SORT 28 (1) January-June 2004, 1-8

On-line nonparametric estimation⋆

R. Khasminskii*

Wayne State University, USA

Abstract

A survey of some recent results on nonparametric on-line estimation is presented. The first result deals with an on-line estimation for a smooth signal *S(t)* in the classic 'signal plus Gaussian white noise' model. Then an analogous on-line estimator for the regression estimation problem with equidistant design is described and justified. Finally some preliminary results related to the on-line estimation for the diffusion observed process are described.

MSC: 62G05, 62G08, 62M05.

Keywords: White Gaussian Noise, on-line estimation, nonlinear filter.

1 On-line estimation of a signal in a Gaussian white noise model

We consider an observation process $X^\varepsilon(t)$ having the form

$$X^\varepsilon(t) = \int_0^t S(s)ds + \varepsilon W(t),\ t \in [0,1]. \tag{1.1}$$

Here $W(t)$ is a standard Wiener process and $\varepsilon > 0$ is a small parameter. Denote by $\Sigma(\beta, L)$ a class of functions $S(t), t \in [0,T]$ having k derivatives on $(0,T)$ with k-th derivative $S^{(k)}(t)$ satisfying the Hölder condition with the exponent $\alpha \in (0,1] (\beta = k + \alpha)$:

$$|S^{(k)}(t+h) - S^{(k)}(t)| \leq L|h|^\alpha.$$

* *Address for correspondence:* Department of Mathematics, Wayne State University, Detroit, MI 48202, USA. E-mail: rafail@math.wayne.edu

⋆ Partially supported by NSF Grant DMS 9971608.

Received: October 2003

Accepted: January 2004

The following problem was considered by Ibragimov and Khasminskii (1981): what is the rate of convergence to 0 for the best estimators of $S, S^{(1)}, \ldots, S^{(k)}$, as $\varepsilon \to 0$, and how can estimators with this rate be created? It was shown in Ibragimov and Khasminskii (1980a, 1981) that the kernel and projection estimators $\widehat{S}_\varepsilon(t)$ for a suitable choice of parameters have a property

$$\sup_{S\in\Sigma(\beta,L)} E\Big(\Big[\frac{\widehat{S}_\varepsilon(t)-S(t)}{\varepsilon^{2\beta/(2\beta+1)}}\Big]^2+\sum_{j=1}^{k}\Big[\frac{\widehat{S}^{(j)}_\varepsilon(t)-S^{(j)}(t)}{\varepsilon^{2(\beta-j)/(2\beta+1)}}\Big]^2\Big)\le C, \tag{1.2}$$

and there are no estimators with uniformly in $\Sigma(\beta, L)$ better rate of convergence to 0 risks (here and below we denote by C, C_i generic positive constants, which may be different and do not depend on ε).

In some applications it is necessary to create a tracking (or on-line) type of estimator for S, that is estimators with the property: $\widehat{S}_\varepsilon(t+h)$ is based on $\widehat{S}_\varepsilon(t)$ and observation process on the time interval $[t, t+h]$ only. Unfortunately the well-known kernel and projection estimators do not have this property.

The tracking estimator for the model (1.1) was proposed by Chow *et al.* (1997). This estimator has the structure of a Kalman filter. Heuristically this estimator is based on the auxiliary filtering model

$$\begin{aligned} dS(t) &= S^{(1)}(t)dt \\ dS^{(j)}(t) &= S^{(j+1)}(t)dt,\ \ j=1,\ldots,k-1 \\ dS^{(k)}(t) &= \sigma_\varepsilon dW'(t) \\ dX_t &= S(t)dt+\varepsilon dW(t). \end{aligned} \tag{1.3}$$

