Ann. Inst. Statist. Math.
Vol. 40, No. 4, 799–820 (1988)

ROBUST M-ESTIMATORS IN DIFFUSION PROCESSES

NAKAHIRO YOSHIDA

Department of Applied Mathematics, Faculty of Engineering Science, Osaka University, Toyonaka, Osaka 560, Japan

(Received September 24, 1987; revised April 25, 1988)

Abstract. Methods of robust estimation in diffusion processes are given by means of M-estimation. It is shown that the asymptotic variance of an M-estimator is obtained by applying a certain integral operator to the influence function and integrating its square. Under the condition of boundedness of the influence function, the existence of an optimal robust M-estimator is shown and an approximately optimal practical method is given. Moreover, as another criterion of robustness we consider a norm of integral type and show that the corresponding optimal robust M-estimator is obtained by solving a boundary value problem of a second order differential equation. Finally, as an illustrative example the Ornstein-Uhlenbeck process is discussed.

Key words and phrases: Diffusion process, M-estimator, influence function, Sobolev space, second order differential equation.

1. Introduction

The purpose of the present paper is to give methods of robust estimation in diffusion processes realized by stochastic differential equations discussed in Section 2. For that purpose, we consider the method of M-estimation, i.e., the estimators are defined as solutions of some estimating equations. Their asymptotic properties are given in Section 3.

Contaminations of stationary distributions of diffusion processes make M-estimators biased. In Section 4, we calculate the influence function, the measure of influence of infinitesimal contamination. Then it is interesting to see that the asymptotic variance of the M-estimator is not the square integral of its influence function but that of its image by some integral operator. For that reason it does not seem to have an explicit Huber-type representation as in Künsch (1984). Theoretically, we show the existence of an optimal robust M-estimator under the condition of uniform boundedness of influence functions. We propose a practical method which gives an approximately optimal M-estimator in Section 5.

On the other hand, we can consider other optimality problems by using different criteria for robustness. In Section 6, we adopt an integral-type norm as our measure of influence functions. It is shown that the optimal robust M-estimator is given as a solution of a boundary value problem of a second order differential equation. Finally, an illustrative example of this method is given.

2. Model and maximum likelihood estimator

We treat the following stochastic differential equation:

$$\begin{aligned} dX(t) &= \sigma(X(t))dW(t) + f(X(t),\theta)dt\,, \\ X(0) &= x(0)\,, \end{aligned} \tag{2.1}$$

where θ is a parameter in Θ, an interval in $R, f(x,\theta)$ and $\sigma(x)$ are functions of class C^1, $\sigma(x) > 0$ and $W(t)$ is a standard Wiener process. A sufficient condition for the existence and uniqueness of a (strong) solution of the stochastic differential equation is Lipschitz continuity and linear growth of the coefficients $f(\cdot,\theta)$ and $\sigma(\cdot,\theta)$ (see e.g., Ikeda and Watanabe (1981)). We assume that our diffusion process is a solution of the equation (2.1).

Moreover, the diffusion processes with ergodicity are treated here. Let the boundaries $\pm\infty$ of the phase space be inaccessible,

$$B(x,\theta) = \int_0^x 2\sigma^{-2}(y)f(y,\theta)dy\,, \tag{2.2}$$

$$m(x,\theta) = \int_0^x 2\sigma^{-2}(y)\exp B(y,\theta)dy\,. \tag{2.3}$$

Then if $\tilde{m} = m(\infty,\theta) - m(-\infty,\theta) < \infty$, the diffusion process defined by (2.1) has the ergodic property and its stationary distribution is given by $\nu(x,\theta) = [m(x,\theta) - m(-\infty,\theta)]/\tilde{m}$ with density $\mu(x,\theta) = \nu(dx,\theta)/dx$ (see Itô and McKean (1965), Mandl (1968) and Gihman and Skorohod (1972)).

The log-likelihood ratio based on the observation $(X(t); 0 \le t \le T)$ is given by the formula

$$\begin{aligned} \Lambda(T,\theta) = &\int_0^T f(X(t),\theta)\sigma^{-2}(X(t))dX(t) \\ &- \frac{1}{2}\int_0^T f(X(t),\theta)^2\sigma^{-2}(X(t))dt\,, \end{aligned} \tag{2.4}$$

(e.g., Liptser and Shiryayev (1977)). If $H(x,\theta) := B(x,\theta)/2$, by Itô's formula, as in Lánska (1979) we have

$$\Lambda(T,\theta) = H(X(T),\theta) - H(X(0),\theta) - \int_0^T h(X(t),\theta)dt\,, \tag{2.5}$$

where

$$h(x,\theta) = [f(x,\theta)\sigma(x)^{-1}]^2/2 + \sigma^2(x)\partial/\partial x[f(x,\theta)\sigma(x)^{-2}]/2\,. \tag{2.6}$$

Therefore, under some regularity conditions, the maximum likelihood estimator (MLE) is a solution of the following estimating equation:

$$\dot\Lambda(T,\theta) = \dot H(X(T),\theta) - \dot H(X(0),\theta) - \int_0^T L(\theta)\dot H(X(t),\theta)dt = 0\,, \tag{2.7}$$

where "$\cdot$" stands for derivatives with respect to the parameter, $L(\theta)$ is the generator corresponding to the diffusion process defined by (2.1):

$$L(\theta) = f(x,\theta)D + \frac{1}{2}\sigma^2(x)D^2, \qquad D = \frac{\partial}{\partial x}\,, \tag{2.8}$$

and

$$L(\theta)\dot H(x,\theta) = \dot h = f\dot f\sigma^{-2} + (1/2)\sigma^2 D[\dot f\sigma^{-2}]\,. \tag{2.9}$$

It is well known that $\dot\Lambda(T,\theta)$ given in (2.7) is a martingale when θ is true. We can calculate the MLE from (2.7) in practice. For the consistency, asymptotic normality and efficiency of the MLE's in diffusion processes, refer to Kutoyants (1977, 1978, 1984), Lánska (1979) and Rao and Rubin (1981). McKeague (1984) discusses the asymptotic behaviour of the MLE's in misspecified models.

3. M-estimation and asymptotic behaviour

Practically, we can hardly get clean data generated from a model (2.1) in the strict sense, since it is natural to regard it as contaminated by some noises and misspecification of the true models.

Although we often adopt Gaussian models, which are feasible and relatively easily treated theoretically, if the observations are contaminated by additive noises of long-tailed distributions, they do not belong to the class.

Consider the Gaussian system X defined by the following stochastic differential equation:

$$dX(t) = -AX(t)dt + dW(t), \qquad X(0) = x(0)\,,$$

where A is a positive definite matrix. If small independent effects $Z(t)$ with long-tailed distributions are added to $X(t)$:

$$Y(t) = X(t) + Z(t),$$

then the observation $Y(t)$ has long-tailed distribution and not Gaussian.

The second type of contamination is caused by innovation outliers. For example, let X be a 1-dimensional nonlinear stochastic dynamical system with equilibrium point $x = 0$ satisfying the following stochastic differential equation:

$$dX(t) = -\theta X(t)dt + \sigma(X(t))dW(t), \qquad X(0) = x(0),$$

where θ is a positive constant. Assume that we estimate the parameter θ by using the locally linearized model

$$dX(t) = -\theta X(t)dt + b dW(t), \qquad b > 0,$$

either for lack of knowledge about the form of the diffusion coefficient $\sigma(x)$ or for computational convenience. Then the stationary distribution of the locally linearized model is normal $N(0, b^2/2\theta)$. On the other hand, in the case where $\sigma(x) = (b_1^2 + ax^2)^{1/2}$, $a > 0$, the true model has a distribution whose density is proportional to $(ax^2 + b_1^2)^{-(1+\theta/a)}$ and heavy-tailed.

