statistics 16 (1985) 2, 263–275

Estimation of the Drift for Diffusion Process[1]

B. L. S. PRAKASA RAO

Indian Statistical Institute, New Delhi

Summary. A brief survey of recent results in the area of parametric and nonparametric estimation of the drift coefficient of a diffusion process are presented. Some open problems are stated.

1. Introduction

Let $\{X_t, t \geqq 0\}$ be a stochastic process defined on a probability space $(\Omega, \mathfrak{F}, \mathsf{P})$ and satisfying the ITO stochastic differential equation

$$dX_t = a(X_t)\,dt + \sigma(X_t)\,dW_t, \quad t \geqq 0, \quad X_0 = X \tag{1.1}$$

where $\{W_t, t \geqq 0\}$ is the standard WIENER process defined on $(\Omega, \mathfrak{F}, \mathsf{P})$ and X_0 independent of $\{W_t, t \geqq 0\}$. Suppose the following conditions hold:

(1) $\mathsf{E}X^2 < \infty$, and

(2) $a(x)$ and $\sigma(x)$ are real valued BOREL-measurable functions satisfying the LIPSCHITZ conditions

$$|a(x) - a(y)| \leqq K\,|x-y|$$
$$|\sigma(x) - \sigma(y)| \leqq K\,|x-y|$$

and the linear growth conditions

$$|a(x)| \leqq K\,(1+x^2)^{\frac{1}{2}},$$
$$|\sigma(x)| \leqq K\,(1+x^2)^{\frac{1}{2}}$$

where $K > 0$.

It is known that the equation (1.1) has a unique solution with probability one and it is a strong MARKOV process under the conditions (1) and (2). (See GIKHMAN and SKOROKOD (1972), WONG (1971)). Such a solution is called a Diffusion process. Let $p_{X_t|X_0=a}$ denote the transition density of X_t given $X_0 = a$. Under some additional conditions, this transition density converges to a limiting density $p(\cdot)$ as $t \to \infty$. Suppose the initial density of $X_0 = X$ is $p(\cdot)$. Then the process $\{X_t\}$ is a stationary

[1] Paper presented at the 6th International Summer School on Problems of Model Choice and Parameter Estimation in Regression Analysis, November 1983, Sellin, GDR.

MARKOV process. It is again known that $(a(\cdot), \sigma(\cdot))$ are connected to $p(\cdot)$ by

$$\left(\frac{1}{2}\,\sigma^2 p\right)^{(1)} = ap\,, \tag{1.2}$$

where $f^{(1)}$ denotes the derivative of f (See WONG (1971)).

In the following discussion, we shall assume that either $\sigma(.)$ is known or it is an unknown constant. It is known that the sample paths of the process X satisfying (1.1) are continuous with probability one. Let μ_X^T be the measure induced on $C\,(0,\, T]$ by the process X observed on $[0,\, T]$. $C[0,\, T]$ is the space of continuous functions on $[0,\, T]$ endowed with the sup norm topology. Let μ_W^T be the measure induced on $C[0,\, T]$ by the standard WIENER process on $[0,\, T]$. If

$$\mathsf{P}\left\{\int_0^T \left[\frac{a(X_t)}{\sigma(X_t)}\right]^2 \mathrm{d}t < \infty\right\} = 1$$

then $\mu_X^T \ll \mu_W^T$ and

$$\log \frac{\mathrm{d}\mu_X^T}{\mathrm{d}\mu_W^T} = \left[\int_0^T \frac{a(X_t)}{\sigma(X_t)}\,\mathrm{d}X_t - \frac{1}{2}\int_0^T \left[\frac{a(X_t)}{\sigma(X_t)}\right]^2 \mathrm{d}t\right] \quad \text{a.e.} \quad [\mu_W^T]\,. \tag{1.3}$$

Using this explicit expression for the RADON-NIKODYM derivative, one can study inference problems concerning either estimation or testing of the drift coefficient of a diffusion process. We shall now describe some recent work in this area by this author and others. This survey is not exhaustive.

2. Parametric Inference

Let us consider the stochastic differential equation (1.1) in the form viz.,

$$\mathrm{d}X_t = a\,(\theta,\, X_t)\,\mathrm{d}t + \sigma(X_t)\,\mathrm{d}W_t, \quad t \geqq 0, \quad X_0 = X \tag{2.1}$$

where $a(.,.)$ and $\sigma(.)$ are known functions and $\theta \in \Theta \subset \mathrm{R}$ is unknown. Maximum likelihood estimation (MLE) of θ given that the process X_t is observed over $[0,\, T]$ continuously has been discussed by several authors both when $a(.,\, x)$ is linear as well as when $a(.,\, x)$ is non-linear in θ. An extensive survey of this discussion in the linear case is given in BASAWA and PRAKASA RAO (1980a, b). We do not go into the details here. For related work in the nonlinear case, see KUTOYANTS (1984), IBRAGIMOV and HASMINSKII (1981), PRAKASA RAO and RUBIN (1981) and BASU (1983). The probabilistic techniques used to study asymptotic theory in both PRAKASA RAO and RUBIN (1981) and BASU (1983b), are of independent interest. For simplicity, we shall assume that $\sigma(.) \equiv 1$ and suppose that

$$\mathsf{P}\left\{\int_0^T [a(\theta,\, X_t)]^2\,\mathrm{d}t < \infty\right\} = 1 \tag{2.2}$$

for all $T \geqq 0$. Then, it is known that $\mu_\theta^T \ll \mu_W^T$ where μ_θ^T denotes the measure induced on $C[0,\, T]$ by the process X when the process X is observed over $[0,\, T]$ and

