ELSEVIER

Statistics & Probability Letters 38 (1998) 255–261

STATISTICS & PROBABILITY LETTERS

Asymptotic variance of M-estimators for dependent Gaussian random variables [1]

Marc G. Genton *

Department of Mathematics, Massachusetts Institute of Technology, Cambridge, MA 02139-4307, USA

Received 1 March 1997; received in revised form 1 August 1997

Abstract

This paper discusses the asymptotic behavior of M-estimators for dependent Gaussian random variables. We show that for a Gaussian distribution, the asymptotic variance of an M-estimator of scale is minimal in the independent case and must necessarily increase for dependent data. This is not true for location estimation where the asymptotic variance can increase or decrease for dependent observations, depending on the sign of the correlation. Several examples are analyzed, showing that the asymptotic variance of the maximum likelihood estimator varies widely under dependencies.

Keywords: Robustness; M-estimator; Location; Scale; Asymptotic variance; Dependent data; Asymptotic efficiency

1. Introduction

The term M-estimator denotes a broad class of estimators of maximum likelihood type, which play an important role in robust statistics. At first introduced for location estimation by Huber (1964), who studied their robustness properties by means of a minimax theorem for the asymptotic variance, they have since been extended to many other situations. The most important ones are scale estimation, regression models and tests. The theory of M-estimators for independently distributed observations is for the most part known. Some of their properties are discussed in Huber (1981), Hampel et al. (1986), Genton and Rousseeuw (1995). The case of dependent data received less attention. It seems that pioneers in this field were Gastwirth and Rubin (1975) with a paper investigating the effect of serial dependence in the data on the efficiency of some robust location estimators. This theme was followed up by Portnoy (1977, 1979), who studied approximately optimal estimators, in the asymptotic minimax sense of Huber (1964, 1972, 1981), in dependent situations. Most of the results are for the location problem, whereas in this paper, we examine the case of scale estimation, which produces interesting results for the Gaussian situation.

* E-mail: genton@math.mit.edu.

[1] This research was supported by the Swiss National Science Foundation.

PII S0167-7152(98)00026-1

Let $X_1,\dots,X_n$ be identically distributed observations according to a parametric distribution F_θ. An M-estimator $T_n(X_1,\dots,X_n)$ of the parameter θ is defined by the implicit equation

$$\frac{1}{n}\sum_{i=1}^{n}\eta(X_i,T_n)=0, \tag{1}$$

and corresponds asymptotically to the statistical functional $T(F)$ defined by

$$\int \eta(x,T(F))\mathrm{d}F(x)=0, \tag{2}$$

where η is a real and sufficiently regular function.

Under regularity conditions, T_n is consistent, i.e. $T_n \to T(F)$ in probability as $n\to\infty$. Moreover, $\sqrt{n}(T_n - T(F))$ is asymptotically normal with zero expectation and variance $V^*(T,F,F^{(k)})$, given by (Portnoy, 1977)

$$V^*(T,F,F^{(k)})=\frac{A(\eta,F)+2\sum_{k=1}^{\infty}A^*(\eta,F^{(k)})}{B^2(\eta,F)}T^2(F), \tag{3}$$

where

$$A(\eta,F)=\int \eta^2(x/T(F))\mathrm{d}F(x), \tag{4}$$

$$B(\eta,F)=\int (x/T(F))\eta'(x/T(F))\mathrm{d}F(x), \tag{5}$$

$$A^*(\eta,F^{(k)})=\int\int \eta(x_1/T(F))\eta(x_2/T(F))\mathrm{d}F^{(k)}(x_1,x_2), \tag{6}$$

and $F^{(k)}$ is the bivariate distribution of the pair (X_1,X_{1+k}). Regularity conditions for consistency and asymptotic normality are given by Huber (1967) for the independent case and by Portnoy (1977, 1979) and Bustos (1982) for the dependent case. In this latter situation, mixing conditions like α-mixing or ϕ-mixing are sufficient (Billingsley, 1968; Doukhan, 1994).