Moreover, there exists another type of contamination when we treat processes, caused by patchy outliers. Martin and Yohai (1986) stress this point; however, we will not discuss it in the present paper.

Contamination of the stationary distribution of the observation process makes the MLE biased. So our question is how to get robust estimators in some sense for contaminated data.

The problem of robust estimation has been studied for i.i.d. models. We can refer to Huber (1981) and Hampel *et al.* (1986). For dependent models, many authors study it in time series (see Denby and Martin (1979), Kleiner *et al.* (1979), Bustos (1982), Künsch (1984), Martin and Yohai (1985, 1986), and Bustos and Yohai (1986)). They use M-estimators, GM-estimators, etc. to construct robust estimators. Künsch (1984) and Martin and Yohai (1986) are based on influence functions and influence functionals, respectively. In this paper, we consider the problem of robust estimation in diffusion models in the way of M-estimation, i.e., the estimator is given as a null point of an estimating equation.

DEFINITION 3.1. For functions $a(x, \theta)$ and $A(x, \theta)$, an estimator $\hat{\theta}(T)$ at which

$$M(T,\theta) := A(X(T),\theta) - A(X(0),\theta) - \int_0^T a(X(t),\theta)dt\,, \tag{3.1}$$

vanishes, is called an M-estimator corresponding to (a, A).

The MLE is an M-estimator with

$$a = L(\theta)\dot{H} = \dot{h} \quad \text{and} \quad A = \dot{H}\,. \tag{3.2}$$

Note that our estimating function (3.1) and the differential of the contrast function defined in Lánska (1979) are the same. The asymptotic behaviour of the M-estimator is described in the same manner, but to construct robust estimators it is convenient to treat a wider class of functions a and A, i.e., Sobolev-type space. Especially, we remove the continuity of $a(x,\theta)$ in x, so we briefly show the asymptotic behaviour of our M-estimators here.

For simplicity, we assume that $\int \sigma^2(x)\mu(x,\theta)dx < \infty$ for all $\theta \in \Theta$ in this paper. Let $F = \{A;\ DA \in L^2(\sigma^2\mu(\cdot,\theta)),\ L(\theta)A \in L^2(\mu(\cdot,\theta)),\ \theta \in \Theta\}$, where the derivatives w.r.t. x are in the sense of Schwartz's distributions. We treat stationary diffusion processes for simplicity and consider the following conditions.

(I) $A(\cdot,\theta) \in F$, $\theta \in \Theta$.

(II-1) For $\theta \in \Theta$ and $\theta' \in \operatorname{Int}\Theta$, $a(\cdot,\theta) \in L^1(\mu(\cdot,\theta'))$.

(II-2) $(\partial/\partial\theta)\int a(x,\theta)\mu(x,\theta')dx|_{\theta=\theta'} = \int \dot{a}(x,\theta')\mu(x,\theta')dx \neq 0$, $\theta' \in \operatorname{Int}\Theta$.

(III) For $\theta' \in \operatorname{Int}\Theta$, $\theta \to M(T,\theta)$ is continuous at θ', $T > 0$, a.s. $P_{\theta'}$.

(IV) $\dot{a}(\cdot,\theta) \in L^1(\mu(\cdot,\theta))$ and for $\theta \in \operatorname{Int}\Theta$. $\dot{M}(T,\theta)/T \to -\int \dot{a}(x,\theta) \cdot\mu(x,\theta)dx$ in P_θ continuously, i.e., even if θ in $\dot{M}$ is replaced by an arbitrary sequence $s(T) \to \theta$, the convergence holds.

(V) $DG(\cdot,\theta) \in L^2(\sigma^2\mu(\cdot,\theta))$, $\theta \in \Theta$, where G is defined by

$$G(x,\theta) = -\int_0^x \exp(-B(y,\theta))dy \cdot \int_y^\infty 2a(u,\theta)\sigma^{-2}(u)\exp(B(u,\theta))du\,. \tag{3.3}$$

Note that $L(\theta)G(\cdot,\theta) = a(\cdot,\theta)$ and $M(T,\theta)$ is a local martingale under P_θ when $A = G$.

Remarks 3.1. (1) Lánska's condition that $\int a(x,\theta)\mu(x,\theta)dx = 0$ for Fisher consistency is an essential one. Our assumption that $\int \sigma^2(x)\mu(x,\theta)dx < \infty$ and (V) lead to the condition (see the proof of Theorem 3.1).

(2) For the following two theorems it is sufficient to take F as

$$\{A; DA \in L^2(\sigma^2\mu(\cdot,\theta)), L(\theta)A \in L^1(\mu(\cdot,\theta)), \theta \in \Theta\}.$$

THEOREM 3.1. *If* (I), (II-1, 2), (III) *and* (V) *hold and* $\theta \in \operatorname{Int} \Theta$ *is true, a consistent M-estimator* $\hat{\theta}(T)$ *exists.*

First, we prepare

LEMMA 3.1. *When* $\theta' \in \operatorname{Int} \Theta$, *for* $\theta \in \Theta$

$$[A(X(T),\theta) - A(X(0),\theta)]/T \to 0,$$

as $T \to \infty$ *a.s.* $(P_{\theta'})$.

PROOF. By the extended version of Itô's formula (Krylov (1980)),

$$\begin{aligned} (3.4)\quad & [A(X(T),\theta) - A(X(0),\theta)]/T \\ &= \frac{1}{T}\int_0^T L(\theta')A(X(t),\theta)dt + \frac{1}{T}\int_0^T DA(X(t),\theta)\sigma(X(t))dW(t). \end{aligned}$$

In order to prove the lemma, it is sufficient to show that the first term in the r.h.s. of (3.4) converges to zero a.s., because the last term converges to zero a.s. by Lepingle's strong law of large numbers for martingales (Lepingle (1978)). Moreover, it is sufficient to show that $\int L(\theta')A(x,\theta) \cdot\mu(x,\theta')dx = 0$, which is its a.s. limit. By (I) and Schwartz's inequality,

$$\begin{aligned} \int \sigma^2(x)|DA(x,\theta)| \cdot \mu(x,\theta')dx &\le \left\{\int \sigma^2(x)\mu(x,\theta')dx\right\}^{1/2} \\ &\quad \cdot \left\{\int |DA(x,\theta)|^2\sigma^2(x)\mu(x,\theta')dx\right\}^{1/2} < \infty. \end{aligned}$$

It is easy to see that for $\tilde{A}(x,\theta) := DA(x,\theta)$

$$\begin{aligned} \tilde{A}(x,\theta) &= \exp(-B(x,\theta')) \\ &\quad \cdot \left[\int_{-\infty}^{x} 2\sigma^{-2}L'(\theta')\tilde{A}(y,\theta) \cdot \exp B(y,\theta')dy + c\right], \end{aligned}$$

where c is a constant and $L'(\theta') = f(\cdot,\theta') + (1/2)\sigma^2(\cdot)D$. If

$$k := \int_{-\infty}^{\infty} 2\sigma^{-2}L'(\theta')\tilde{A}(y,\theta) \cdot \exp B(y,\theta')dy + c \neq 0,$$

$\exp B(x,\theta') \cdot \tilde{A}(x,\theta) \to k$ as $x \to \infty$ and that contradicts the integrability of $\sigma^2(x)\mu(x,\theta')\tilde{A}(x,\theta)$. Therefore, $k=0$. Similarly, $c=0$, considering the case $x \to -\infty$. Hence, $\int L'(\theta')\tilde{A}(x,\theta)\mu(x,\theta')dx = 0$ and the proof is completed.