μ_W^T is as defined earlier. Further more

$$\log L_T(\theta) \equiv \log \frac{d\mu_\theta^T}{d\mu_W^T} = \int_0^T a(\theta, X_t)\, dX_t - \frac{1}{2} \int_0^T [a(\theta, X_t)]^2\, dt\,. \tag{2.3}$$

Maximum Probability Estimation (MPE)

Let

$$Z_T(\theta) = \int_{\theta - T^{-\frac{1}{2}}}^{\theta + T^{-\frac{1}{2}}} L_T(t)\, dt\,. \tag{2.4}$$

Any measurable function $\tilde{\theta}_T$ for which the integrated likelihood $Z_T(\theta)$ is maximized with respect to θ is called a *maximum probability estimator* (MPE) of θ based on the sample path of the process X on $[0, T]$.

Existence and asymptotic properties of this estimator were studied in PRAKASA RAO (1982).

Suppose the solution $\{X_t\}$ of (1.1) is stationary and ergodic. Further suppose that $a(.,.)$ is differentiable with respect to θ with derivative $a^{(1)}(., x)$ and $0 < \sigma^2(\theta) = \mathsf{E}_\theta\, [a^{(1)}(\theta, X_0)]^2 < \infty$. Under some further regularity conditions, it was shown that there exists a MPE $\tilde{\theta}_T$ of θ which is $\sqrt{T}$-consistent i.e., $T^{\frac{1}{2}}(\tilde{\theta}_T - \theta) = O_p(1)$ and

$$T^{\frac{1}{2}}(\tilde{\theta}_T - \theta) \xrightarrow{\mathfrak{L}} N\left(0, \frac{1}{\sigma^2(\theta)}\right) \quad \text{as} \quad T \to \infty\,. \tag{2.5}$$

Further more, this estimator is asymptotically efficient in the sense of WEISS and WOLFOWITZ (1974). For details, see PRAKASA RAO (1982).

Least Squares Estimation (LSE)

A basic assumption, in the study of either maximum likelihood estimation or maximum probability estimation of parameters of diffusion processes, is that the process can be observed continuously in time. It is obvious that this assumption is too strong and impossible to meet in actual practice. For convenience, write $X(t)$ for X_t.

Suppose the process X is observed at the points $t_k = k\,\frac{T}{n}$ for $k = 0, 1, \ldots, n$ where $\frac{T}{n} \to 0$ and $T \to \infty$. Let

$$Q_n(\theta) = \sum_{k=0}^{n-1} \left[X(t_{k+1}) - X(t_k) - a(\theta, X(t_k))\, \frac{T}{n} \right]^2 \tag{2.6}$$

and $\hat{\theta}_{n,T}$ be defined to be a measurable function from $\Omega \to \Theta$ such that

$$Q_n(\hat{\theta}_{n,T}) = \inf_{\theta \in \Theta} Q_n(\theta) \tag{2.7}$$

where Θ is the parameter space. Clearly a measurable solution $\hat{\theta}_{n,T}$ exists if $a(\theta, x)$ is continuous in θ and Θ is compact. DOROGOVCEV (1976) has given sufficient conditions for weak consistency of the estimator $\hat{\theta}_{n,T}$. Under some further conditions, it was proved in PRAKASA RAO (1983a) that

$$T^{\frac{1}{2}} (\hat{\theta}_{n,T} - \theta) \xrightarrow{\mathcal{L}} N\left(0, \frac{1}{\sigma^2(\theta)}\right) \tag{2.8}$$

as $T \to \infty$ and $n^{-\frac{1}{2}} T \to 0$.

Following RAO (1973), an estimator θ_T of θ is said to be *first order efficient* if there exist non-random functions $\alpha(\theta)$ and $\beta(\theta)$ such that

$$T^{-\frac{1}{2}} \frac{\mathrm{d} \log L_T(\theta)}{\mathrm{d}\theta} - \alpha(\theta) - \beta(\theta)\, T^{\frac{1}{2}} (\theta_T - \theta)' \xrightarrow{P} 0 \quad \text{as} \quad T \to \infty . \tag{2.9}$$

It has been shown in PRAKASA RAO (1983b) that the LSE is first order efficient as $T \to \infty$ and $n^{-\frac{1}{2}} T \to 0$.