In the framework of the location model $F_\theta(x)=F(x-\theta)$, $\theta\in\mathbb{R}$, it is natural to use $\eta(x,T(F))=\psi(x-T(F))$ in Eqs. (1) and (2), where ψ is odd. Some examples of M-estimators of location are given in Section 3.

2. Scale estimation

The scale model is given by $F_\sigma(x)=F(x/\sigma)$, $\sigma>0$. In this context, it is natural to use the function $\eta(x,S(F))=\chi(x/S(F))$ in Eqs. (1) and (2), where χ is even. Some examples of M-estimators of scale are given in Section 3. The following lemma and theorem show that for dependent Gaussian random variables, the asymptotic variance of an M-estimator of scale is necessarily greater than for independent ones. This is true even if the Gaussian random variables are negatively correlated. Let us denote by Φ the standard Gaussian distribution and by Φ_τ, $-1<\tau<1$, the Gaussian bivariate distribution with variance matrix

$$\Sigma=\begin{pmatrix}1 & \tau\\ \tau & 1\end{pmatrix}. \tag{7}$$

Its density function is

$$\phi(x_1,x_2)=\frac{1}{2\pi\sqrt{1-\tau^2}}\exp\left(-\frac{1}{2}(x_1\,x_2)\Sigma^{-1}(x_1\,x_2)^{\mathrm{T}}\right). \tag{8}$$

Note that further restriction on τ may be necessary in order to insure positive definiteness of the complete variance-covariance matrix of the sample $X_1,\dots,X_n$.

Lemma 1. *Let X_1 and X_2 be two dependent random variables having a Gaussian bivariate distribution Φ_τ, $-1<\tau<1$. For any even χ, the inequality*

$$A^*(\chi,\Phi_\tau)\geqslant 0,$$

holds, with equality if and only if $\tau=0$.

Proof. The proof is based on an inequality of Dudley (1973), generalized by Gutmann (1978). If $\tau=0$, then the bivariate density function can be written as $\phi(x_1,x_2)=\phi(x_1)\phi(x_2)$, where ϕ is the standard density function. This yields $A^*(\chi,\Phi_0)=0$. If the covariance $\mathrm{Cov}(X_1,X_2)=\tau>0$, then Gutmann (1978) has shown that

$$\mathrm{Cov}(h(X_1),h(X_2))>0, \tag{9}$$

for any real function h such that $0<\mathrm{Var}(h(X_1))<\infty$. This is the case for our function χ. If the covariance $\mathrm{Cov}(X_1,X_2)=\tau<0$, then $\mathrm{Cov}(-X_1,X_2)=-\tau>0$, and we have $\mathrm{Cov}(\chi(-X_1),\chi(X_2))=\mathrm{Cov}(\chi(X_1),\chi(X_2))>0$, as the function χ is symmetric. This proves the desired inequality. □

Theorem 1. *For every symmetric M-estimator of scale based on dependent data, with marginal distribution $F=\Phi$ and bivariate distributions $F^{(k)}=\Phi_{\tau_k}$, $k\geqslant 1$, the inequality*

$$V^*(S,\Phi,\Phi_{\tau_k})\geqslant V(S,\Phi),$$

holds, with equality if and only if $\tau_k=0$, $\forall k\geqslant 1$.

Proof. Using the previous lemma, we have $\sum_{k=1}^{\infty}A^*(\chi,\Phi_{\tau_k})\geqslant 0$, and hence

$$V^*(S,\Phi,\Phi_{\tau_k})=\frac{A(\chi,\Phi)+2\sum_{k=1}^{\infty}A^*(\chi,\Phi_{\tau_k})}{B^2(\chi,\Phi)}\geqslant\frac{A(\chi,\Phi)}{B^2(\chi,\Phi)}=V(S,\Phi). \qquad \square$$

