A NOTE ON OPTIMAL ESTIMATION IN THE PRESENCE OF OUTLIERS

John N. Haddad

Department of Mathematics and Statistics, Notre Dame University - Louaize, Zouk Mosbeh, Lebanon.

(Received 26 September 2016 - Accepted 30 November 2016)

AMS Subject Classification: Primary 62M10

ABSTRACT

Haddad, J. 2017. A Note on optimal estimation in the presence of outliers. Lebanese Science Journal, 18(1): 136-141.

The basic estimation problem of the mean and standard deviation of a random normal process in the presence of an outlying observation is considered. The value of the outlier is taken as a constraint imposed on the maximization problem of the log likelihood. It is shown that the optimal solution of the maximization problem exists and expressions for the estimates are given. Applications to estimation in the presence of outliers and outlier detection are discussed and illustrated through a simulation study and analysis of trade data.

Keywords: Optimal estimation, Non Linear programming, Outliers.

INTRODUCTION

Quality Control charts are used to detect sudden and subtle changes that may occur in a purely measurement process. That is the model:

$$X_t = \mu_t + \varepsilon_t \ \ if \ \ t = 1 \ldots n \qquad (1)$$

where $\mu_t s$ are assumed to have the same value μ except for few values of t. Random errors ε_t are necessarily independent with zero mean and a constant variance σ^2. Additional assumptions about the distribution of the errors may be added later.

The common treatment for the above problem in (1) is done through the 3-sigma method, where observations are compared to lie outside the interval $(\mu - 3\sigma/\sqrt{n}, \mu + 3\sigma/\sqrt{n})$ as given in Hogg et al (2005). A further improvement is to use

http://dx.doi.org/10.22453/LSJ-018.1.136141

Cumulative Sums known as CUSUM charts as discussed in Page (1961) and in Taylor (2000).

Other methods that may be utilized for making inference for model (1) is to use robust techniques that are resistant to the presence of observations that have departed from the underlying assumptions and then outlying observations are detected where characterization of these observations would be considered.

A constrained likelihood approach is used here to draw optimal estimates for the parameters of the process given in Equation (1). A non linear optimization problem, NLP, is formulated. An NLP optimization is shown to be a useful tool in the context of hypothesis testing where the restriction of the null hypothesis is taken as the constraint. For more details about constrained tests see for instance Abraham and Yatawara (1988).

In this manuscript, the formulation of the NLP is discussed in section (2). Section (3) has a Monte Carlo study to validate the proposed method. Finally section (4) has extensions to outlier detection and then it is applied to Trade data.

THE ESTIMATION PROCESS

For the model given in (1) under the assumption of equal means μ and normally distributed errors, the log likelihood is given by

$$l(\mu,\sigma) \propto -n\ln(\sigma) - \sum_t (X_t - \mu)^2 / 2\sigma^2 \qquad (2)$$

and then the maximum likelihood estimates of μ and σ are the values that maximize $l(\mu,\sigma)$ or equivalently minimize $-l(\mu,\sigma)$. However, the minimization process has to be constrained by the distributional properties. One such constraint is $|(X_t - \mu)/\sigma| < q$, for all $t = 1,\ldots,n$ and some quantile value which is about 3 in the this case. If one observation, say $X_j - \mu = q\sigma$, then the following NLP can be considered where the objective is to minimize the Lagrangian function

$$(n-1)\ln(\sigma) + \sum_{t\neq j} \frac{(X_t - \mu)^2}{2\sigma^2} + \lambda_j [\frac{(X_j - \mu)}{\sigma} - q] \qquad (3)$$

where $\lambda_j \geq 0$. The existence of an optimal solution has been discussed in Winston (2004) and it is the solution of the following set of equations:

$$(n-1)/\sigma - \sum_{t\neq j} (X_t - \mu)^2 / \sigma^3 - [\lambda_j (X_j - \mu)/\sigma^2] = 0 \qquad (4)$$

$$-2\sum_{t\neq j}(X_t-\mu)/\sigma-(\lambda_j/\sigma)=0 \qquad (5)$$

If the Lagrangian λ_j is zero, then the usual estimators $\overline{X}$ and $s^2=\sum_{i=1}^{n}(X_i-\overline{X})^2/n$ are the optimal solution of the equations (4) and (5), respectively. On the other hand, if the binding constraint has a nonzero "shadow Price", then we must have

$$X_j-\mu=q\sigma \qquad (6)$$

Solving equations (6) for μ, and equation (5) for λ_j, and substituting in equation (4) gives $\hat{\sigma}$ as the positive root of the quadratic equation,

$$(n-1)\hat{\sigma}^2-qS_1\hat{\sigma}-S_2=0, \qquad (7)$$

where $S_1=\sum_{t=1}^{n}(X_t-X_j)$ and $S_2=\sum_{t=1}^{n}(X_t-X_j)^2$. Thus the following is the optimal solution:

$$\hat{\sigma}=\frac{qS_1/(n-1)+\sqrt{(qS_1/(n-1))^2+4S_2/(n-1)}}{2} \qquad (8)$$

and

$$\hat{\mu}=X_j-q\hat{\sigma} \qquad (9)$$

The value q is not known apriori, but one may resort to deletion of the j^{th} observation, then an estimate of μ and σ are obtained to compute a value for q. However, there is an alternative method of approximating the value of q; it is estimated through the moments of the two statistics S_1 and S_2, respectively. It can be shown (see appendix) that the following statistic

$$\frac{S_2-S_1^2/(n-1)}{n-2} \qquad (10)$$

is an unbiased estimator for σ^2. Then the value of q is obtained using equation (7) and finally μ is obtained from equation (6). It is worth pointing out here that the estimation process is done in the reverse order as compared to the usual ones. That is σ is estimated first then "residual" q then μ at the end.