We might mention here that LE BRETON (1975) studied asymptotic properties of the descritized versions of MLE for linear parametric stochastic differential equations with $a(\theta, x) = \theta A(x)$. He obtained the closeness of these to the MLE when the total sample path is available. Such results are unknown for non-linear stochastic differential equations. Results of BERRY-ESSEEN type bounds for distribution of LSE as well as MLE are worth investigating in both the cases of descritized sampling as well as in the case of continuous sampling.

BAYES Estimation (BE)

Suppose Λ is a prior probability measure on $(\Theta, \mathfrak{B})$ where $\mathfrak{B}$ is the σ-algebra of BOREL subsets of Ω. Assume that Λ has a density $\lambda(.)$ which with respect to the LEBESGUE measure and the density is continuous and positive in an open neighbourhood of θ, the true parameter. Let θ_0 be an arbitrary but fixed value of θ.

We assume that the process X has been observed continuously on $[0, T]$. The posterior density of θ, given $X_0^T \equiv \{X(t) : 0 \leq t \leq T\}$ is defined by

$$p(\theta \mid X_0^T) = \frac{\dfrac{\mathrm{d}\mu_\theta^T}{\mathrm{d}\mu_{\theta_0}^T} \lambda(\theta)}{\displaystyle\int_\Theta \frac{\mathrm{d}\mu_\theta^T}{\mathrm{d}\mu_{\theta_0}^T} \lambda(\theta)\, \mathrm{d}\theta} \tag{2.10}$$

where $\log \dfrac{\mathrm{d}\mu_\theta^T}{\mathrm{d}\mu_W^T}$ is given by (2.3).

Let $p^*(t \mid X(t) : 0 \leq t \leq T\}$ be the posterior density of $T^{\frac{1}{2}}(\theta - \hat{\theta}_T)$ where $\hat{\theta}_T$ is an MLE of θ and define

$$\gamma_T(t) = \frac{\mathrm{d}\mu^T_{\hat{\theta}_T + tT^{-\frac{1}{2}}}}{\mathrm{d}\mu^T_{\hat{\theta}_T}} . \tag{2.11}$$

Suppose

$$\log \gamma_T(t) \to -\frac{1}{2}\beta t^2 \quad \text{a.s.} \quad \text{as} \quad T \to \infty \tag{2.12}$$

for each t where $\beta > 0$ and there exists $0 < \varepsilon < \beta$ such that

$$\int_{-\infty}^{\infty} K(t) \exp\left(-(\beta-\varepsilon)\frac{t^2}{2}\right) dt < \infty . \tag{2.13}$$

Under some additional conditions, BASU (1983a) proved that

$$\lim_{T\to\infty} \int_{-\infty}^{\infty} K(t) \left| p^*(t \mid X(t): 0 \leq t \leq T) - \left(\frac{\beta}{2\pi}\right)^{\frac{1}{2}} \exp\left(-\frac{1}{2}\beta t^2\right)\right| dt = 0 \quad \text{a.s.} \tag{2.14}$$

which is the BERNSTEIN-VON MISES theorem. PRAKASA RAO (1981) proved the above result for linear parametric stochastic differential equations and it was proved for diffusion fields in PRAKASA RAO (1984). As a consequence it can be shown that BAYES estimator (BE) θ_T^* and MLE $\hat{\theta}_T$ have the same asymptotic properties and asymptotic distributions for suitable class of loss functions of the type $l(\theta, \Phi) = l(|\theta - \Phi|)$ and $l(.)$ non-decreasing.

3. Nonparametric Inference

Suppose the functional form of the drift coefficient in (1.1) is unknown and the problem of interest again now is to estimate the drift $a(.)$ based on either discrete or continuous sampling of the process X satisfying (1.1) on $[0, T]$.

Continuous sampling

Suppose a complete observation of the process X over $[0, T]$ is available.

Method of kernels: Suppose that the solution $\{X_t, t \geq 0\}$ of (1.1) is a stationary MARKOV process. For each $t \in [0, \infty)$, define the transition probability operator

$$H_t f = \mathsf{E}(f(X_t) \mid X_0) \tag{3.1}$$

for any bounded BOREL-measurable function f. Define

$$|H_t|_2 = \sup_{f: \mathsf{E} f(X_0) = 0} \frac{\mathsf{E}^{\frac{1}{2}}(H_t f)^2(X)}{\mathsf{E}^{\frac{1}{2}} f^2(X)} . \tag{3.2}$$

The operator H_t is said to satisfy the *condition* $G_2(s, \alpha)$ if there exists some $s > 0$ such that

$$|H_s|_2 \leq \alpha \tag{3.3}$$

where $0 < \alpha < 1$. Since $\{X_t\}$ is a stationary MARKOV process, the family of operators $\{H_t, t \geq 0\}$ is a semi-group and it can be shown (See BANON (1977)) that

$$|H_t|_2 < \frac{\beta^t}{\alpha} \quad \text{with} \quad \beta = \alpha^{1/s} < 1 . \tag{3.4}$$

Let $K(.)$ be a bounded probability density on R and h be a bounded positive function on R_+ such that

(i) $h_t \downarrow 0 \quad \text{as} \quad t \to \infty$ (3.5a)