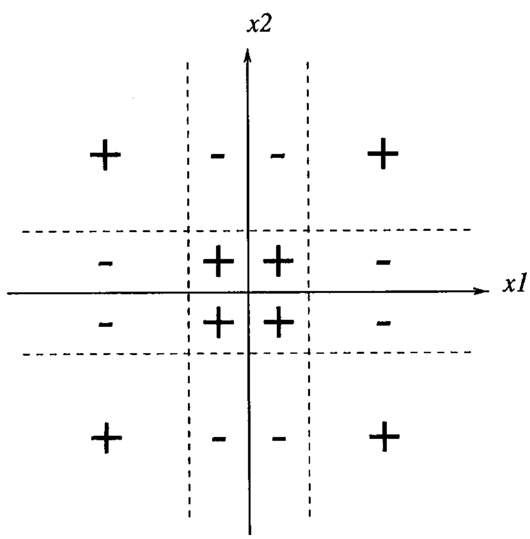

Fig. 1. The sign of the function $\chi(x_1)\chi(x_2)$. The negative parts are represented by the symbol $-$ and the positive ones by the symbol $+$.

Note that this result is not necessarily true if the underlying distribution is not Gaussian and if the observations are negatively correlated. Effectively, the proof of Theorem 1 is based on the inequality (9), which becomes tricky under non-Gaussian distributions. In fact, as χ is an even function, the sign of the product $\chi(x_1)\chi(x_2)$ takes the structure presented in Fig. 1. Thus, one can construct a bivariate distribution which does not satisfy the inequality (9). For example, a distribution with heavy weights along the axes (in the negative parts of $\chi(x_1)\chi(x_2)$), or a multi-modal distribution with modes in the negative parts of $\chi(x_1)\chi(x_2)$, will produce a negative covariance between the random variables $\chi(X_1)$ and $\chi(X_2)$.

3. Examples

In this section, we analyze the behavior of the asymptotic variance of some typical location and scale estimators. First, we consider the simple case where $\tau_1=\tau$ and $\tau_k=0$, $\forall k\geqslant 2$. This corresponds to an MA(1) dependence structure (Brockwell and Davis, 1987), and needs the additional constraint $-\frac{1}{2}<\tau<\frac{1}{2}$.

3.1. The maximum likelihood estimator

In the location model, the maximum likelihood estimator (MLE) at $F=\Phi$ is defined by the function $\psi(x)=x$, and corresponds to the arithmetic mean. Straightforward computation yields $A^*(\psi,\Phi_\tau)=\tau$ and $V^*(T,\Phi,\Phi_\tau)=1+2\tau$. Note in this case that the asymptotic variance may decrease if the correlation is negative.

In the scale model, the maximum likelihood estimator (MLE) at $F=\Phi$ is defined by the function $\chi(x)=x^2-1$, and corresponds to the classical standard deviation. This yields $A^*(\chi,\Phi_\tau)=2\tau^2$ and $V^*(S,\Phi,\Phi_\tau)=\frac{1}{2}+\tau^2$. Note in this case that the asymptotic variance may not decrease if the correlation is negative.

3.2. The median estimator

In the location model, the median estimator at $F=\Phi$ is defined by the function $\psi(x)=\mathrm{sign}(x)$. This yields

$$A^*(\psi,\Phi_\tau)=\frac{2}{\pi}\arctan\left(\frac{\tau}{\sqrt{1-\tau^2}}\right),$$

$$V^*(T,\Phi,\Phi_\tau)=\frac{\pi}{2}+2\arctan\left(\frac{\tau}{\sqrt{1-\tau^2}}\right).$$

3.3. The median absolute deviation estimator

In the scale model, the median absolute deviation estimator (MAD) at $F=\Phi$ is defined by the function $\chi(x)=\mathrm{sign}(|x|-q)$, where $q=\Phi^{-1}(3/4)$. This yields

$$A^*(\chi,\Phi_\tau)=8\int_{-q}^{q}\Phi\left(\frac{q-\tau x}{\sqrt{1-\tau^2}}\right)\mathrm{d}\Phi(x)-3,$$

and

$$V^*(S,\Phi,\Phi_\tau)=\frac{1}{q^2\phi^2(q)}\left[\int_{-q}^{q}\Phi\left(\frac{q-\tau x}{\sqrt{1-\tau^2}}\right)\mathrm{d}\Phi(x)-\frac{5}{16}\right],$$

where ϕ is the standard Gaussian density function.