PROOF OF THEOREM 3.1. For any sufficiently small $\delta > 0$, by Lemma 3.1, $M(T,\theta \pm \delta)/T$ converges to

$$-\int a(x,\theta\pm\delta)\mu(x,\theta)dx$$

as $T \to \infty$ a.s. (P_θ). Set

$$k = -\int_{-\infty}^{\infty} \tilde{m}a(x,\theta)\mu(x,\theta)dx\,.$$

From (3.3),

$$DG(x,\theta) = -\exp(-B(x,\theta))\int_x^{\infty} \tilde{m}a(u,\theta)\mu(u,\theta)du\,.$$

If $k \neq 0$, $\exp B(x,\theta) \cdot DG(x,\theta) \to k$ as $x \to -\infty$ and

$$\int \sigma^2(x)DG(x,\theta)\cdot\mu(x,\theta)dx = \pm\infty\,.$$

This is a contradiction. In fact, by (V),

$$\left|\int \sigma^2(x)DG(x,\theta)\cdot\mu(x,\theta)dx\right| \le \left\{\int \sigma^2(x)\mu(x,\theta)dx\right\}^{1/2} \cdot \left\{\int |DG(x,\theta)|^2\sigma^2(x)\mu(x,\theta)dx\right\}^{1/2} < \infty\,.$$

Hence, $\int a(x,\theta)\mu(x,\theta)dx = 0$. The signs of $M(T,\theta\pm\delta)$ are different for large T by (II-2), so there exists a $\hat{\theta}(T)$ in $(\theta-\delta,\theta+\delta)$ from (III).

THEOREM 3.2. *If* (I)–(V) *hold, when* $\theta \in \operatorname{Int}\Theta$ *is true,* $\sqrt{T}(\hat{\theta}(T)-\theta) \to N(0,\Sigma)$ *in law as* $T \to \infty$, *where* $\Sigma = \Delta/U^2$,

$$\Delta = \int [DG(x,\theta)\sigma(x)]^2\mu(x,\theta)dx\,,$$

$$U = U(\theta) = -\int \dot{a}(x,\theta)\mu(x,\theta)dx\,.$$

PROOF. By Taylor expansion,

$$-T^{-1/2}M(T,\theta)=T^{-1}\dot{M}(T,\bar{\theta})T^{1/2}(\hat{\theta}(T)-\theta)\,,$$

where $\bar{\theta}$ is between $\hat{\theta}(T)$ and θ. With (IV) this implies that

$$UT^{1/2}(\hat{\theta}(T)-\theta)+T^{-1/2}M(T,\theta)\to 0 \tag{3.5}$$

in P_θ. On the other hand,

$$\begin{aligned} T^{-1/2}M(T,\theta)=\ & T^{-1/2}[A(X(T),\theta)-A(X(0),\theta)] \\ & -T^{-1/2}[G(X(T),\theta)-G(X(0),\theta)] \\ & +T^{-1/2}\int_0^T DG(X(t),\theta)\cdot\sigma(X(t))dW(t)\,. \end{aligned} \tag{3.6}$$

The first two terms in the r.h.s. of (3.6) converge to zero in probability because of stationarity, and the third term converges in distribution to $N(0,\Delta)$ by a central limit theorem for martingales, e.g., Feigin (1985). The result follows from (3.5) and (3.6).

4. Influence function and robustness

Let κ be the stationary distribution of $X=(X(t))$. Under some conditions, we have

$$\frac{1}{T}M(T,\theta)\to -\int a(x,\theta)d\kappa \tag{4.1}$$

as $T\to\infty$ a.s. If the value θ^* at which the r.h.s. of (4.1) vanishes is unique, we write $Y(\kappa)$ for this θ^*. Moreover, if the convergence in (4.1) is in the sense of uniform topology on $C(\Theta)$, then the M-estimator $\theta(T)$ converges to $Y(\kappa)$ as $t\to\infty$ a.s. To ensure that, for example, we may consider conditions of uniform boundedness and the equicontinuity of the family $\{M(T,\cdot)/T;\,T\ge 0\}$. For such a and κ,

$$\int a(x,Y(\kappa))d\kappa=0\,. \tag{4.2}$$

Here note that κ does not have to be a member of $\{\nu(\cdot,\theta)\}$. Fisher consistency is written by $Y(\nu(\cdot,\theta))=\theta$, for all θ, and it is assumed for our M-estimators.

Consider the case where the stationary distribution $\nu(\cdot,\theta)$ of $(X(t))$ under P_θ changes into $(1-\varepsilon)\nu(\cdot,\theta)+\varepsilon\kappa$, where κ is a probability distribution on R. For example, let Z_t be a stochastic process such that $X+Z$ is stationary with marginal distribution κ. The process Z is a contamination and κ may be an arbitrary distribution. Let V_t be a $\{0,1\}$-valued Markov process with stationary distribution $P\{V_t=0\}=1-\varepsilon$, $P\{V_t=1\}=\varepsilon$. Suppose that V and (X,Z) are independent. When our contaminated observed process Y is a mixture of the true one and the contaminated one:

$$Y_t=(1-V_t)X_t+V_t(X_t+Z_t)=X_t+V_tZ_t\,,$$

it is easy to show that the marginal distribution of the observed process Y is $(1-\varepsilon)\nu+\varepsilon\kappa$.

From (4.2)

$$\int a(x, Y((1-t)\nu(\cdot,\theta)+t\kappa))d((1-t)\nu(\cdot,\theta)+t\kappa)=0 \tag{4.3}$$

holds. Differentiating with respect to t at $t=0$ we obtain

$$\begin{aligned} Y'(\kappa,a,\nu(\cdot,\theta)) &:= \frac{\partial}{\partial t} Y((1-t)\nu(\cdot,\theta)+t\kappa)|_{t=0} \\ &= U^{-1}\int a(x,\theta)d\kappa\,, \end{aligned} \tag{4.4}$$

where

$$U=U(\theta)=-\int \dot a(x,\theta)\mu(x,\theta)dx\,. \tag{4.5}$$

When $\nu=\delta_x$,

$$Y'(x,a,\nu(\cdot,\theta)) := Y'(\delta_x,a,\nu(\cdot,\theta))=U^{-1}a(x,\theta) \tag{4.6}$$

is called the influence function of the M-estimator at $\nu(\cdot,\theta)$. The influence function is a measure of sensitivity of estimators to changes in the stationary distribution of X.

From Fisher consistency,

$$\int a(x,\theta)\mu(x,\theta)dx=0\,,$$

and by differentiating we get a differential representation of it:

$$-U+\int a(x,\theta)\dot\mu(x,\theta)dx=0\,, \tag{4.7}$$

where $\dot\mu$ is the derivative of $\mu(x,\theta)$ with respect to θ.