(ii) $\gamma_t = \int_0^t h_s \, ds < \infty$ (3.5b)

and

(iii) $\gamma_t \to \infty \quad \text{as} \quad t \to \infty .$ (3.5c)

For $t > 0$, let

$$p_t(x_0) = \frac{1}{\gamma_t} \int_0^t K\left(\frac{x_0 - X_s}{h_s}\right) ds . \tag{3.6}$$

$p_t(x_0)$ can be considered as a kernel type estimator of the marginal density $p(x_0)$ of X_0. Banon (1978) studied consistency properties of such density estimators. He has also given sufficient conditions under which a process X which is a solution of (1.1) satisfies the condition $G_2(s, \alpha)$ for some $s > 0$ and $0 < \alpha < 1$. Relation (1.2) implies that

$$p(x)\, a(x) = \frac{1}{2} \sigma^2(x)\, p^{(1)}(x) + \frac{1}{2} \sigma^2(x)^{(1)}\, p(x)$$

and hence

$$a(x) = \frac{1}{2} \left\{\sigma^2(x) \frac{p^{(1)}(x)}{p(x)} + \sigma^2(x)^{(1)}\right\} \tag{3.7}$$

provided $p(x) \neq 0$. One can estimate the derivative $p^{(1)}(x)$ of the density $p(x)$ again by the method of kernels (see Prakasa Rao (1983b).

Theorem 3.1. (Banon (1978)). *Let $K_1(.)$ be a bounded probability density function and $K_2(.)$ be a continuous probability density function of bounded variation such that $K_2^{(1)}(.)$ is bounded. Let h_t be a bounded positive function such that*

(i) $h_t \downarrow 0 \quad as \quad t \to \infty ,$ (3.8a)

(ii) $\gamma_t = \int_0^t h_s ds < \infty$ (3.8b)

(iii) $\gamma_t \to \infty \quad as \quad t \to \infty \quad and$ (3.8c)

(iv) $h_t^2 \gamma_t \to \infty \quad as \quad t \to \infty .$ (3.8d)

Define

$$q_t(x) = \frac{\int_0^t \frac{1}{h_s} K_2^{(1)}\left(\frac{x - X_s}{h_s}\right) ds}{\int_0^t K_1\left(\frac{x - X_s}{h_s}\right) ds + \varepsilon} \tag{3.9}$$

for some fixed $\varepsilon > 0$. If $p^{(1)}(.)$ is continuous and bounded then

$$q_t(x) \xrightarrow{P} q(x) = \frac{p^{(1)}(x)}{p(x)} \quad as \quad t \to \infty . \tag{3.10}$$

In particular, one can estimate the drift coefficient $a(x)$ by

$$a_t(x) = \frac{1}{2} \{\sigma^2(x)^{(1)} + \sigma^2(x)\, q_t(x)\} \tag{3.11}$$

and

$$a_t(x) \xrightarrow{P} a(x) \quad \text{as} \quad t \to \infty \tag{3.12}$$

If $\sigma(\cdot)$ is unknown but a constant, then one can obtain an estimator

$$\sigma_n^2 = \frac{1}{n} \sum_{i=1}^{n} \frac{1}{\tau_i} (X_{t_i+\tau_i} - X_{t_i})^2 \tag{3.13}$$

of σ^2 where $\{\tau_i\}$ is a positive sequence tending to zero and $t_1 \geqq 0$ and $t_i + \tau_i \leqq t_{i+1}$, $i = 1, 2, \ldots$ If $\mathsf{E}(X^4) < \infty$, then it can be shown that

$$\sigma_n^2 \xrightarrow{q.m} \sigma^2 \quad \text{as} \quad n \to \infty$$

by methods in WONG and ZAKAI (1965) (see BANON (1978)). Substituting σ_n^2 as an estimator of σ^2, we have

$$a_{t,n}(x) = \frac{1}{2} \sigma_n^2 q_t(x) \tag{3.14}$$

as an estimator of $a(x)$ and

$$a_{t,n}(x) \xrightarrow{P} a(x) \quad \text{as} \quad t \to \infty \quad \text{and} \quad n \to \infty\,. \tag{3.15}$$

Strongly consistent estimators of $a(x)$ can also be obtained (see BANON and NGUYEN (1978)) by considering a wider class of estimators of $p(x)$ of the type

$$p_t(x) = \left\{\int_0^t h(s)\, H(h(s))\, ds\right\}^{-1} \left\{\int_0^t H(h(s))\, K\left(\frac{X_s - x}{h(s)}\right) ds\right\}. \tag{3.16}$$

Observe that the estimators suggested are recursive in nature and hence are easy to compute.

Method of delta-families

A family $\{\delta_h, h > 0\}$ of non-negative L_∞-functions is called a *delta-family of positive type* α if

(i) there exist constants $A > 0$, $B > 0$ such that

$$\left| 1 - \int_{-A}^{B} \delta_h(x)\, dx \right| = O(h^\alpha) \tag{3.17a}$$

(ii) $$\sup \{|\delta_h(x)| : |x| \geqq h^\alpha\} = O(h^\alpha) \tag{3.17b}$$

(iii) $$\|\delta_h\|_\infty \cong h^{-1} \tag{3.17c}$$

as $h \to 0$.