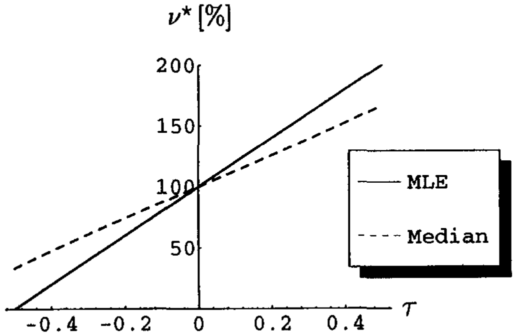

Fig. 2. The asymptotic efficiency relative to independence $\nu^*(\Phi_\tau)$ in the location model and MA(1) dependence structure for the MLE and the median.

3.4. The Welsch estimator

In the scale model, the generalized Welsch estimator at $F = \Phi$ is defined by the function $\chi(x) = \sqrt{\frac{d}{d+2}} - \exp(-x^2/d)$, with $d > 0$. This yields

$$A^*(\chi, \Phi_\tau) = \frac{d}{\sqrt{4(1-\tau^2) + d(d+4)}} - \frac{d}{d+2}.$$

The classical Welsch estimator corresponds to $d = 2/3$ and yields

$$A^*(\chi, \Phi_\tau) = \frac{1}{\sqrt{16 - 9\tau^2}} - \frac{1}{4},$$

$$V^*(S, \Phi, \Phi_\tau) = \frac{128}{9\sqrt{16 - 9\tau^2}} + \frac{64}{9\sqrt{7}} - \frac{16}{3}.$$

3.5. The effects of dependencies

In order to analyze the effects of the dependencies on the asymptotic variance, we define the *asymptotic efficiency relative to independence* $\nu^*(\Phi_\tau)$ of an M-estimator S at Φ_τ as being the ratio

$$\nu^*(\Phi_\tau) = \frac{V^*(S, \Phi, \Phi_\tau)}{V(S, \Phi)}.$$

Note that in the case of M-estimators of scale, this quantity is always greater than 1.

Fig. 2 shows the asymptotic efficiency relative to independence $\nu^*(\Phi_\tau)$ in the location model for the MLE and the median. The asymptotic efficiency relative to independence of the maximum likelihood estimator depends quite strongly on τ. Moreover, the asymptotic variance can increase or decrease for dependent observations, depending on the sign of the correlation. Note that the asymptotic variance of the MLE could theoretically be reduced to zero by letting $\tau \to -\frac{1}{2}$, whereas the asymptotic variance of the median could not, because

$$\lim_{\tau \to -1/2} \left[\frac{\pi}{2} + 2 \arctan\left(\frac{\tau}{\sqrt{1-\tau^2}} \right) \right] = \frac{\pi}{6}.$$

The asymptotic efficiency relative to independence $\nu^*(\Phi_\tau)$ in the scale model is shown in Fig. 3 for the MLE, the Welsch estimator and the MAD. Here again, we notice that the asymptotic efficiency relative to

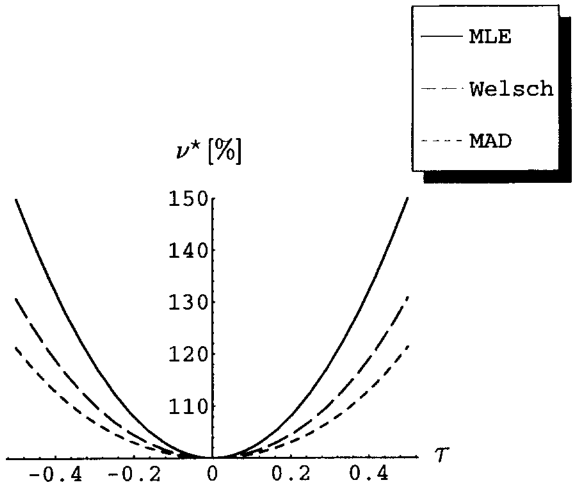

Fig. 3. The asymptotic efficiency relative to independence $\nu^*(\Phi_\tau)$ in the scale model and MA(1) dependence structure for the MLE, the Welsch estimator and the MAD.

independence of the maximum likelihood estimator varies quite a lot. However, the asymptotic variance must necessarily increase for dependent observations, even if the data is negatively correlated.

