{\sigma}^2}\left\{n\sigma^2+(n-k)(\mu_0-\hat{\mu}_0)^2+\sum_{j=1}^{k}(\mu_j-\hat{\mu}_j)^2-\frac{n}{2}\right\}\right]\end{aligned}$$

where E_x denotes the expectation under the assumed distribution $f(x|\,J,\theta)$ of the data. However, since the maximum likelihood estimates $\hat{\mu}_1,\cdots,\hat{\mu}_k$ depend in a complicated way on the magnitudes of $x_{i_1},\cdots,x_{i_k}$, it is difficult to evaluate $C_{k,n}$ analytically. Therefore for the present

purpose of evaluation of the bias, we will use the simple estimator $\hat{\mu}_j=x_{i_j}$ $(j=1,\cdots,k)$ irrespectively of the ordering of the magnitudes of $x_{i_1},\cdots,x_{i_k}$. This is equivalent to assuming that our model satisfies the natural ordering $x_{i_1}\leqq x_{i_2}\leqq\cdots\leqq x_{i_k}$ and will produce a reasonable approximation when $\sigma^{-1}(\mu_i-\mu_{i-1})$ $(i=2,\cdots,k)$ are sufficiently large. The effect of this assumption will be checked in Section 6.

Under the assumption that the original data x was drawn from the distribution $f(x|J,\theta)$, we have $\mathrm{E}_x[\hat{\sigma}^{-2}]=n(n-k-3)^{-1}\sigma^{-2}$, $\mathrm{E}_x[(\hat{\mu}_0-\mu_0)^2]=(n-k)^{-1}\sigma^2$ and $\mathrm{E}_x[(\hat{\mu}_i-\mu_i)^2]=\sigma^2$, and it follows that

$$C_{k,n}=\frac{n(k+2)}{n-k-3}\ .$$

An unbiased estimate of the expected log likelihood of the estimated model is now obtained as

$$\log f(x|J,\hat{\theta})-C_{k,n}$$

and the predictive likelihood of the estimated model under the assumption of J is defined by

$$p(x|J)=\exp\{\log f(x|J,\hat{\theta})-C_{k,n}\}\ .$$

3. The prior and posterior probabilities

We assume that we have no information initially to say that a specific k, the number of outliers, is more likely than others. Hence we put $p(0)=p(1)=\cdots=p(n)=1/(n+1)$, where $p(k)$ denotes the prior probability that there are k outliers. Given that there are k outliers, there are ${}_nC_k$ ways of specifying k observations as outliers out of the n. Thus the prior probability that a specific set of k observations are the outliers is given by

$$\frac{1}{n+1}{}_nC_k^{-1}=\frac{(n-k)!k!}{(n+1)!}\ .$$

This prior probability admits another derivation. We assume that an observation is an outlier with probability α. The probability that a particular set of k observations are the outliers is then given by $\alpha^k(1-\alpha)^{n-k}$. By integrating $\alpha^k(1-\alpha)^{n-k}$ over 0 through 1, we obtain the above prior probability.

For each set of k assumed outliers there are $k!$ ways of assigning them to the k distributions specified by the means μ_j $(j=1,\cdots,k)$. By assuming every configurations to be equally probable, we obtain the prior probability of the model specified by $J=\{i_1,\cdots,i_k\}$ as

$$\pi(j)=\frac{(n-k)!}{(n+1)!}\ .$$

The posterior probability of the model specified by J is then given by $\pi(J|x)=p(x)^{-1}p(x|J)\pi(J)$, where $p(x)=\sum_J p(x|J)\pi(J)$ and $p(x|J)$ denotes the predictive likelihood of the model defined in the preceding section. The posterior probability of $x_{i_1},\cdots,x_{i_k}$ being the outliers is given by $\sum\pi(J|x)$, where the summation extends over the set of $k!$ J's obtained by permuting $\{i_1,\cdots,i_k\}$.

4. Algorithm

In this section we will describe an algorithm for the computation of posterior probabilities.

(1) *Specification of outliers*

For $m=0,1,\cdots,2^n-1$, put $ind(i)$ $(i=1,\cdots,n)$ equal to the ith bit of the binary expansion of m, i.e., $\sum_{i=1}^{n}2^{i-1}ind(i)=m$. The number of outliers is given by $k=\sum_{i=1}^{n}ind(i)$ and the set of outliers $\{x_{i_j};\ j=1,\cdots,k\}$ is specified by putting $i_j=i$ for the jth non-zero $ind(i)$.