The influence function of the MLE is often unbounded and then from (4.4) it is seen that contamination of the stationary distribution of the process, especially for large $|x|$, greatly influences its bias. Recall the second example of Section 3. There the contaminated stationary distribution is long-tailed and the influence function of the MLE is a quadratic function of x. Following Hampel (1974), we require the boundedness of the influence function as a criterion of robustness. Then our task is to choose, under the boundedness condition of the influence function, a pair of functions a and A possessing the minimum of the asymptotic variance of the corresponding M-estimator or a pair possessing a relatively small variance comparable to MLE, the minimum variance estimator whose influence function is unbounded in general.

As we have seen in Section 3, the asymptotic variances depend not on A but on a, essentially. So we take G as A in the sequel. Our approach based on influence functions is quite different from Huber's minimax approach (Huber (1981)). Our procedure is similar to that of Künsch (1984). In the first step, our optimal robust problem is the following:

(P1) Minimize $\int[\sigma(x)DG/U]^2\mu(x,\theta)dx$ under

$$\int L(\theta)G/U\cdot\mu(x,\theta)dx=0\,, \tag{4.8}$$

$$\text{ess.sup}\,|L(\theta)G/U|\le c(\theta)\quad\text{and} \tag{4.9}$$

$$\int L(\theta)G/U\cdot\dot\mu(x,\theta)dx=1\,, \tag{4.10}$$

where $c(\theta)$ is a constant and

$$G\in F(\theta):=\{G;\,DG\in L^2(\sigma^2\mu(\cdot,\theta)),\,L(\theta)G\in L^2(\mu(\cdot,\theta))\}\,. \tag{4.11}$$

Note that $L(\theta)G=a$ and the asymptotic variance is not equal to the square integral of the influence function. The condition (4.10) is provided for the purpose of normalizing G/U.

If we set $\xi=DG/U$, the problem (P1) is equivalent to the following problem:

(P2) Minimize $\int\xi^2\sigma^2\mu(x,\theta)dx$ under

$$\int L'(\theta)\xi\cdot\mu(x,\theta)dx=0\,, \tag{4.8$'$}$$

$$\text{ess.sup}\,|L'(\theta)\xi|\le c(\theta)\quad\text{and} \tag{4.9$'$}$$

$$\int L'(\theta)\xi\cdot\dot\mu(x,\theta)dx=1\,, \tag{4.10$'$}$$

where $c(\theta)$ is a constant and

$$\xi\in F'(\theta):=\{\xi;\,\xi\in L^2(\sigma^2\mu(\cdot,\theta)),\,L'(\theta)\xi\in L^2(\mu(\cdot,\theta))\}\,. \tag{4.11$'$}$$

Let $\dot{\mu}(\cdot,\theta)/\mu(\cdot,\theta) \in L^2(\mu(\cdot,\theta))$ and C be a subset of $F'(\theta)$ whose elements satisfy the conditions (4.8′)–(4.10′). Then we have

THEOREM 4.1. *If C is non-empty, there exists an optimal solution of* (P2) *in C.*

The space $F' := F'(\theta)$ endowed with Hilbertian norm

$$|\xi|_{F'}^2 = \int \xi^2\sigma^2\mu(x,\theta)dx + \int (L'(\theta)\xi)^2\mu(x,\theta)dx$$

is a Hilbert space. We write $|\xi|_2^2 = \int \xi^2\sigma^2\mu(x,\theta)dx$ and $L^2 = L^2(\sigma^2\mu(\cdot,\theta))$.

Before the proof of this theorem we prepare the following lemmas.

LEMMA 4.1. *Let $(\xi(n); n \geq 1)$ be an arbitrary sequence in C. If $\xi(n)$ converges weakly to $\xi \in F'$ in F',*

(1) ess.sup $|L'(\theta)\xi| \leq c(\theta)$,

(2) *there exists a subsequence $(\xi(n'))$ of $(\xi(n))$ such that $\xi(n')$ converges weakly to ξ in L^2 and $L'(\theta)\xi(n')$ converges weakly to $L'(\theta)\xi$ in $L^2(\mu(\cdot,\theta))$.*

PROOF. We omit the symbol θ. Since the sequence $(\xi(n))$ is weakly bounded, $(\xi(n))$ is bounded in F' by Banach-Steinhaus' theorem. By definition $|\cdot|_2 \leq |\cdot|_{F'}$, and so $(\xi(n))$ is also bounded in L^2. Therefore, there exists a subsequence $(\xi(n'))$ of $(\xi(n))$ which converges weakly to some $\xi' \in L^2$. Similarly, from boundedness of $(L'\xi(n))$ in $L^2(\mu)$, we can suppose that $L'\xi(n')$ also converges to some $\eta \in L^2(\mu)$ weakly in $L^2(\mu)$. Let $\mathscr{D}$ be the space of all smooth functions with compact support, then for $\psi \in \mathscr{D}$, $\langle L'\xi', \psi\rangle = \langle \eta, \psi\rangle$ holds from weak convergences of $(\xi(n'))$ and $(L'\xi(n'))$, where $\langle\ ,\ \rangle$ is the canonical bilinear form on $\mathscr{D}' \times \mathscr{D}$ and $\mathscr{D}'$ is the space of Schwartz's distributions. Therefore, $L'\xi' = \eta$ and so $\xi(n')$ converges weakly to ξ' in F'. Uniqueness of weak limits implies $\xi' = \xi$.

For an arbitrary $\psi \in \mathscr{D}$, by convergence of $L'\xi(n')$ to $L'\xi$ in distribution sense, $|\langle |L'\xi|, \psi\rangle| \leq c|\psi|_{L(R)}$, and so ess.sup $|L'\xi| \leq c$. This completes the proof.

LEMMA 4.2. *C is weakly closed in F'.*

PROOF. Let $\xi(n) \in C$ converge weakly to ξ in F'. It is shown in Lemma 4.1 that $|L'\xi| \leq c$. There exists a subsequence $(\xi(n'))$ of $(\xi(n))$ such that $L'(\theta)\xi(n')$ converges weakly to $L'(\theta)\xi$ in $L^2(\mu(\cdot,\theta))$. Since 1 and $\dot{\mu}(\cdot,\theta)/\mu(\cdot,\theta) \in L^2(\mu(\cdot,\theta))$, (4.8′) and (4.10′) are easily seen; therefore, $\xi \in C$.

PROOF OF THEOREM 4.1. Let $k = \inf\{|\xi|_2; \xi \in C\}$, then there exists a sequence $(\xi(n))$ in C such that $|\xi(n)|_2 \to k$ as $n \to \infty$. Since in $C|L'\xi| \le c$, $|\xi(n)|_{F'}$ is bounded, and there exists a subsequence $(\xi(n'))$ of $(\xi(n))$ which converges weakly to some ξ in F'. Lemma 4.2 implies that $\xi \in C$. By Lemma 4.1 there exists a subsequence $(\xi(n''))$ of $(\xi(n'))$ which converges weakly to ξ in L^2. From the Hilbert space theory, $|\xi|_2 \le \liminf |\xi(n'')|_2 = k$. Hence, $|\xi|_2 = k$, and this ξ is an optimal solution.

Remark 4.1. As seen from the proof of Lemma 3.1 the condition that $\xi \in L^2(\sigma^2\mu(\cdot, \theta))$ implies the consistency condition (4.8′), and so it is not necessary. But to clarify consistency we added the condition.

