Sharp Adaptive Estimation of the Trend Coefficient of an Ergodic Diffusion

Arnak S. Dalalyan

Humboldt Universität zu Berlin

The Model

Let $X = (X_t,\, t \geq 0)$ be a random process defined by SDE

$$dX_t = S(X_t)\, dt + dW_t, \tag{1}$$

where W_t is a standard Wiener process and $X_0 = \xi$ is a random variable independent of W. Let Σ_0 be the set of all functions $S(\cdot) \in \mathrm{C}^1$ such that

$$\overline{\lim_{|x|\to\infty}} \operatorname{sgn}(x)\, S(x) < 0, \quad \big|S(x)\big| \leq C(1+x)^{\nu}, \tag{2}$$

for some positive constants C and ν. Then the SDE (1) has a unique solution; in addition, this solution is ergodic.

For simplicity, we suppose that $X_0 = \xi$ follows the invariant law; the probability density of this law is given by

$$f_S(y) = \frac{1}{G(S)} \exp\left\{2\int_0^y S(v)\,\mathrm{d}v\right\}.$$

The statistical problem we are interested in:

- The observation is a continuous path x^T of X over $[0, T]$.
- The unknown function is the trend coefficient $S(\cdot)$.
- The function of interest is $S(\cdot)$.
- We are interested in the behavior of the estimators as $T \to \infty$.
- The quality of estimation is measured by $L^2(\mathbb{R}, f_S^2)$-risk:

$$R_T(\bar{S}_T, S) = \mathbf{E}_S \int_{\mathbb{R}} \left(\bar{S}_T(x) - S(x)\right)^2 f_S^2(x)\,dx.$$

Historical Background

- Pham, T. D. (1981), Prakasa Rao, B. L. S. (1990), Van Zanten, J. H. (2001) studied the rate of convergence of a kernel estimator and its asymptotic normality. If $S(\cdot) \in H^{\beta}$, then the (optimal) rate of convergence is proved to be $T^{\frac{2\beta}{2\beta+1}}$.
- Spokoiny, V. G. (2000) constructed an adaptive, almost rate optimal estimator of the trend coefficient via locally linear approximation of log-likelihood.
- Galtchouk L. and Pergamenshchikov S. (2001_a, 2001_b) considered the problem of trend estimation, when the diffusion is observed up to a stopping time.

Local Minimax Risk

1. The parameter space. Let $\Sigma(k) = \Sigma_0 \cap C^k$ for any $k \in \mathbb{N}$, be the set of all k times differentiable trend coefficients satisfying condition (2). We fix an $S_0 \in \Sigma(k)$, $k \in \mathbb{N}^*$, $R > 0$ and define

$$\Sigma_\delta = \left\{ S \in V_\delta(S_0) : \left\| (S^{(k)} - S_0^{(k)}) f_S \right\|_2^2 \le R \right\}$$

Thus $\Sigma_\delta = \Sigma_\delta(S_0, k, R)$.

2. The local risk. Let $\bar{S}_T(\cdot) = \bar{S}_T(\cdot, x^T)$ be an estimator of the trend coefficient $S(\cdot)$, then

$$R_T(\bar{S}_T, \Sigma_\delta) = \sup_{S \in \Sigma_\delta} \mathbf{E}_S \left\| (\bar{S}_T - S) f_S \right\|_2^2 .$$

3. The minimax approach. We study the minimax risk

$$r_T(\Sigma_\delta) = \inf_{\bar{S}_T} \sup_{S \in \Sigma_\delta} R_T(\bar{S}_T, S).$$

The asymptotic behavior of this quantity is described by following

Theorem 1 (Dalalyan, A. S. & Kutoyants, Yu. A. (2001))**.** *Let $k \geq 1$ and the order of smoothness of the density f_0 corresponding to the central function S_0 be $> k+1$, then*

$$\lim_{\delta \to 0} \lim_{T \to \infty} T^{\frac{2k}{2k+1}} r_T(\Sigma_\delta) = P(k, R),$$

where $P(k, R)$ is the Pinsker constant (Pinsker, M. S. (1980)):

$$P(k, R) = (2k+1)\left(\frac{k}{\pi(k+1)(2k+1)}\right)^{\frac{2k}{2k+1}} R^{\frac{1}{2k+1}}.$$

Construction of Estimator

Let K_T be a smooth approximation of the Dirac measure at zero δ_0. (e. g. $K_T(x) = h_T^{-1} K(x/h_T)$).