It is clear that the last equation in (1.3) is equivalent to (1.1). Assuming that the standard Wiener processes $W(t)$ and $W'(t)$ are independent and choosing a constant σ_ε in a suitable way, we arrive at the following estimator for $S(t) = S^{(0)}(t), \ldots, S^{(k)}(t)$ (see details in Chow *et al.* (1997))

$$\begin{aligned} d\widehat{S}^{(j)}_\varepsilon(t) &= \widehat{S}^{(j+1)}_\varepsilon(t)dt+\frac{q_j}{\varepsilon^{2(j+1)/(2\beta+1)}}(dX^\varepsilon(t)-\widehat{S}_\varepsilon(t)dt), \\ j &= 0,1,\ldots,k-1 \\ d\widehat{S}^{(k)}_\varepsilon(t) &= \frac{q_k}{\varepsilon^{2(k+1)/(2\beta+1)}}(dX^\varepsilon(t)-\widehat{S}_\varepsilon(t)dt), \end{aligned} \tag{1.4}$$

subject to the initial conditions $\widehat{S}(0) = S_\circ, \widehat{S}^{(j)}(0) = S^j_\circ$, $j = 1, \ldots, k$, which reflect a priori information on $S(0)$, $S^{(j)}(0)$, $j = 1, \ldots, k$.

Denote by $p_k(\lambda)$ the polynomial

$$p_k(\lambda)=\lambda^{k+1}+q_0\lambda^k+\ldots+q_{k-1}\lambda+q_k.$$

The following result was proven in Chow *et al.* (1997):

Theorem 1.1 *For any choice* $q_0, \ldots, q_k$ *such that all roots of the polynomial* $p_k(\lambda)$ *have negative real parts, and for arbitrary bounded initial conditions* $S_\circ, \ldots, S_\circ^k$ *the tracking filter (1.4) has the property: there exists an initial boundary layer* $\Delta_\varepsilon = C_1 \varepsilon^{2/2\beta+1} \log(1/\varepsilon)$ *such that for* $t \geq \Delta_\varepsilon$ *the inequality (1.2) is valid.*

Remark 1.2 *It is proven in Chow* et al. *(1997) also, that the initial boundary layer of the order* $\varepsilon^{2/2\beta+1}$ *is inevitable for any tracking type estimator.*

Remark 1.3 *An analogous result was proven in Chow* et al. *(2001) for the estimation of a time dependent spatial signal observed in a cylindrical Gaussian white noise model of the small intensity* ε*. It is proven in Chow* et al. *(2001) that outside of the inevitable boundary layer the symbiosis of a projection estimator in the space variables and tracking type estimator in the time variable also has an optimal rate of convergence of risks to* 0*, as* $\varepsilon \to 0$*, for a suitable choice parameters of a tracking filter and a projection estimator.*

2 On-line estimation of a smooth regression function

It is well known that the model (1.1) is a natural approximation for the regression estimation model with equidistant design. In more detail, consider the following statistical model. Let $f(t) \in \mathbb{R}^1$, $t \in [0,1]$, be a function from $\Sigma(\beta, L)$, $t_{in} = \frac{i}{n}$, $i = 1, \ldots, n$, and the observation model has the form

$$X_{in} = f(t_{in}) + \sigma(t_{in})\xi_{in}, \tag{2.1}$$

where $(\xi_{in})_{i \leq n}$ is a sequence of i.i.d. random variables with $E\xi_{in} = 0$, $E\xi_{in}^2 = 1$ and $\sigma^2(t_{in}) < C$. The natural analogy of an estimator (1.4) is the tracking estimator (hereafter we write for brevity t_i instead of t_{in} and X_i instead of X_{in})

$$\begin{aligned} \widehat{f_n^{(j)}}(t_i) &= \widehat{f_n^{(j)}}(t_{i-1}) + \frac{1}{n}\widehat{f_n^{(j+1)}}(t_{i-1}) + \frac{q_j}{n^{\frac{(2\beta-j)}{2\beta+1}}}(X_i - \widehat{f_n^{(0)}}(t_{i-1})) \\ j &= 0, 1, \ldots, k-1 \\ \widehat{f_n^{(k)}}(t_i) &= \widehat{f_n^{(k)}}(t_{i-1}) + \frac{q_k}{n^{\frac{(2\beta-k)}{2\beta+1}}}(X_i - \widehat{f_n^{(0)}}(t_{i-1})) \end{aligned} \tag{2.2}$$

subject to some initial conditions $\widehat{f_n^{(0)}}(0), \widehat{f_n^{(1)}}(0), \ldots, \widehat{f_n^{(k)}}(0)$.The following theorem, analogous to Theorem 1.1, was proven by Khasminskii and Liptser (2002):