Let $h(t)$ be a non-negative real valued function such that $h(t) \downarrow 0$ as $t \to \infty$ and $h(t)$ is locally integrable i.e.,

$$\gamma(t) = \int_0^t h(s)\, ds < \infty, \quad \gamma(t) \to \infty \quad \text{as} \quad t \to \infty\,. \tag{3.18}$$

Define

$$p_t^*(x) = \frac{1}{\gamma(t)} \int_0^t h(s)\, \delta_{h(s)}\, (x - X_s)\, ds\,. \tag{3.19}$$

Prakasa Rao (1979b) studied asymptotic properties such as consistency and asymptotic normality of this estimator. If the family $\{\delta_h,\ h > 0\}$ is such that the above conditions (i) to (iii) hold and, further if

(iv) $\delta_h(.)$ is differentiable for every $h > 0$ such that

$\|\delta_h^{(1)}\|_\infty \cong h^{-1}$,

then

$$q_t(x) = \frac{\int_0^t h(s)\, \delta_{h(s)}^{(1)}\, (x - X_s)\, ds}{\int_0^t h(s)\, \delta_{h(s)}\, (y - X_s)\, ds} \tag{3.20}$$

is an estimator of $\frac{p^{(1)}(x)}{p(x)}$ whenever $p(x) \neq 0$ and one can use this estimator $q_t(x)$ to construct estimators of $a(x)$ as before both in case $\sigma(.)$ is known or $\sigma(.)$ is unknown but constant. For details regarding consistency and asymptotic distribution, see Prakasa Rao (1979b, 1983b).

Method of Sieves

If the drift $a(.)$ is completely unknown, but we are interested in estimating it at a point, then we have proposed estimators above. Let us now consider the case of stochastic differential equation

$$dX_t = a(X_t)\, dt + dW_t, \quad t \geqq 0, \quad X_0 = X \tag{3.21}$$

where we would like to estimate $a(x)$ for $x \in [-\lambda, \lambda]$ for some fixed $\lambda > 0$. Let

$$u(x) = \int_0^x \exp\left(-2 \int_0^y a(z)\, dz\right) dy\,. \tag{3.22}$$

Suppose

$$\lim_{x \to +\infty} u(x) = +\infty \quad \text{and} \quad \lim_{x \to -\infty} u(x) = -\infty\,. \tag{3.23}$$

The process X_t will be recurrent in view of the above condition (see Friedman (1975), Ch. 9). Let

$$e_1 = \inf\{t \geqq 0,\ |X_t| = \lambda\}\,,$$
$$R_1 = \inf\{t \geqq e_1,\ X_t = 0\}\,,$$
$$e_{i+1} = \inf\{t \geqq R_i : |X_t| = \lambda\}, \quad i = 1, 2, \ldots$$
$$R_{i+1} = \inf\{t \geqq e_{i+1} : X_t = 0\}, \quad i = 1, 2, \ldots$$

and

$$X_t^1 = I_{[0, e_1]}(t)\, X_t\,,$$
$$X_t^i = I_{[0, e_i - R_{i-1}]}(t)\, X_{R_{i-1} + t}, \quad i = 2, 3, \ldots$$

Observe that $X_t^1, X_t^2, \ldots$ are i.i.d random processes. Here $I_A(t)$ is the indicator function of the set A. The parameter space Θ here is the space of LIPSCHITZ continuous functions i.e.,

$$\Theta = \{a(x) : \exists L \text{ such that } |a(x)-a(y)| \leq L\,|x-y| \text{ for all } x, y \in \mathrm{R}\} \tag{3.24}$$

endowed with the metric

$$d(\alpha, \beta) = \left(\mathsf{E} \int_0^{e_1} |\alpha(X_s) - \beta(X_s)|^2 \, \mathrm{d}s\right)^{\frac{1}{2}}$$

Let

$$S_m = \left\{ \sum_{k=-m}^{m} a_k e^{ik\frac{\pi}{\lambda}x} : \sum_{-m}^{m} |a_k| \leq K \log m,\ a_k = \bar{a}_{-k} \text{ for each } k \right\} \tag{3.25}$$

for some constant K. Then $\bigcup_{m=1}^{\infty} S_m$ is dense in Θ since any $a \in \Theta$ can be uniformly approximated by the trigonometric polynomials on $[-\lambda, \lambda]$. The RADON-NIKODYM derivative of the measure generated by the process X with respect to the WIENER process given that the process X is observed upto time e_1 is

$$f(x, a) \equiv \exp\left(\int_0^{e_1} a(X_s)\, \mathrm{d}X_s - \frac{1}{2} \int_0^{e_1} a^2(X_s)\, \mathrm{d}s\right)$$

where e_1 is the first exit time of the process X_t for $[-\lambda, \lambda]$. Define

$$L_n(a) = \prod_{i=1}^{n} f(x_t^i, a) \tag{3.26}$$

and

$$M_m^n = \{\alpha \in S_m \mid L_n(\alpha) = \sup_{\beta \in S_m} L_n(\beta)\}\,. \tag{3.27}$$

Hence M_m^n is the set of all maximum likelihood estimators in S_m of α given the processes $X_t^1, X_t^2, \ldots, X_t^n$. GEMAN and HWANG (1979) proved the following theorem.