(2) *Computation of the log posterior probability of the naturally ordered model*

For the given combination $\{i_1,\cdots,i_k\}$ of assumed outliers, compute the logarithm of the posterior probability of the naturally ordered model by

$$\log p(m|x)=-\frac{n}{2}\log\sigma^2(m)-\frac{n(k+2)}{n-k-3}-\log n!+\log(n-k)!\ ,$$

where a common additive constant is ignored and

$$\sigma^2(m)=\frac{1}{n}\sum_{ind(j)=0}(x_j-\mu(m))^2\ ,\qquad \mu(m)=\frac{1}{n-k}\sum_{ind(i)=0}x_j\ .$$

(3) *Computation of the posterior probability*

The posterior probability of another model, specified by $J(j)=\{jnd\cdot(1),\cdots,jnd(k)\}$ which is obtained by rearranging $\{i_1,\cdots,i_k\}$, is obtained by using the weight, $w(j,m)$, relative to that of the naturally ordered model

$$w(j,m)=\left\{\frac{n\sigma^2(m)+\sum_{i=1}^{k}(x_{jnd(i)}-\hat{\mu}_i)^2}{n\sigma^2(m)}\right\}^{-n/2},$$

where the maximum likelihood estimates $\hat{\mu}_i$ $(i=1,\cdots,k)$ are obtained by the following pool-adjacent-violators algorithm (Barlow et al. [5]).

The algorithm starts with the initial estimates $\mu_i=x_{jnd(i)}$ $(i=1,\cdots,k)$. If the initial estimates satisfy the order condition, they are the final estimates $\hat{\mu}_i$ $(i=1,\cdots,k)$. If not, select all the sequence of violators of the ordering; that is, select all the pairs of p and q such that $\mu_{p-1}\leqq\mu_p>\mu_{p+1}>\cdots>\mu_q\leqq\mu_{q+1}$. For every pair of p and q, replace the estimates μ_i $(i=p,\cdots,q)$ by the pooled one $\frac{1}{q-p+1}\sum_{r=p}^{q}\mu_r$. If the resulting μ_i's do not yet satisfy the order condition, repeat the above step until to produce the final estimates $\hat{\mu}_i$ $(i=1,\cdots,k)$.

The posterior probability of x_{i_j} $(j=1,\cdots,k)$ being the outliers is obtained by

$$\pi(m\,|\,x)\propto p(m\,|\,x)\sum_{j}w(j,\,m)\ ,$$

where the summation extends over $k!$ possible models specified by $J(j)$'s.

5. Examples

To check the performance of the present procedure, we applied it to two sets of familiar data. The computation was performed by a computer program, OUTLAP, developed by Kitagawa [14].

Example 1 (Darwin's data).

Table 1. Darwin's data

−67	−48	6	8	14	16	23	24
28	29	41	49	56	60	75	

This data set has been discussed by many researchers, such as Box and Tiao [7], Abraham and Box [1] and Freeman [8]. Here the data $x_1, x_2,\cdots, x_n$ are ar-

Table 2. Prior and posterior probabilities of some combinations of possible outliers (Darwin's data)

Number of outliers	Possible outliers	Prior probability	Posterior probability	w
2	−67, −48	0.0016	0.515	1.787
0	none	0.1667	0.140	1.
1	−67	0.0111	0.107	1.
3	−67, −48, 75	3.7×10^{-4}	0.085	1.697
1	−48	0.0111	0.017	1.
3	−67, −48, 6	3.7×10^{-4}	0.014	1.988
3	−67, −48, 60	3.7×10^{-4}	0.012	1.761
3	−67, −48, 8	3.7×10^{-4}	0.011	1.973
3	−67, −48, 56	3.7×10^{-4}	0.009	1.772
4	−67, −48, 60, 75	1.2×10^{-4}	0.007	2.795