3.6. *The* AR(1) *dependence structure*

The case where $\tau_k = \theta^k$, $\forall k \geqslant 1$ corresponds to an AR(1) dependence structure (Brockwell and Davis, 1987) with $-1 < \theta < 1$. No additional constraint is needed for the correlation since $-1 < \tau_k < 1$, $\forall k \geqslant 1$. The maximum likelihood estimator (MLE) allows explicit computations of the infinite sums in equation (3) by means of elementary results on geometric series. In the location model, the MLE at $F = \Phi$ yields $V^*(T, \Phi, \Phi_\theta) = 1 + 2\frac{\theta}{1-\theta}$ for the asymptotic variance, which may decrease if the correlation is negative. In the scale model, the MLE at $F = \Phi$ yields $V^*(S, \Phi, \Phi_\theta) = \frac{1}{2} + \frac{\theta^2}{1-\theta^2}$ for the asymptotic variance, which may not decrease if the correlation is negative. Both models (location and scale) show larger variations of the asymptotic variance for the AR(1) dependence structure than for the MA(1).

4. Conclusion

In this paper, the asymptotic variance of M-estimators for dependent Gaussian random variables has been studied. We showed that for a Gaussian distribution, the asymptotic variance of an M-estimator of scale is minimal in the independent case and must necessarily increase for dependent data. However, this is not true for location estimation where the asymptotic variance can increase or decrease for dependent observations, depending on the sign of the correlation. The asymptotic variance under dependencies has been computed for several examples of M-estimators, showing that the asymptotic variance of the maximum likelihood estimator suffers quite a large variation due to dependencies.

Acknowledgements

This work contains parts of my Ph.D. dissertation, which was written under the generous guidance of Prof. Stephan Morgenthaler at the Swiss Federal Institute of Technology, Lausanne, Switzerland.

References

Billingsley, P., 1968. Convergence of Probability Measures. Wiley, New York.

Brockwell, P.J., Davis, R.A., 1987. Time series: Theory and Methods. Springer, Berlin.

Bustos, O.H., 1982. General M-estimates for contaminated pth order autoregressive processes: consistency and asymptotic normality. Zeitschrift für Wahrscheinlichkeitstheorie und verwandte Gebiete. 59, 491–504.

Doukhan, P., 1994. Mixing: Properties and Examples. Springer, Berlin.

Dudley, R., 1973. Sample functions of the Gaussian process. Ann. Probab. 1, 66–103.

Gastwirth, J.L., Rubin, H., 1975. The behavior of robust estimators on dependent data. Ann. Statist. 3, 1070–1100.

Genton, M.G., Rousseeuw, P.J., 1995. The change-of-variance function of M-estimators of scale under general contamination. J. Comput. Appl. Math. 64, 69–80.

Gutmann, S., 1978. Correlations of functions of normal variables. J. Multivariate Anal. 8, 575–578.

Hampel, F.R., Ronchetti, E.M., Rousseeuw, P.J., Stahel, W.A., 1986. Robust Statistics: The Approach based on Influence Functions. Wiley, New York.

Huber, P.J., 1964. Robust estimation of a location parameter. Ann. Math. Statist. 35, 73–101.

Huber, P.J., 1967. The behavior of maximum likelihood estimates under non-standard conditions. Proc. 5th Berkeley Symp. on Mathematical Statistics and Probability, vol. 1, pp. 221–233.

Huber, P.J., 1972. Robust statistics: a review. Ann. Math. Statist. 43, 1041–1067.

Huber, P.J., 1981. Robust Statistics. Wiley, New York.

Portnoy, S., 1977. Robust estimation in dependent situations. Ann. Statist. 5, 22–43.

Portnoy, S., 1979. Further remarks on robust estimation in dependent situations. Ann. Statist. 7, 224–231.