5. A simple method for constructing robust estimators

Problem (P2) is a linearly constrained minimization problem and its numerical approach has been studied by many authors (e.g., Powell (1982)), but there is another way to solve it approximately, which is simple and gives relatively good results.

$\dot{h}$ is proportional to the influence function of the MLE and we shall modify it in $[\upsilon(+), \infty)$ and $(-\infty, -\upsilon(-)]$ to be bounded, where $\upsilon(+)$ and $\upsilon(-)$ are constants in $\bar{R}^+ = [0, \infty]$. It is not necessary to take them as $\upsilon(+) = \upsilon(-)$. They may depend on θ in general. For notational simplicity we often omit θ in the sequel. Let $\mu(x, \theta)$ be differentiable in θ. Let

$$\phi = \begin{cases} I_-(x) & \text{if} \quad x < -\upsilon(-)\,, \\ \lambda(= -\dot{f}\sigma^{-2}) & \text{if} \quad -\upsilon(-) \le x \le \upsilon(+)\,, \\ I_+(x) & \text{if} \quad \upsilon(+) < x\,, \end{cases} \tag{5.1}$$

where

$$I_-(x) = \exp(-B(x,\theta)) \int_{-\infty}^{x} 2k_-\sigma^{-2}(y) \exp(B(y,\theta))dy\,, \tag{5.2}$$

$$I_+(x) = -\exp(-B(x,\theta)) \int_{x}^{\infty} 2k_+\sigma^{-2}(y) \exp(B(y,\theta))dy\,, \tag{5.3}$$

and k_- and k_+ are constants defined by

$$\lambda(-\upsilon(-)) = I_-(-\upsilon(-)) \quad \text{and} \tag{5.4}$$

$$\lambda(\upsilon(+)) = I_+(\upsilon(+))\,. \tag{5.5}$$

Then ϕ is a continuous function and $L'(\theta)\phi$ is equal to k_-, $-\dot{h}$ and k_+ for $x < -\upsilon(-)$, $-\upsilon(-) \le x \le \upsilon(+)$ and $\upsilon(+) < x$, respectively.

Assume that $l_+ I_{[0,\infty)}$, $l_- I_{(-\infty,0]}$, $\lambda = -\dot{f}\sigma^{-2}$ and $1 \in L^2(\sigma^2\mu(\cdot,\theta))$ then by definition

$$
\begin{aligned}
(5.6)\qquad S(\phi) &:= \int \phi\lambda\sigma^2\mu dx \\
&= \int_{-\upsilon(-)}^{\upsilon(+)} (\dot{f}/\sigma)^2 \mu dx - \int_{\upsilon(+)}^{\infty} l_+ \dot{f}\mu dx \\
&\quad - \int_{-\infty}^{-\upsilon(-)} l_- \dot{f}\mu dx ,
\end{aligned}
$$

$$
\begin{aligned}
(5.7)\qquad Q(\phi) &:= \int \phi^2\sigma^2\mu dx \\
&= \int_{-\upsilon(-)}^{\upsilon(+)} (\dot{f}/\sigma)^2 \mu dx + \int_{\upsilon(+)}^{\infty} l_+^2 \sigma^2\mu dx \\
&\quad + \int_{-\infty}^{-\upsilon(-)} l_-^2 \sigma^2\mu dx .
\end{aligned}
$$

Our candidate for an approximately optimal M-estimator corresponds to

$$(5.8)\qquad \xi = DG/U = \phi/S(\phi) ,$$

whose influence function is given by

$$(5.9)\qquad L'(\theta)\xi = \begin{cases} k_-/S(\phi) & \text{if} \quad x < -\upsilon(-) , \\ -\dot{h}/S(\phi) & \text{if} \quad -\upsilon(-) \le x \le \upsilon(+) , \\ k_+/S(\phi) & \text{if} \quad \upsilon(+) < x . \end{cases}$$

For ξ defined in (5.8) we have

PROPOSITION 5.1. *If* $\lim_{x\to\pm\infty} \phi\sigma^2\dot{\mu} = 0$, (4.8′)–(4.11′) *hold with*

$$(5.10)\quad c(\theta) := \sup\{|k_+|, |k_-|, |\dot{h}(x,\theta)|; -\upsilon(-) \le x \le \upsilon(+)\}/|S(\phi)| .$$

Moreover, $\int \xi^2\sigma^2\mu dx$ *(the asymptotic variance of the M-estimator corresponding to* $L'(\theta)\xi$*) is given by* $Q(\phi)/S(\phi)^2$*, which converges to* $\left[\int (\dot{f}/\sigma)^2 \mu dx\right]^{-1}$*, i.e., the reciprocal of the Fisher information, as* $\upsilon(+)$*,* $\upsilon(-) \to \infty$.

Remark 5.1. This limit of the asymptotic variance is equal to that of the MLE, the minimum asymptotic variance of estimators.

LEMMA 5.1. *$L'(\theta)^*\dot{\mu}\cdot\sigma^{-2}\mu^{-1} = -\dot{f}\sigma^{-2} \equiv \lambda$, where $L'(\theta)^*$ is the conjugate operator of $L'(\theta)$, i.e., $L'(\theta)^* = f(\cdot,\theta) - (1/2)D\sigma^2$.*

PROOF. By (2.2) and (2.3), we obtain

$$\begin{aligned}\dot{\mu}(x,\theta) &= \frac{\partial}{\partial\theta}\left[2\tilde{m}^{-1}\sigma^{-2}(x)\exp\left(\int_0^x 2\sigma^{-2}(u)f(u,\theta)du\right)\right]\\ &= -2\tilde{m}^{-2}\dot{\tilde{m}}\sigma^{-2}(x)\exp\left(\int_0^x 2\sigma^{-2}(u)f(u,\theta)du\right)\\ &\quad + 2\tilde{m}^{-1}\sigma^{-2}(x)\left(\int_0^x 2\sigma^{-2}(u)\dot{f}(u,\theta)du\right)\exp\left(\int_0^x 2\sigma^{-2}(u)f(u,\theta)du\right),\end{aligned}$$

so that

$$\begin{aligned}L'(\theta)^*\dot{\mu}(x,\theta) &= \left[f(x,\theta) - \frac{1}{2}D\sigma^2(x)\right]\dot{\mu}(x,\theta)\\ &= -2\tilde{m}^{-2}\dot{\tilde{m}}f(x,\theta)\sigma^{-2}(x)\exp\left(\int_0^x 2\sigma^{-2}(u)f(u,\theta)du\right)\\ &\quad + 2\tilde{m}^{-1}f(x,\theta)\sigma^{-2}(x)\left(\int_0^x 2\sigma^{-2}(u)\dot{f}(u,\theta)du\right)\\ &\quad\cdot\exp\left(\int_0^x 2\sigma^{-2}(u)f(u,\theta)du\right)\\ &\quad + \tilde{m}^{-2}\dot{\tilde{m}}D\exp\left(\int_0^x 2\sigma^{-2}(u)f(u,\theta)du\right)\\ &\quad - \tilde{m}^{-1}D\left[\int_0^x 2\sigma^{-2}(u)\dot{f}(u,\theta)du\,\exp\left(\int_0^x 2\sigma^{-2}(u)f(u,\theta)du\right)\right]\\ &= -2\tilde{m}^{-1}\sigma^{-2}(x)\dot{f}(x,\theta)\exp\left(\int_0^x 2\sigma^{-2}(u)f(u,\theta)du\right)\\ &\qquad\qquad\text{(using Leibniz's rule for the last term)}\\ &= -\dot{f}(x,\theta)\mu(x,\theta).\end{aligned}$$

This completes the proof.