ranged in ascending order of magnitude. The lowest two observations look rather discrepant from the rest. The posterior probabilities of various possible combinations of outliers are obtained by multiplying the predictive likelihoods, $\exp\{\log f(x|J)-C_{k,n}\}$, by the prior probabilities $\pi(J)$. The ten largest posterior probabilities which together constitute 91.7% of the total probability are listed in Table 2. The highest posterior probability 0.515 occurs at $J=\{1, 2\}$, indicating that -67 and -48 are probably the outliers. The posterior probabilities of $J=\phi$ (no outliers), $\{1\}$ and $\{1, 2, 15\}$ are also considerable. The probabilities of other combinations of assumed outliers are almost negligible. The marginal posterior probabilities $\pi(i|x)$ that the observation x_i is an outlier are obtained as $\pi(1|x)=0.812$, $\pi(2|x)=0.705$, $\pi(15|x)=0.120$, $\pi(14|x)=0.030$ and $\pi(i|x)<0.022$ for $i=3, 4, \cdots, 13$. The meaning of the values in the last column denoted by w will be explained in Section 6.

Example 2 (Herndon's data).

The second example is Herndon's data discussed by Grubbs [9], Tietjen and Moore [15] and Kitagawa [13]: Here the extreme values

Table 3. Herndon's data

−1.40	−0.44	−0.30	−0.24	−0.22	−0.13	−0.05	0.06
0.10	0.18	0.20	0.39	0.48	0.63	1.01	

on both sides are suspicious. The ten largest posterior probabilities which together constitute 85.1% of the total probability is listed in Table 3. The highest posterior probability 0.414 occurs at $J=\{1\}$, indicating the -1.40 is the outlier. The posterior probabilities that -1.40 and 1.01 are the outliers and that there is no outliers are not negligible but those of other cases are very small. The marginal posterior prob-

Table 4. Prior and posterior probabilities of some combinations of possible outliers (Herndon's data)

Number of outliers	Possible outliers	Prior probability	Posterior probability	w
1	−1.40	0.0111	0.414	1.0
2	−1.40, 1.01	0.0016	0.223	1.0
0	none	0.1667	0.152	1.0
3	−1.40, 0.63, 1.01	3.7×10^{-4}	0.035	1.553
2	−1.40, −0.44	0.0016	0.019	1.175
1	1.01	0.0111	0.018	1.0
2	−1.40, 0.63	0.0016	0.013	1.003
3	−1.40, −0.44, 1.01	3.7×10^{-4}	0.010	1.060
2	−1.40, −0.30	0.0016	0.010	1.126
3	−1.40, 0.48, 1.01	3.7×10^{-4}	0.009	1.388

abilities $\pi(i|x)$ that the observation x_i is an outlier are respectively given by $\pi(1|x)=0.802$, $\pi(15|x)=0.326$, $\pi(14|x)=0.064$, $\pi(2|x)=0.040$, $\pi(13|x)=0.027$ and $\pi(i|x)<0.021$ for $i=3, 4, \cdots, 12$.

6. Discussion

In Section 2, we evaluated the average increase of the maximum log likelihood due to the increase of the number of parameters under the simplifying assumption that the condition $x_{i_1} \leqq x_{i_2} \leqq \cdots \leqq x_{i_k}$ always holds. Now we will present a result of our empirical study on the effect of this assumption.

The following two situations were considered:

(1) $k=2$, $J=\{1, 2\}$

$$f(x|J, \mu_0, \mu_j, \sigma)=\sigma^{-n}\phi\left(\frac{x_1-\mu_1}{\sigma}\right)\phi\left(\frac{x_2-\mu_2}{\sigma}\right)\prod_{i=3}^{n}\phi\left(\frac{x_i-\mu_0}{\sigma}\right)$$

with $a=\mu_2-\mu_1 \geqq 0$, $\mu_0=0$ and $\sigma=1$.

(2) $k=3$, $J=\{1, 2, 3\}$

$$f(x|J, \mu_0, \mu_j, \sigma)=\sigma^{-n}\phi\left(\frac{x_1-\mu_1}{\sigma}\right)\phi\left(\frac{x_2-\mu_2}{\sigma}\right)\phi\left(\frac{x_3-\mu_3}{\sigma}\right)\prod_{i=4}^{n}\phi\left(\frac{x_i-\mu_0}{\sigma}\right)$$

with $a=\mu_2-\mu_1=\mu_3-\mu_2 \geqq 0$, $\mu_0=0$ and $\sigma=1$.