Theorem 3.2. *If $m_n \cong n^{1-\varepsilon}$ for some $\varepsilon > 0$, then $M_{m_n}^n \to a$ a.s.*

The method proposed above is known as "The method of sieves" suggested by GRENANDER (1980). Here "Sieve" refers to the subspace over which the likelihood function is maximized. The properties of sequence of estimators so obtained depend on the growth of the sieve as compared to the growth of the sample size. The reader is referred to GRENANDER (1980) and GEMAN and HWANG (1979) for more details. NGUYEN and PHAM (1980) used this method for the estimation of $a(.)$ in the stochastic differential equation

$$\mathrm{d}X(t) = a(t)\, X(t)\, \mathrm{d}t + \mathrm{d}W(t), \quad X(0) = X, \quad t \geq 0 \tag{3.28}$$

where

$$a(.) \in L^2([0, T])$$

for every $T > 0$.

Discrete sampling

In the following discussion, we shall assume that $\sigma(.)\equiv 1$. We shall now describe estimators of $a(.)$ when a sampled version of the process $\{X_t,\ t\geqq 0\}$, say, $\{X_i,\ 1\leqq i\leqq n\}$ where $X_i=X_{t_i};\ 1\leqq i\leqq n$ is available. In analogy with the estimator suggested in (3.14), one can estimate $a(x)$ by

$$a_n(x)=\left[\frac{1}{2}\sum_{i=1}^{n}\frac{1}{h_i}K_2^{(1)}\left(\frac{x-X_i}{h_i}\right)\right]\left[\varepsilon+\sum_{i=1}^{n}K_1\left(\frac{x-X_i}{h_i}\right)\right]^{-1} \tag{3.29}$$

when the discrete set of observation $X_i,\ 1\leqq i\leqq n$ are available where the kernels K_1 and K_2 are bounded densities on R, $\{h_i,\ i\geqq 1\}$ a decreasing sequence of positive numbers such that $h_i\downarrow 0$ as $i\to\infty$ and ε is a fixed positive number. BANON (1977) studied the weak consistency and asymptotic normality of such estimators. For simulation purposes, one can use

$$\begin{aligned} &t_i=iT, \quad T>0\\ &h_i=h_1 i^{-\beta} \quad \text{where} \quad h_1>0 \quad \text{and} \quad \beta=\frac{1}{5},\\ &K_1(x)=x \quad \text{for} \quad -1<x\leqq 1\\ &\qquad\ =0 \quad \text{otherwise}\end{aligned}$$

and

$$\begin{aligned} &K_2(x)=2\ (1-|x|) \quad \text{for} \quad -1<x\leqq 1\\ &\qquad\ =0 \qquad\qquad\quad \text{otherwise}\end{aligned}$$

Let $a_{n1}(x)$ be the estimator so obtained.

Since the drift $a(x)$ can be obtained as a limit of conditional expectation i.e.,

$$a(x)=\lim_{t\to 0}\frac{1}{t}\,\mathsf{E}\,(X_t-X_0\mid X_0=x), \tag{3.30}$$

another estimator of $a(x)$ at x for which $p(x)>0$ can be chosen as

$$\hat{a}_n(x)=\left[\sum_{i=1}^{n}(X_{i+1}-X_i)\,K\left(\frac{x-X_i}{h_i}\right)\right]\left[\varepsilon+\sum_{i=1}^{n}\tau_i K\left(\frac{x-X_i}{h_i}\right)\right]^{-1} \tag{3.31}$$

where K is a square integrable kernel with bounded support, $0<\tau_i\downarrow 0$ as $i\to\infty$, $t_i=\sum_{j=1}^{i}\tau_j$ and $h_i=h(t_i)$ where $h:\mathrm{R}^+\to\mathrm{R}^+$ and $h(t)\downarrow 0$ as $t\to\infty$ and $\varepsilon>0$. This estimator was suggested by NGUYEN and PHAM (1981). They showed that the estimator is asymptotically normal under some conditions. For simulation purposes, one can take

$$\tau_i=\tau_1 i^{-\alpha}, \quad h(t)=h_1 t^{-\beta}\,.$$

Suppose we choose $\alpha=\frac{2}{5}$ and $\beta=\frac{1}{3}$ for comparison purposes. Let $\hat{a}_{n2}(x)$ be the estimator obtained for these specific parameters.

An estimator based on nonparametric estimation of regression function

$$\mathsf{E}\,\{X_{t+\Delta}\mid X_t=x\} \tag{3.32}$$

given a MARKOV sequence sampled from stationary MARKOV process is suggested by PHAM (1981). We do not discuss the details here.