PROOF OF PROPOSITION 5.1. To see (4.9′) and (4.11′) is trivial and (4.8′) holds by Remark 4.1. By integration by parts and Lemma 5.1,

$$\int_{-N}^{N} L'(\theta)\phi\cdot\dot{\mu}dx = \int_{-N}^{N}\phi L'(\theta)^*\dot{\mu}dx + [\phi\sigma^2\dot{\mu}/2]_{-N}^{N}$$

$$= \int_{-N}^{N} \phi\lambda\sigma^2 \mu dx + [\phi\sigma^2\dot{\mu}/2]_{-N}^{N} .$$

When $N \to \infty$, we have

$$\int L'(\theta)\phi \cdot \dot{\mu} dx = \int \phi\lambda\sigma^2 \mu dx \equiv S(\phi) ,$$

and (4.10′). The last part is obvious from (5.6) and (5.7).

In our case, the integrability condition implies the Fisher consistency, as stated above. However, for ξ defined by (5.8), we can show the consistency condition (4.8′) directly. In fact, by using (5.9) and (5.2)–(5.5),

$$\begin{aligned} S(\phi)\int L'(\theta)\xi \cdot \mu dx &= k_- \int_{-\infty}^{-\upsilon(-)} \mu dx + k_+ \int_{\upsilon(+)}^{\infty} \mu dx - \int_{-\upsilon(-)}^{\upsilon(+)} \dot{h}\mu dx \\ &= -\tilde{m}^{-1}\dot{f}(-\upsilon(-),\theta)\sigma^{-2}(-\upsilon(-)) \\ &\qquad \cdot \exp B(-\upsilon(-),\theta) \\ &\quad + \tilde{m}^{-1}\dot{f}(\upsilon(+),\theta)\sigma^{-2}(\upsilon(+)) \\ &\qquad \cdot \exp B(\upsilon(+),\theta) - \int_{-\upsilon(-)}^{\upsilon(+)} \dot{h}\mu dx . \end{aligned}$$

On the other hand, from (2.9) and integration by parts, we obtain

$$\begin{aligned} &\int_{-\upsilon(-)}^{\upsilon(+)} \dot{h}\mu dx \\ &\quad = \tilde{m}^{-1}\int_{-\upsilon(-)}^{\upsilon(+)} \left(f\dot{f}\sigma^{-2} + \frac{1}{2}\sigma^2 D[\dot{f}\sigma^{-2}] \right) 2\sigma^{-2} \exp\left(\int_0^x 2f\sigma^{-2} \right) dx \\ &\quad = \tilde{m}^{-1}\int_{-\upsilon(-)}^{\upsilon(+)} 2f\dot{f}\sigma^{-4} \exp\left(\int_0^x 2f\sigma^{-2} \right) dx \\ &\qquad - \tilde{m}^{-1}\int_{-\upsilon(-)}^{\upsilon(+)} \dot{f}\sigma^{-2} D \exp\left(\int_0^x 2f\sigma^{-2} \right) dx \\ &\qquad + \tilde{m}^{-1}\dot{f}(\upsilon(+),\theta)\sigma^{-2}(\upsilon(+)) \exp\left(\int_0^{\upsilon(+)} 2f\sigma^{-2} \right) \\ &\qquad - \tilde{m}^{-1}\dot{f}(-\upsilon(-),\theta)\sigma^{-2}(-\upsilon(-)) \exp\left(\int_0^{-\upsilon(-)} 2f\sigma^{-2} \right) \\ &\quad = \tilde{m}^{-1}\dot{f}(\upsilon(+),\theta)\sigma^{-2}(\upsilon(+)) \exp B(\upsilon(+),\theta) \\ &\qquad - \tilde{m}^{-1}\dot{f}(-\upsilon(-),\theta)\sigma^{-2}(-\upsilon(-)) \exp B(-\upsilon(-),\theta) . \end{aligned}$$

Therefore,

$$\int L'(\theta)\xi \cdot \mu dx = 0 .$$

More generally, if we choose bounded continuous functions $k_-(x,\theta)$ and $k_+(x,\theta)$ on $(-\infty, -\upsilon(-)]$ and $[\upsilon(+),\infty)$, respectively, such that I_- and I_+ defined by (5.2) and (5.3) satisfy (5.4) and (5.5), we have the same result as in Proposition 5.1 replacing (5.10) with

$$c(\theta) = \sup [\{|k_-|; x < -\upsilon(-)\} \cup \{|k_+|; \upsilon(+) < x\} \\ \cup \{|\dot{h}|; -\upsilon(-) \leq x \leq \upsilon(+)\}]/|S(\phi)| .$$

If $\mu(x,\theta)$ is decreasing exponentially as $|x| \to \infty$ and $I_\pm$ and $\dot{f}/\sigma$ are of polynomial order, then from (5.6) and (5.7) it is clear that when $\upsilon(+)$, $\upsilon(-) \to \infty$, $S(\phi)$ and $Q(\phi)$ rapidly converge to the information and our M-estimator has an almost minimum asymptotic variance, while $c(\theta)$ are of polynomial orders and not so large in many cases.

Finally, we give an example. Let X be a diffusion process (Ornstein-Uhlenbeck process) defined by the following stochastic differential equation:

$$dX(t) = -\theta X(t)dt + dW(t), \qquad X(0) = x(0) ,$$

$\theta > 0$. Then

$$\begin{aligned} f(x,\theta) &= -\theta x, & \sigma(x) &= 1 , \\ \mu(x,\theta) &= (\theta/\pi)^{1/2} \exp(-\theta x^2), & \lambda &= \lambda(x,\theta) = -\dot{f}\sigma^{-2} = x , \\ h(x,\theta) &= \theta^2 x^2/2 - \theta/2, & \dot{h} &= \theta x^2 - 1/2 . \end{aligned}$$

For ξ defined by (5.1)-(5.8)

$$\xi = \begin{cases} I_-(x)/S(\phi) = \exp(\theta x^2) \int_{-\infty}^{x} 2k_- \\ \qquad \cdot \exp(-\theta y^2) dy/S(\phi) & \text{if} \quad x < -\upsilon(-) , \\ x/S(\phi) & \text{if} \quad -\upsilon(-) \leq x \leq \upsilon(+) , \\ I_+(x)/S(\phi) = -\exp(\theta x^2) \int_{x}^{\infty} 2k_+ \\ \qquad \cdot \exp(-\theta y^2) dy/S(\phi) & \text{if} \quad \upsilon(+) < x , \end{cases}$$

where $k_\pm$ are given by

$$k_- = -v(-)\left[2\exp(\theta v(-)^2)\int_{v(-)}^{\infty}\exp(-\theta y^2)dy\right]^{-1},$$
$$k_+ = -v(+)\left[2\exp(\theta v(+)^2)\int_{v(+)}^{\infty}\exp(-\theta y^2)dy\right]^{-1},$$

and

$$\begin{aligned}S(\phi) &= \int_{-v(-)}^{v(+)} x^2(\theta/\pi)^{1/2}\exp(-\theta x^2)dx \\ &\quad -\int_{v(+)}^{\infty} 2k_+x(\theta/\pi)^{1/2}\int_x^{\infty}\exp(-\theta y^2)dydx \\ &\quad +\int_{-\infty}^{-v(-)} 2k_-x(\theta/\pi)^{1/2}\int_{-x}^{\infty}\exp(-\theta y^2)dydx .\end{aligned}$$