In Fig. 1 each dot shows the sample means of $C^*_{k,n}=M^{-1}\sum_x\{\log f(x|J,$

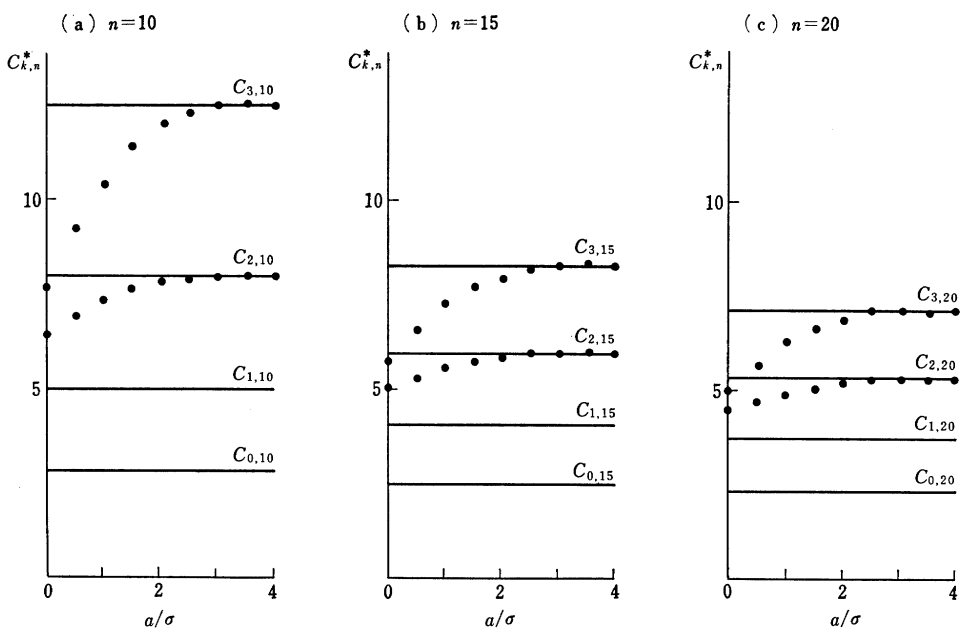

Fig. 1. Sample means of the average increase of the maximum log likelihood due to the increase of the number of parameters

$\hat{\mu}_0, \hat{\mu}_j, \hat{\sigma})-\mathrm{E}_y \log f(y|J, \hat{\mu}_0, \hat{\mu}_j, \hat{\sigma})\}$ obtained from $M=40000$ simulations for each combination of $k=2$ and 3, $n=10$, 15 and 20 and $a=0.0, 0.5, 1.0, \cdots, 4.0$. The solid lines show $C_{k,n}$ for $k=0$, 1, 2 and 3.

From the figure we can see that if $a/\sigma=\mu_i-\mu_{i-1}\geqq 2$, the difference of $C^*_{k,n}$ and $C_{k,n}$ is negligible compared with that of $C^*_{k,n}$ and $C^*_{k+1,n}$. On the other hand, if a/σ is much smaller than 2 the difference of $C^*_{k,n}$ and $C_{k,n}$ becomes considerable. It shows that if some of the outliers are located closely each other then the posterior probability that they are simultaneously outliers is underestimated. This indicates that our procedure may not work well for the detection of a cluster of several outliers, particularly when they are located close to the main distribution. For the detection of these outliers, we should use a modified model which allows the situation where some of the outliers are from a common distribution.

There are some practical ways of simplifying the algorithm to obtain reasonable approximations to the posterior probabilities. In an actual computation, we first rearrange the data in order of increasing magnitude as $x_{(1)}\leqq\cdots\leqq x_{(n)}$ and specify the maximum K of the number of possible outliers, so that, if $K<i\leqq n-K$, the posterior probability that $x_{(i)}$ is an outlier is very small. Then we set $\pi(J|x)=0$ if $\{x_i;\ i\in J\}$ contains any of $x_{(i)}$ with $K<i\leqq n-K$. With this approximation, the number of combinations of possible outliers is reduced to 2^{2K} from 2^n. This will avoid the combinatorial explosion of the number of models.