- A natural estimator of the distribution function is

$$\hat{F}_T(x) = \frac{1}{T} \int_0^T \mathbb{1}_{\{X_t \leq x\}} dt.$$

- A natural estimator of the invariant density is the convolution

$$f_{K,T}(x) = (K * \hat{F}_T)(x) = \frac{1}{T} \int_0^T K_T(x - X_t)\, dt.$$

- A natural estimator of f_S' is the convolution

$$f_{K,T}^{(1)}(x) = (K' * \hat{F}_T)(x) = \frac{1}{T} \int_0^T K_T'(x - X_t)\, dt.$$

Using the explicit form of the invariant density, we get

$$S(x) = \frac{f_S'(x)}{2f_S(x)} \ .$$

It provides a natural way of construction of an estimator:

$$\bar{S}_T(x) = \frac{\text{estimator of } \ f_S'(x)}{2 \times \text{estimator of } \ f_S(x)} = \frac{f_{K,T}^{(1)}(x)}{2f_{K,T}(x)} \ .$$

The problem with this estimator is that at some points the denominator can be equal to zero while the numerator is $\neq 0$. We avoid it using the following modified estimator

$$\bar{S}_{K,T}(x) = \frac{f_{K,T}^{(1)}(x)}{2f_{K,T}(x) + \nu_T(x)} \ ,$$

where $\nu_T(x) = \varepsilon_T e^{-l_T|x|}$ with $l_T = \frac{1}{(\log T)}$ and $\varepsilon_T = T^{\frac{1}{\sqrt{\log T}} - \frac{1}{2}}$.

Theorem 2 (Dalalyan & Kutoyants, 2001). *For any symmetric non-negative smooth function* $K(\cdot)$ *satisfying* $\int_{\mathbb{R}} K(u)du = 1$ *and for any* $S \in \Sigma_\delta$, *we have*

$$R_T(\bar{S}_{K,T}, S) \sim \frac{1}{4}\,\mathbf{E}_S \int_{\mathbb{R}} \left(f_{K,T}^{(1)}(x) - f_S'(x)\right)^2 dx,$$

as $T \to \infty$.

For any function $h \in \mathbb{L}^2(\mathbb{R})$, we define the Fourier transform

$$\hat{h}(\lambda) = \frac{1}{2\pi} \int_{\mathbb{R}} e^{i\lambda x} h(x)\, dx.$$

Thus $\hat{K}_T$ and $\hat{f}_S$ are the Fourier transforms of the kernel K_T and the invariant density f_S. We set also

$$\hat{\varphi}_T(\lambda) = \frac{1}{T} \int_0^T e^{i\lambda X_t} dt.$$

The choice of the minimax kernel. The Parseval identity yields

$$R_T(\bar{S}_{K,T}, S) \sim \frac{1}{8\pi}\,\mathbf{E}_S \int_{\mathbb{R}} \lambda^2 \left|\hat{K}_T(\lambda)\hat{\varphi}_T(\lambda) - \hat{f}_S(\lambda)\right|^2 d\lambda$$

Using the fact that $\hat{\varphi}_T$ is an unbiased estimate of $\hat{f}_S$ and the relation

$$|\lambda|^2 \mathbf{Var}_S[\hat{\varphi}_T(\lambda)] \sim 4T^{-1}$$

we obtain that the risk $R_T(\bar{S}_{K,T}, S)$ is equivalent to

$$\Delta_T(\hat{K}, |\hat{f}|) = \frac{1}{8\pi} \int_{\mathbb{R}} \lambda^2 \left|\hat{K}_T(\lambda) - 1\right|^2 \left|\hat{f}_S(\lambda)\right|^2 d\lambda + \frac{1}{2\pi T} \int_{\mathbb{R}} \left|\hat{K}_T(\lambda)\right|^2 d\lambda$$

In the same time, our conditions imply that

$$\int_{\mathbb{R}} |\lambda|^{2k+2} \left|\hat{f}(\lambda) - \hat{f}_0(\lambda)\right|^2 d\lambda \le 8\pi R.$$

The functional Δ_T has a saddle point, which provides the optimal (minimax) kernel

$$\hat{K}_T^*(\lambda) = \left(1 - \left|\lambda \alpha_T^*\right|^k\right)_+$$

with optimal bandwidth

$$\alpha_T^* = T^{-\frac{1}{2k+1}} \left(\frac{4k}{\pi R (k+1)(2k+1)} \right)^{\frac{1}{2k+1}} .$$

The estimator of the trend coefficient S constructed via this kernel K_T^* is asymptotically optimal, but it can not be realised if we do not know the smoothness order of the unknown function S.