Theorem 2.1 *Let* $q_0, \ldots, q_k$ *are chosen so that all roots of the polynomial* $p_k(\lambda)$ *have negative real parts. Let an observation model has the form (2.1),* $f \in \Sigma(\beta, L)$ *and* $\sigma^2(t) < C$*. Then the estimator (2.2) with arbitrary bounded initial conditions*

$\widehat{f_n^{(0)}}(0), \widehat{f_n^{(1)}}(0), \ldots, \widehat{f_n^{(k)}}(0)$ *possesses the property: for* $t_l > C_1 n^{-\frac{1}{2\beta+1}} \log n := \delta_n$

$$\sup_{f\in\Sigma(\beta,L)} \sum_{j=0}^{k} E(f^{(j)}(t_\ell) - \widehat{f_n^{(j)}}(t_\ell))^2 n^{\frac{2(\beta-j)}{2\beta+1}} \le C_2. \tag{2.3}$$

Remark 2.2 *Similar the proof an analogous property of the estimator (1.4) it is easy to conclude from the results in Stone (1980) and Ibragimov and Khasminskii (1980b) that the rate of convergence of risks to zero for* $n \to \infty$ *in (2.3) is unimprovable. The boundary layer of order* $n^{-\frac{1}{2\beta+1}}$ *is also inevitable for any on-line estimator.*

Remark 2.3 *It is easy to apply the estimator (2.2) for the estimation* f *with the best rate of convergence of risk to* 0 *for* all $t \in [\delta_n, 1]$. *It is enough to set, for instance,* $\widehat{f_n^{(j)}}(t) = \widehat{f_n^{(j)}}(t_\ell)$ *for* $t_l \le t < t_{l+1}$.

Proof of Theorems 1.1 and 2.1 are similar. Making use of the choice parameters $q_0, \ldots, q_k$ and the recursive form of estimators (1.4), (2.2) one can find the suitable upper bounds for the bias and variance of these estimators. As an illustration, consider the simplest case of the estimation problem (2.1) with $f \in \Sigma(1, L)$ ($\beta = 1$). Then the estimator (2.2) takes the form

$$\widehat{f_n}(t_\ell) = \widehat{f_n}(t_{\ell-1}) + \frac{q_0}{n^{2/3}}(X_\ell - \widehat{f_n}(t_{\ell-1})); \; \widehat{f_n}(0) = f_0 \tag{2.4}$$

with arbitrary bounded f_0 and positive bounded q_0. Making use of (2.1) and notations $\Delta_n(\ell) = \widehat{f_n}(t_\ell) - f(t_\ell)$, $\Delta f(t_\ell) = f(t_{\ell+1}) - f(t_\ell)$, one can rewrite (2.4) as

$$\Delta_n(\ell) = (1 - \frac{q_0}{n^{2/3}})(\Delta_n(\ell-1) - (1 - \frac{q_0}{n^{2/3}})\Delta f(t_{\ell-1}) + \frac{q_0\sigma(t_\ell)\xi_\ell}{n^{2/3}}. \tag{2.5}$$

It follows from (2.5) that

$$\Delta_n(\ell) = (1 - \frac{q_0}{n^{2/3}})^\ell \Delta_n(0) - \sum_{i=0}^{\ell-1}(1 - \frac{q_0}{n^{2/3}})^{\ell-i}\Delta f(t_i) \tag{2.6}$$
$$+ \frac{q_0}{n^{2/3}} \sum_{i=0}^{\ell-1}(1 - \frac{q_0}{n^{2/3}})^{\ell-i}\sigma(t_i)\xi_i.$$

It follows from the assumption $\beta = 1$ that $|\Delta f(t_i)| \le L/n$. Thus we have from (2.6) that

$$|E\Delta_n(\ell)| \le |\Delta_n(0)| \exp\{-\frac{q_0\ell}{n^{2/3}}\} + Cn^{-1/3}.$$

So $|E\Delta_n(\ell)| \le Cn^{-1/3}$, as $\ell \ge C_1 n^{2/3} \log n$, or, equivalently, as $t_\ell = l/n \ge C_1 n^{-1/3} \log n$. Analogously one can obtain

$$Var\Delta_n(\ell) \le \frac{C}{n^{4/3}} \sum_{i=0}^{\ell-1}(1 - \frac{q_0}{n^{2/3}})^{2(\ell-i)} \le Cn^{-2/3}.$$

These upper bounds for $|E\Delta_n(\ell)|$, $Var\Delta_n(\ell)$ imply the assertion of the Theorem 2.1 for the case $\beta = 1$. □

3 On-line estimation for the diffusion observed process

Recently we started (together with Y. Golubev) to study the problem of on-line estimation of an unknown signal $S(t)$ for the case of a diffusion observed process. A preliminary result concerns estimating a signal of the smoothness $\alpha, 0 < \alpha \leq 1$ only.