A SIMULATED EXAMPLE

To demonstrate the viability of the derived solution of the NLP problem, a simulation study has been carried out where contamination is taken at a known location. A random realization is generated from a standard normal variates of length n. The "middle" observation is contaminated by adding $q\sigma$ to the actual value of the mean. The mean and the standard deviation are computed before and after the contaminants are added. Then the optimal estimates are computed using the expressions that are derived in section 2 and given in the following sequence: $\hat{\sigma}$ is obtained from equation (10), then $\hat{q}$ is determined from equation (7), and finally $\hat{\mu}$ is estimated from equation (6). The following table shows the results of a 10000 runs of a data generated from a standard normal.

TABLE 1

Comparison of the Three Estimates

n	q	No outliers	With outliers	Optimal
10	3	-0.00515, 0.96983	0.29387, 1.33005	0.96757, 3.29787, 0.003047
	5	0.00069, 0.96785	0.50077, 1.83529	0.96424, 5.54277 , 0.02297
50	3	0.00053 , 0.99345	0.06024, 1.07155	0.99319, 3.04585, 0.00704
	5	0.00137, 0.99465	0.10110, 1.21338	0.99451, 5.07507, 0.00521

It is quite noticeable that $\hat{\sigma}$ is almost the same as if no contamination were added. However, the mean value estimate $\hat{\mu}$ is closer to the actual μ and does better than $\overline{X}$. Further the value of q gets closer to the actual value as n increases.

OUTLIER DETECTION

It is worth pointing out here that the estimating equation in (7) is based on the two statistics S_1 and S_2 where they take the values of 0 and σ^2, respectively, when X_j is equal to μ and departs significantly when an outlier is present. Thus if the location of the outlier is known, $\hat{\sigma}$ is stable and recovers more or less the value of σ as demonstrated in the previous section (3). On the other hand if the outlier location is unknown then one would compute $\hat{\sigma}(j)$, for $j = 1, \cdots, n$. Hence the observation with the smallest $\hat{\sigma}$ will be declared as the location of an outlying observation.

To apply the proposed procedure, the Trade data given in Toylor (2000) is used. The data set has 24 values The following table reports the values of $\hat{\sigma}(j)$:

TABLE 2

Estimate of $\sigma(j)$ of the Trade Data.

j	1	2	3	4	5	6	7	8
$\hat{\sigma}(j)$	1.963	1.956	1.963	1.963	1.961	1.923	1.881	1.922
j	9	10	11	12	13	14	15	16
$\hat{\sigma}(j)$	1.961	1.771	1.963	1.962	1.959	1.963	1.875	1.950
j	17	18	19	20	21	22	23	24
$\hat{\sigma}(j)$	1.882	1.963	1.962	1.963	1.942	1.965	1.962	1.962

It is quite clear to conclude that the 10^{th} observation is a potential outlying observation. This is the same observation that was identified by Taylor's (2000) CUSUM method.

APPENDIX

For a fixed value X_j, $X_t - X_j : N(\mu - X_j, \sigma^2)$,

Now since $S_1 = \sum_t (X_t - X_j)$ then

$$E(S_1) = (n-1)(\mu - X_j) \tag{11}$$

and

$$V(S_1) = (n-1)\sigma^2 \tag{12}$$

Moreover for $S_2 = \sum_t (X_t - X_j)^2$, then

$$\begin{aligned} E(S_2) &= \sum_t (V(X_t - X_j) + (E(X_t - X_j))^2) \\ &= (n-1)\sigma^2 + (n-1)(\mu - X_j)^2 \end{aligned} \tag{13}$$

Therefore $E(S_2 - S_1^2/(n-1)) = (n-2)\sigma^2$ and the result of equation (10) follows.

REFERENCES

Abraham and Yatawara. 1988. A Score Test for Detection of Time Series Outliers. *Journal of Time Series Analysis*, 9: 109-119.

Hogg, McKean and Craig. 2005. Introduction to Mathematical Statistics. Pearson, 6^{th} Edition, New York.

Page, E.S. 1961. Cumulative Sum Charts. *Technometr ics*, 3: 1-9.

Taylor, W. 2000. Change-Point Analysis: A Powerful New Tool for Detecting Changes.

Winston, W. 2004. Operations Research, Applications and Algorithms. Fourth Edition, Thomson, Brooks/Coles Belmot, CA.