Generally $w(m)$ $(=\sum w(j, m))$ takes a value between 1 and $k!$. But we can see from Tables 2 and 4 that the values of $w(m)$ of models with significant posterior probabilities are close to 1. Thus we may put $w=1$ to get a reasonable approximation. With this approximation the algorithm can be simplified greatly. The value of $w(m)$ takes a significant value only when there are outliers with nearly equal values.

It is easy to extend the present quasi Bayesian approach to the detection of outliers of large variance type. The model in this case is specified by the conditional data distribution

$$p(x|J, \mu, \sigma, \tau)=\prod_{i\notin J}\frac{1}{\sigma}\phi\left(\frac{x_i-\mu}{\sigma}\right)\prod_{j\in J}\frac{1}{\tau}\phi\left(\frac{x_i-\mu}{\tau}\right),$$

and the prior probability

$$\pi(J)=\frac{(n-k)!k!}{(n+1)!}.$$

By our experience, however, the procedure based on this model is generally insensitive to outliers compared with the procedure discussed in this paper.

The advantage of the present quasi Bayesian approach to the out-

lier problem is that it allows a natural definition of the outlier correction. When an observation x_i is considered to be an outlier a natural correction will be to replace it with the sample mean of the normal observations. Since the posterior probability of the model specified by J is given by $\pi(J|x)$, the final corrected values are obtained by $z_i=\sum_J \pi(J|x)x_i(J)$ $(i=1,\cdots,n)$, where $x_i(J)$ is the corrected value of x_i under the assumption of the model specified by J. This will find a wide application in the area of automatic extreme value correction. Computer program based on this idea is already given in Kitagawa [14].

Acknowledgement

The work reported in this paper was partly supported by a grant from the Asahi Shimbun Publishing Company.

THE INSTITUTE OF STATISTICAL MATHEMATICS

REFERENCES

[1] Abraham, B. and Box, G. E. P. (1978). Linear models and spurious observations, *Appl. Statist.*, **27**, 131-138.

[2] Akaike, H. (1973). Information theory and an extension of the maximum likelihood principle, *2nd Int. Symp. on Information Theory* (eds. B. N. Petrov and F. Csaki), Akademiai Kiado, Bdapest, 267-281.

[3] Akaike, H. (1977). On entropy maximization principle, *Applications of Statistics* (ed. P. R. Krishnaiah), North-Holland, 27-41.

[4] Akaike, H. (1980). On the use of the predictive likelihood of a Gaussian model, *Ann. Inst. Statist. Math.*, **32**, A, 311-324.

[5] Barlow, R. E., Bartholomew, D. J., Bremner, J. M. and Brunk, H. D. (1972). *Statistical Inference under Order Restrictions*, John Wiley and Sons.

[6] Barnett, V. (1978). The study of outliers: Purpose and model, *Appl. Statist.*, **27**, 242-250.

[7] Box, G. E. P. and Tiao, G. G. (1968). A Bayesian approach to some outlier problems, *Biometrika*, **55**, 119-129.

[8] Freeman, F. R. (1979). On the number of outliers in data from a linear model, Int. Meeting on Bayesian Statistic, May 28-June 2, Valencia, Spain, to be published in a special issue of *Trab. Estadist.*

[9] Grubbs, F. E. (1969). Procedures for detecting outlying observations in samples, *Technometrics*, **11**, 1-21.

[10] Guttman, I. (1973). Care and handling of univariate or multivariate outliers in detecting spuriosity—A Bayesian approach, *Technometrics*, **15**, 723-738.

[11] Guttman, I., Dutter, R. and Freeman, P. R. (1978). Care and handling of univariate outliers in the general linear model to detect spuriosity—A Bayesian approach, *Technometrics*, **20**, 187-193.

[12] Hawkins, D. M. (1980). *Identification of Outliers*, Chapman and Hall.

[13] Kitagawa, G. (1979). On the use of AIC for the detection of outliers, *Technometrics*, **21**, 193-199.

[14] Kitagawa, G. (1980). OUTLAP, an outlier analysis program, *Computer Science Monographs*, No. 15, Institute of Statistical Mathematics, Tokyo.

[15] Tiejen, G. L. and Moore, R. H. (1972). Some Grubbs-type statistics for the detection of several outliers, *Technometrics*, **14**, 583-597.