The influence function of this ξ is given by

$$L'(\theta)\xi = \begin{cases} k_-/S(\phi) & \text{if} \quad x < -v(-), \\ (-\theta x^2 + 1/2)/S(\phi) & \text{if} \quad -v(-) \le x \le v(+), \\ k_+/S(\phi) & \text{if} \quad v(+) < x . \end{cases}$$

Especially, for the MLE the counterparts are given by

$$S(\phi_{\mathrm{MLE}}) = \int \lambda^2 \mu dx = 1/(2\theta),$$
$$\xi_{\mathrm{MLE}} = \lambda / S(\phi_{\mathrm{MLE}}) = 2\theta x,$$

and the influence function

$$L'(\theta)\xi_{\mathrm{MLE}} = -\dot{h}/S(\phi_{\mathrm{MLE}}) = -2\theta(\theta x^2 - 1/2) = -2\theta^2 x^2 + \theta .$$

Then the influence function of the MLE is a quadratic function and contamination of the stationary distribution, especially for large $|x|$, influences its bias greatly. We get the following table for our M-estimators when $\theta = 1/2$ and the variance of the MLE (i.e., the minimum variance) equals unity. Our simple M-estimator has a small asymptotic variance and it is preferable with respect to robustness of estimation.

Remarks 5.2. (1) The influence function of our M-estimator is discontinuous in general.

(2) We must check the conditions of Section 3 for the function ξ after making the function ξ. In the above example they are easily checked for constant $v(\pm)$ and $\Theta = [a, b]$ $(a, b > 0)$.

Table 1. Dividing points, bounds of influence functions and asymptotic variances.

$v(-)\,(=v(+))$	$c(\theta)$	Asymptotic variance
0.50	1.88	1.68
1.00	1.30	1.26
1.50	1.75	1.09
2.00	2.45	1.02
2.50	3.43	1.00
3.00	4.68	1.00
3.50	6.18	1.00

6. Optimization problem with another criterion

In Sections 4 and 5, we have treated gross error sensitivity, that is, we have required uniform boundedness of influence functions of estimators. For robustness of estimators, their influence functions should be small in some sense, and we can view the problem in another way by using a different measure of influence functions.

Let $w(x)$ be a positive function of class C^1 on R. Using the same notations as before, the influence function of the M-estimator corresponding to ξ is equal to $L'(\theta)\xi$ and we take as a measure of it the seminorm

$$\int |L'(\theta)\xi|^2 w(x)dx\,. \tag{6.1}$$

Since the influence of contamination of the stationary distribution of the process X on bias of M-estimators is given by (4.4) and if $d\kappa(x) \le w(x)dx$,

$$|Y'(\kappa, a, v(\cdot,\theta))| = \left|\int L'(\theta)\xi d\kappa(x)\right| \le \left\{\int |L'(\theta)\xi|^2 w(x)dx\right\}^{1/2}.$$

Therefore, change rates of M-estimators by contamination dominated by $w(x)dx$ are estimated by the seminorm (6.1).

We want to choose ξ for which the seminorm (6.1) is small and simultaneously the asymptotic variance

$$|\xi|_2^2 = \int \xi^2\sigma^2\mu dx\,, \tag{6.2}$$

is also small. So we define the norm

$$|\xi|_w^2 = \int \xi^2\sigma^2\mu dx + \int |L'(\theta)\xi|^2 w(x)dx\,, \tag{6.3}$$

and we seek ξ which attains the minimum of $|\cdot|_w$-norm. We call the M-

estimator corresponding to such ξ an optimal w-robust M-estimator. The MLE is an optimal zero-robust estimator in some regular classes of estimators.

We assume that for $\theta \in \Theta, \int \lambda^2\sigma^2\mu dx < \infty$ and $\int \sigma^2 \mu dx < \infty$. Let $F(w,\theta)$ be the completion of $\mathscr{D}$ with respect to the norm $|\cdot|_w$. Then $(F(w,\theta), |\cdot|_w)$ is a Hilbert space. Let $C(w,\theta) = F(w,\theta) \cap \left\{ \xi; \int \xi\lambda\sigma^2\mu dx = 1 \right\}$ (this integral corresponds to that of condition (4.10′) by Lemma 5.1). Note that for $\xi \in F(w,\theta)$ the consistency condition holds. Now, our problem is

(P3) Minimize $|\xi|_w$ in $\xi \in C(w,\theta)$.

By the inequality (6.5) below, it is shown that the linear functional $\xi \to \int \xi\lambda\sigma^2\mu dx$ is bounded on $F(w,\theta)$. Hence, $C(w,\theta)$ is a closed convex set in the Hilbert space $F(w,\theta)$, and it is well known that there exists a unique element ξ_0 in $C(w,\theta)$ such that $|\xi_0| = \min\{|\xi|; \xi \in C(w,\theta)\}$.

We can, however, get this optimal ξ_0 by solving a boundary-value problem directly. Consider the second order differential equation

$$L'(\theta)^*[w(x)L'(\theta)\xi] + \sigma^2(x)\mu(x,\theta)\xi = \lambda(x,\theta)\sigma^2(x)\mu(x,\theta)\,, \tag{6.4}$$

where $L'(\theta)$, $L'(\theta)^*$, $\mu(x,\theta)$ etc. are the same as before and differentials with respect to x are in the sense of distributions.

THEOREM 6.1. *There exists a unique solution of* (6.4) *in* $F(w,\theta)$.

PROOF. The following estimate holds: for $\xi \in F(w,\theta)$

$$\begin{aligned} \left| \int \xi\lambda\sigma^2\mu dx \right| &\le \left\{ \int \lambda^2\sigma^2\mu dx \right\}^{1/2} \left\{ \int \xi^2\sigma^2\mu dx \right\}^{1/2} \\ &\le \left\{ \int \lambda^2\sigma^2\mu dx \right\}^{1/2} |\xi|_w\,. \end{aligned} \tag{6.5}$$

Therefore, the linear functional $F(w,\theta) \ni \xi \to \int \xi\lambda\sigma^2\mu dx$ is bounded. By Riesz' theorem there exists a unique ξ_1 in $F(w,\theta)$ such that for $\xi \in F(w,\theta)$

$$\int \xi\xi_1\sigma^2\mu dx + \int (L'(\theta)\xi)(L'(\theta)\xi_1)w dx = \int \xi\lambda\sigma^2\mu dx\,. \tag{6.6}$$

Especially, for $\xi \in \mathscr{D}$

$$\int \xi\{L'(\theta)^*[wL'(\theta)\xi_1] + \xi_1\sigma^2\mu\}dx = \int \xi\lambda\sigma^2\mu dx\,,$$

hence (6.4) is shown. Uniqueness is obvious.

Let $\xi_0 = \xi_1 / |\xi_1|_w^2$. From Theorem 6.1 we have

$$L'(\theta)^*[wL'(\theta)\xi_0] + \sigma^2\mu\xi_0 = |\xi_0|_w^2\lambda\sigma^2\mu \,, \tag{6.7}$$

since $|\xi_0|_w = 1/|\xi_1|_w$.