Assume that an observed process is a solution of the stochastic differential equation on $\mathbb{R}^1$

$$dX_\varepsilon(t) = F(t, X_\varepsilon(t), S(t))dt + \varepsilon\sigma(t, X_\varepsilon(t))dw(t);\ X_\varepsilon(0) = x_0. \tag{3.1}$$

(It is a natural generalization of an observation model (1.1)) Here $S(t) : \mathbb{R}^1 \mapsto \mathbb{R}^1$ is an unknown function, and the problem is to estimate this function on the interval $(0, T)$ making use of $X_\varepsilon(t), 0 \leq t \leq T$. Let the following conditions hold:

A1. The functions F, σ are Lipschitzian with respect to all variables, and σ is bounded.

A2. The function $S(t)$ satisfies the Hölder condition

$$|S(t + h) - S(t)| \leq L|h|^\alpha,\ 0 < \alpha \leq 1.$$

A3. For some positive C_i and all $0 \leq t \leq T, x \in \mathbb{R}^1, S \in \mathbb{R}^1$ the inequality $C_1 \leq |\frac{\partial F(t,x,S)}{\partial S}| \leq C_2$ holds.

We consider the following on-line estimator $S_\varepsilon(t)$

$$dS_\varepsilon(t) = \frac{dX_\varepsilon(t) - F(t, X_\varepsilon(t), S_\varepsilon(t))dt}{\gamma_\varepsilon F'_S(t, X_\varepsilon(t), S_\varepsilon(t))};\ S_\varepsilon(0) = S^{(0)}. \tag{3.2}$$

Theorem 3.1 *Under conditions* $A1 - A3$ *the estimator (3.2) with* $\gamma_\varepsilon = k\varepsilon^{\frac{2}{2\alpha+1}}$ *(k is an arbitrary positive constant) has the property*

$$E|S_\varepsilon(t) - S(t)|^2 \leq C\varepsilon^{\frac{4\alpha}{2\alpha+1}} \tag{3.3}$$

as $t > C\varepsilon^{\frac{2}{2\alpha+1}} \log(1/\varepsilon)$ *(here C is large enough, but independent of* ε*).*

Proof. Introduce $S_\delta(t) = (2\delta)^{-1} \int_{\mathbb{R}^1} \exp\{-\frac{|t-u|}{\delta}\}S(u)du$. It is easy to see from A2 that

$$|S_\delta(t) - S(t)| \leq c_3\delta^\alpha,\ |S'_\delta(t)| \leq c_3\delta^{\alpha-1}. \tag{3.4}$$

Introduce a new process $x_\delta(t) = S_\varepsilon(t) - S_\delta(t)$. Then we have from (3.1) and (3.2)

$$dx_\delta(t) = \frac{1}{\gamma_\varepsilon}\Delta_\varepsilon(t)dt + \frac{\varepsilon\sigma(t, X_\varepsilon(t))}{\gamma_\varepsilon F'_S(t, X_\varepsilon(t), S_\varepsilon(t))}dw(t) - S'_\delta(t)dt. \tag{3.5}$$

Here we denote

$$\Delta_\varepsilon(t) = \frac{F(t, X_\varepsilon(t), S(t)) - F(t, X_\varepsilon(t), S_\delta(t) + x_\delta(t))}{F'_S(t, X_\varepsilon(t), S_\delta(t) + x_\delta(t))} \tag{3.6}$$

The equation (3.5) and Ito formula imply

$$\begin{aligned} d[x_\delta(t)]^2 &= \frac{2}{\gamma_\varepsilon} x_\delta(t)\Delta_\varepsilon(t)dt + 2\frac{\varepsilon x_\delta(t)\sigma(t, X_\varepsilon(t))}{\gamma_\varepsilon F'_S(t, X_\varepsilon(t), S_\varepsilon(t))} dw(t) \\ &+ [\frac{\varepsilon\sigma(t, X_\varepsilon(t))}{\gamma_\varepsilon F'_S(t, X_\varepsilon(t), S_\varepsilon(t))}]^2 dt - 2x_\delta(t)S'_\delta(t)dt. \end{aligned} \tag{3.7}$$