Noting that $a(x)$ is the limit of a conditions expectation, GEMAN (1979) suggested the estimator

$$\bar{a}_n(x)=\frac{1}{n}\sum_{i=1}^{n}\frac{1}{\tau_i}(X_i-x) \tag{3.33}$$

where $\tau_i \downarrow 0$ as $i\to\infty$ and $X_i=X_{t_i+\tau_i}$ with $\{t_i\}$, a sequence of stopping times such that

$$t_1=\inf\{t\geqq 0,\ X_t^1=x\},\qquad t_{i+1}=\inf\{t\geqq t_i+\tau_i,\ X_t=x\}\ .$$

He has proved that $\bar{a}_n(x)$ is consistent both in strong and weak sense and asymptotically normal under some conditions. Let $\tau_i=\tau_1 i^{-\alpha}$. The optimal choice is $\alpha=\frac{1}{3}$ in the sense of minimizing mean square error. Let $\bar{a}_{n3}(x)$ be the estimator obtained with $\alpha=\frac{1}{3}$.

BANON and NGUYEN (1981a, b) observed through simulation studies that the estimator $a_{n1}(x)$ gives the "best" results. If $\sigma(.)\equiv\sigma$ constant but unknown, then one can use either $\hat{a}_{n2}(x)$ or $\bar{a}_{n3}(x)$ but computation of $\bar{a}_{n3}(x)$ is expensive. For further details about simulation studies of the above estimators see, BANON and NGUYEN (1981a, b).

4. Remarks and Open Problems

The problem of estimation of the rate of convergence the distribution of either the MLE $\hat{\theta}_T$ or the MPE $\tilde{\theta}_T$ or that of the LSE $\hat{\theta}_{n,T}$ to normal distribution is of extreme interest. This problem is solved in the classical i.i.d. case by PFANZAGL (1971) and in PRAKASA RAO (1973) for discrete time stationary MARKOV processes for MLE. No results are known for continuous time processes. Investigation of the BERRY-ESSEEN type bound to obtain the exact rate of convergence in the central limit theorem for stochastic integrals is of independent probabilistic interest. Bounds on the difference $|\tilde{\theta}_T-\theta_T|$ of BE and MLE are of interest to determine the effect of prior on the estimator. Results of this type for discrete time stationary MARKOV processes were obtained in PRAKASA RAO (1979c). For other open problems in the parametric case, see section 2.

Let us now look at the nonparametric aspect of the problem. Consider the stochastic differential equation

$$dX_t=a(X_t)\,dt+dW_t,\quad t\geqq 0,\quad X_0=X\ . \tag{4.1}$$

Suppose it is known that $a(.)$ is monotone increasing. Given that the process is observed upto time T say, how to obtain the MLE of $a(.)$ subject to the condition that $a(.)$ is monotone increasing and what are its asymptotic properties? Results of this type in the classical case are given in PRAKASA RAO (1969, 1970), BARLOW et al. (1972) and GRENANDER (1956). Optimality properties of nonparametric estimators of drift is the subject of investigation in IBRAGIMOV and HASMINSKII (1981), Chapter 7. For a comprehensive survey of non-parametric functional estimation for stochastic processes, see PRAKASA RAO (1983b).

References

BANON, G. (1977). Estimation Non Parametrique de Densite de Probabilite pour les Processus de Markov. Thesis, Universite Paul Sabatier de Toulouse (Sciences).

BANON, G. (1978). Nonparametric identification for diffusion processes. *SIAM J. Control and Optim.* **16**, 380–395.

BANON, G. and NGUYEN, H. T. (1978). Sur l'estimation recurrence de la densite et de sa deriree pour un processus de Markov. *C. R. Acad. Sci. Paris. Ser A* **286**, 691–699.

BANON, G. and NGUYEN, H. T. (1981a). Three methods of nonparametric estimation of the drift term in the diffusion model. (Preprint)

BANON, G. and NGUYEN, H. T. (1981b). Simulation studies of recursive density estimation in dependent observations. (Preprint).

BARLOW, R. E., BARTHOLOMEW, D. J., BREMNER, J. M. and BRUNK, H. D. (1972). *Statistical Inference under Order Restrictions.* Wiley, New York.

BASAWA, I. V. and PRAKASA RAO, B. L. S. (1980a). *Statistical Inference for Stochastic Processes*, Academic Press, London.

BASAWA, I. V. and PRAKASA RAO, B. L. S. (1980b). Asymptotic inference for stochastic processes. *Stochastic Processes and its Appl.* **10**, 221–254.

BASU, A. (1983a). The Bernstein-Von Mises theorem for a certain class of diffusion processes. *Sankhyā, Ser. A*, **45**, 150–160.

BASU, A. (1983b). Asymptotic theory of estimation in non-linear stochastic differential equations for the multi-parameter case. *Sankhyā. Ser. A*, **45**, 56–65.

DOROGOVCEV, A. JA (1976). The consistency of an estimate of a parameter of a stochastic differential equation. *Theory Probab. and Math. Statist.* **10**, 73–82.

FRIEDMAN, A. (1975). *Stochastic Differential Equations and Applications.* Vol. 1, Academic Press, New York.