Our aim is now to construct an adaptive estimator with respect to parameters k and R. It will be done using the method developed by Golubev, G. (1992) and recently used in Cavalier L., Golubev G., Picard D. and Tsybakov A. (2002).

Sharp Adaptive Estimator

The main idea is to replace the function

$$\hat{K}^*_T(\lambda) = \left(1 - \left|\lambda \alpha^*_T\right|^k\right)_+$$

by a random function

$$\tilde{K}_T(\lambda) = \left(1 - \left|\lambda \tilde{\alpha}\right|^{\tilde{\beta}}\right)_+,$$

where $\tilde{\alpha}$ and $\tilde{\beta}$ are data driven (depend on the observation X^T). In order to do it, we define

$$h_{\alpha,\beta}(\lambda) = \left(1 - \left|\lambda \alpha\right|^{\beta}\right)_+$$

$$\mathcal{H}_T = \left\{ h_{\alpha,\beta} \ : \ \alpha \in [T^{-1}, (\log T)^{-1}]; \ \beta \geq 0.5 \right\}.$$

Recall that

$$R_T(\bar{S}_{K,T}, S) \sim \Delta_T\big(\hat{K}_T, |\hat{f}_S|\big).$$

where

$$\Delta_T(h, |\hat{f}|^2) = \frac{1}{8\pi}\int_{\mathbb{R}} |\lambda|^2\big(1 - h(\lambda)\big)^2\big|\hat{f}(\lambda)\big|^2 d\lambda + \frac{1}{2\pi T}\int_{\mathbb{R}} h^2(\lambda)\, d\lambda.$$

In order to choose the values α and β in an adaptive way, we should

- define a good estimator $l_T(h)$ of the functional $\Delta_T(h, \hat{f})$,
- minimise the (random) functional $l_T(h)$ over a suitably chosen finite subset of $\mathcal{H}_T$.

Since Δ_T is a quadratic functional, its estimation by plug-in method is not good. That is why we define

$$l_T(h) = \Delta_T\left(h, \big|\hat{\varphi}_T(\lambda)\big|^2 - \frac{4}{T|\lambda|^2}\right)$$

Theorem 3. *Let the function* S_0 *be such that at an instant* t_0 *the transition density* $p_{t_0}(x,y)$ *is bounded in both variables. Then there exists a subset* $\mathcal{H}'_T$ *of* $\mathcal{H}_T$ *containing* $[\log T]$ *elements, such that the estimator*

$$\tilde{f}_T^{(1)}(x) = \frac{1}{2\pi T}\int_0^T \left[\int_{\mathbb{R}} \lambda \sin\big(\lambda(x - X_t)\big)\tilde{h}_T(\lambda)d\lambda\right]dt$$

where $\tilde{h}_T(\lambda) = \min_{h\in\mathcal{H}'} l_T(h)$, *is a minimax sharp adaptive in the problem of derivative* f'_S *estimation. Therefore,*

$$\tilde{S}_T(x) = \frac{\tilde{f}_T^{(1)}(x)}{2f_{K,T}(x) + \nu_T(x)}$$

is a sharp adaptive estimator of the trend coefficient, i.e.

$$\lim_{\delta\to 0}\lim_{T\to\infty} T^{\frac{2k}{2k+1}} R_T(\tilde{S}_T, \Sigma_\delta) = P(k,R).$$

Concluding Remarks

1°. If the function S is Hölder continuous and satisfies condition (2), then the transition density is bounded at any instant t (cf. Veretennikov, A. (1999)).

2°. In the case where the diffusion coefficient is not identically one, it holds

$$S(x) = \frac{(\sigma^2(x) f_S(x))'}{2 f_S(x)} \ .$$

That is why the extension of the described method to this case is straightforward.

3°. This result can be easily globalised, provided that the conditions are satisfied uniformly on the parameter set.

$$\alpha_i = (1 + \log^{-1} T)^i,$$

$$\beta_i = \left(1 - \frac{i}{\log T}\right)^{-1}.$$