THEOREM 6.2. *ξ_0 is the unique solution of the problem* (P3), *that is, $\xi_0 \in C(w,\theta)$ and $|\xi_0|_w = \min\{|\xi|_w; \xi \in C(w,\theta)\}$.*

PROOF. It is obvious that $\xi_0 \in C(w,\theta)$ from (6.6). For $\xi \in \mathscr{D}$,

$$\begin{aligned} 0 &\le |\xi - \xi_0|_w^2 \\ &= |\xi|_w^2 + |\xi_0|_w^2 - 2\left\{\int \xi\xi_0\sigma^2\mu dx + \int (L'(\theta)\xi)(L'(\theta)\xi_0)wdx\right\} \\ &= |\xi|_w^2 + |\xi_0|_w^2 - 2\left\{\int \xi\xi_0\sigma^2\mu dx + \int \xi L'(\theta)^*[wL'(\theta)\xi_0]dx\right\} \\ &= |\xi|_w^2 + |\xi_0|_w^2 - 2|\xi_0|_w^2 \int \xi\lambda\sigma^2\mu dx \,. \end{aligned} \tag{6.8}$$

For $\xi \in C(w,\theta)$ there exists a sequence $(\phi(n))$ in $\mathscr{D}$ such that $|\xi - \phi(n)|_w \to 0$ as $n \to \infty$ and so $|\phi(n)|_w \to |\xi|_w$ and $\int \phi(n)\lambda\sigma^2\mu dx \to 1$. By approximation argument (6.8) implies that for $\xi \in C(w,\theta)$

$$0 \le |\xi|_w^2 + |\xi_0|_w^2 - 2|\xi_0|_w^2 = |\xi|_w^2 - |\xi_0|_w^2 \,.$$

Moreover, the uniqueness is obvious and we have the result.

Finally, we give an illustrative example. Consider the Ornstein-Uhlenbeck process given in Section 5. For fixed $w(x)$ the optimal equation (6.4) is given by

$$-\xi'' - (w'/w)\xi' + [4\theta^2x^2 + 2(w'/w)\theta x + 2\theta + 4\mu/w]\xi = 4\mu x/w \,, \tag{6.9}$$

where "$'$" stands for differential with respect to x and $\mu = (\theta/\pi)^{1/2} \cdot \exp(-\theta x^2)$. To get a numerical solution of problem (6.9) it is convenient to transform the variable ξ to an appropriate one. However, we will not go into details here. We have a table of asymptotic variances of optimal w-robust estimators for various functions $w(x)$ when $\theta = 1/2$ and the variance of the MLE is unity. The optimal w-robust M-estimator corresponding to fast decreasing $w(x)$ has a relatively small variance.

Table 2. Functions $w(x)$ and asymptotic variances of corresponding optimal w-robust M-estimators.

$w(x)$	Asymptotic variance
1	1.48
$1/2(1+\|x\|^2)$	1.13
$1/2(1+\|x\|^4)$	1.03
$1/2(1+\|x\|^6)$	1.02
$\exp(-x^2/2)$	1.00

6.1 *Concluding remarks*

Theoretically, diffusion coefficients can be estimated without error. But practically it may be necessary to estimate them. It would be an interesting problem to investigate the effect of discretization of continuous observations on estimation of diffusion coefficients and to seek the robust procedures.

Acknowledgements

The author is grateful to Professor Nobuo Inagaki, Dr. Masao Kondo, Dr. Masanobu Taniguchi and Dr. Minoru Yoshida for encouragement and valuable discussions. The example of Section 4 was suggested by the referee. I express my gratitude to the referees for extensive help and also to Mr. Toshiharu Hayashi for help in computer programming.

REFERENCES

Bustos, O. (1982). General M-estimates for contaminated pth-order autoregressive processes: Consistency and asymptotic normality, *Z. Wahrsch. Verw. Gebiete*, **59**, 491-504.

Bustos, O. and Yohai, V. J. (1986). Robust estimates for ARMA models, *J. Amer. Statist. Assoc.*, **81**, 155-168.

Denby, L. and Martin, R. D. (1979). Robust estimation of the first order autoregressive parameter, *J. Amer. Statist. Assoc.*, **74**, 140-146.

Feigin, P. D. (1985). Stable convergence of semimartingales, *Stochastic Process. Appl.*, **19**, 125-134.

Gihman, I. I. and Skorohod, A. V. (1972). *Stochastic Differential Equations*, Springer, Berlin-New York.

Hampel, F. R. (1974). The influence curve and its role in robust estimation, *J. Amer. Statist. Assoc.*, **69**, 383-393.

Hampel, F. R., Ronchetti, E. M., Rousseeuw, P. J. and Stahel, W. A. (1986). *Robust Statistics*, Wiley, New York.

Huber, P. J. (1981). *Robust Statistics*, Wiley, New York.

Ikeda, N. and Watanabe, S. (1981). *Stochastic Differential Equations and Diffusion Processes*, North-Holland, Amsterdam-New York.

Itô, K. and McKean, H. P. Jr. (1965). *Diffusion Processes and Their Sample Paths*, Springer, Berlin-New York.

Kleiner, B., Martin, R. D. and Thomson, D. J. (1979). Robust estimation of power spectra, *J. Roy. Statist. Soc. Ser. B*, **41**, 313-351.

Krylov, N. V. (1980). *Controlled Diffusion Processes*, Springer, Berlin-New York.

Künsch, H. (1984). Infinitesimal robustness for autoregressive processes, *Ann. Statist.*, **12**, 843–863.

Kutoyants, Yu. A. (1977). Estimation of the trend parameter of a diffusion in the smooth case, *Theory Probab. Appl.*, **22**, 399–405.

Kutoyants, Yu. A. (1978). Estimation of a parameter of a diffusion process, *Theory Probab. Appl.*, **23**, 641–649.

Kutoyants, Yu. A. (1984). *Parameter Estimation for Stochastic Processes*, (trans. and ed. B. L. S. Prakasa Rao), Heldermann Verlag, Berlin.

Lánska, V. (1979). Minimum contrast estimation in diffusion processes, *J. Appl. Probab.*, **16**, 65–75.

Lepingle, D. (1978). *Sur le Comportement Asymptotique des Martingales Locales*, Lecture Notes in Mathematics, 649, 148–161, Springer, Berlin-New York.

Liptser, R. Sh. and Shiryayev, A. N. (1977). *Statistics of Random Processes*, Vol. I, Springer, Berlin-New York.

Mandl, P. (1968). *Analytic Treatment of One-Dimensional Markov Processes*, Springer, Prague.

Martin, R. D. and Yohai, V. J. (1985). Robustness in time series and estimating ARMA models, *Handbook of Statistics*, **5**, 119–155.

Martin, R. D. and Yohai, V. J. (1986). Influence functionals for time series, *Ann. Statist.*, **14**, 781–818.

McKeague, I. W. (1984). Estimation for diffusion processes under misspecified models, *J. Appl. Probab.*, **21**, 511–520.

Powell, M. J. D. (ed.) (1982). *Nonlinear Optimization* 1981, Academic Press, New York.

Rao, B. L. S. P. and Rubin, H. (1981). Asymptotic theory of estimation in non-linear stochastic differential equations, *Sankhyā Ser. A*, **43**, 170–189.