GEMAN, S. A. (1979). On a common sense estimator for the drift of a diffusion. Reports in Pattern Analysis No. 79. Division of Applied Math. Brown University.

GEMAN, S. A. and HWANG, C. (1979). Non parametric maximum likelihood estimation by the method of Sieves. Report on Pattern Analysis No. 80, Division of Applied Math. Brown University.

GIKHMAN, I. I. and SKOROKOD, A. V. (1972). *Stochastic Differential Equations.* Springer-Verlag, Berlin.

GRENANDER, U. (1956). On the theory of mortality measurement. Part-II. *Skand. Aktuarietidkskr.* **39**, 126–153.

GRENANDER, U. (1980). *Abstract Inference*, Wiley, New York.

IBRAGIMOV, I. A. and HASMINSKII, R. Z. (1981). *Statistical Estimation.* Springer-Verlag, Berlin.

KUTOYANTS, YU. A. (1984). *Parameter Estimation for Stochastic Processes.* (Translated from Russian and edited by B. L. S. PRAKASA RAO) Heldermann-Verlag, Berlin.

LE BRETON, A. (1975). On continuous and discrete sampling for parameter estimation in diffusion type processes. *Math. Prog. Studies* **5**, 124–144.

NGUYEN, H. T. (1981). Asymptotic normality of recursive density estimators in Markov processes. *Publ. Inst. de Stat.* **26**, 73–93.

NGUYEN, H. T. and PHAM DINH TUAN (1980). Identification of non stationary diffusion model by the method of Sieves. (Preprint).

NGUYEN, H. T. and PHAM DINH TUAN (1981). Nonparametric estimation in diffusion model by discrete sampling. *Publ. Inst. de. Stat.* **26**, 89–109.

PFANZAGL, J. (1971). The Berry-Esseen bound for minimum contrast estimates. *Metrika* **17**, 81–91.

PHAM DINH TUAN (1981). Nonparametric estimation of the drift coefficient in the diffusion equation. *Math. Operationsforsch. u. Statist., ser. statist.* **12**, 61–73.

PRAKASA RAO, B. L. S. (1969). Estimation of a unimodal density. *Sankhyā Ser. A* **31**, 23–36.

PRAKASA RAO, B. L. S. (1970). Estimation for distributions with monotone failure rate. *Ann. Math. Statist.* **41**, 507–519.

PRAKASA RAO, B. L. S. (1973). On the rate of convergence of estimators for Markov processes. *Z. Wahrscheinlichkeitstheorie und verw. Geb.* **26**, 141–152.

PRAKASA RAO, B. L. S. (1979a). Sequential nonparametric estimation of density via delta-sequences. *Sankhya Ser. A* **41**, 82–94.

PRAKASA RAO, B. L. S. (1979b). Nonparametric estimation for continuous time Markov processes via delta families. *Publ. Inst. de Stat.* **24**, 79–97.

PRAKASA RAO, B. L. S. (1979c). The equivalence between (modified) Bayes estimator and maximum likelihood estimator for Markov processes. *Ann. Inst. Statist. Math.* **31**, 499–513.

PRAKASA RAO, B. L. S. (1981). The Bernstein-von Mises theorem for a class of diffusion processes. *The Theory of Random Processes.* **9**, 95–101 (Russian).

PRAKASA RAO, B. L. S. (1984). On Bayes estimation for diffusion fields. In *Statistics: Application and New Directions.* Ed. T. K. GHOSH and T. ROY.) Statistical Publishing Society, Calcutta, 504–511.

PRAKASA RAO, B. L. S. (1982). Maximum probability estimation for diffusion processes. In *Statistics and Probability: Essays in Honour of C. R. Rao* (Ed. G. KALLIANPUR et al.), North Holland 575–590.

PRAKASA RAO, B. L. S. (1983a). Asymptotic theory for non-linear least squares estimators for diffusion processes. *Math. Operationsforsch. u. Statist. ser. statist.* **14**, 195–209.

PRAKASA RAO, B. L. S. (1983b), *Nonparametric Functional Estimation*, Academic Press, New York.

PRAKASA RAO, B. L. S. and RUBIN, H. (1981). Asymptotic theory of estimation in nonlinear stochastic differential equations. *Sankhyā Ser. A* **43**, 170–189.

RAO, C. R. (1973). *Linear Statistical Inference and its Applications.* Wiley, New York (Second Edition).

WEISS, L. and WOLFOWITZ, J. (1974). *Maximum Probability Estimators and Related Topics.* Lecture Notes in Mathematics. Springer-Verlag, Berlin.

WONG, E. (1971). *Stochastic Processes in Information and Dynamical Systems.* McGraw-Hill, New York.

WONG, E. and ZAKAI, M. (1965). The oscillation of stochastic integrals. *Z. Wahrscheinlichkeitstheorie und verw. Geb.* **4**, 103–112.

B. L. S. PRAKASA RAO
Indian Statistical Institute
7, S.J.S. Sansanwal Marg
New Delhi - 110016
India