It follows from A3 and (3.4) that

$$\begin{aligned} x_\delta(t)\Delta_\varepsilon(t) &= x_\delta(t)\frac{F(t, X_\varepsilon(t), S(t)) - F(t, X_\varepsilon(t), S_\delta(t))}{F'_S(t, X_\varepsilon(t), S_\delta(t) + x_\delta(t))} \\ &+ x_\delta(t)\frac{F(t, X_\varepsilon(t), S_\delta(t)) - F(t, X_\varepsilon(t), S_\delta(t) + x_\delta(t))}{F'_S(t, X_\varepsilon(t), S_\delta(t) + x_\delta(t))} \\ &\le \frac{C_2}{C_1}|x_\delta(t)||S(t) - S_\delta(t)| - \frac{C_1}{C_2}|x_\delta(t)|^2 \le C_4|x_\delta(t)|\delta^\alpha - \frac{C_1}{C_2}|x_\delta(t)|^2. \end{aligned} \tag{3.8}$$

Denote $V_\delta(t) = E[x_\delta(t)]^2$. Then it is clear from (3.7), (3.4) and (3.8) that

$$V'_\delta(t) \le -\frac{k_1}{\gamma_\varepsilon} V_\delta(t) + k_2(\frac{\delta^{2\alpha}}{\gamma_\varepsilon} + \frac{\varepsilon^2}{\gamma_\varepsilon^2} + \gamma_\varepsilon \delta^{2\alpha-2}) \tag{3.9}$$

for small enough positive constant k_1 and large enough constant k_2 (both independent of $\varepsilon, \delta, \gamma_\varepsilon$). Now choose the parameters $\delta, \gamma_\varepsilon$ as $\delta \asymp \varepsilon^{\frac{2}{2\alpha+1}}$, $\gamma_\varepsilon \asymp \varepsilon^{\frac{2}{2\alpha+1}}$. Then we obtain from (3.9) for some positive constants k_3, k_4 independent of ε the inequality

$$V'_\delta(t) \le -k_3\varepsilon^{-\frac{2}{2\alpha+1}} V_\delta(t) + k_4\varepsilon^{\frac{4\alpha-2}{2\alpha+1}} \tag{3.10}$$

It follows from (3.10) that

$$V_\delta(t) \le V_\delta(0)\exp\{-k_3\varepsilon^{-\frac{2}{2\alpha+1}t}\} + \frac{k_4}{k_3}\varepsilon^{\frac{4\alpha}{2\alpha+1}}. \tag{3.11}$$

The initial value $V_\delta(0) = S^{(0)} - S_\delta(0)$ is bounded. Thus we can conclude from (3.11) that $V_\delta(t) \le C\varepsilon^{\frac{4\alpha}{2\alpha+1}}$, as $t > C\varepsilon^{\frac{2}{2\alpha+1}}\log(1/\varepsilon)$. Note now that $E|S_\varepsilon(t) - S(t)|^2 \le 2V_\delta(t) + 2|S_\delta(t) - S(t)|^2$. The theorem follows from these upper bounds and (3.4). □

Remark 3.1 *It follows from the s.1 that the rate of convergence in Theorem 3.1 is unimprovable: it is unimprovable even for the case* $F(t, x, S) = S, \sigma(t, x) = 1$.

5 Concluding remark

The estimators (1.4), (2.2), (3.2) can be used for extrapolation too. For instance, the expression

$$\overline{f_n(t_l + h)} = \sum_{j=0}^{k} \frac{h^j}{j!} \widehat{f_n}(t_l)$$

can be used for estimation of $f(t_l + h)$ on the basis of observations $X_{1n}, \ldots, X_{ln}$. It is not hard to check with help of Theorem 2.1 that

$$E|\overline{f_n(t_l + h)} - f(t_l + h)|^2 \leq C \max\{h, n^{-\frac{1}{2\beta+1}}\}^{2\beta},$$

and better rate of convergence of risk is unattainable.

References

Chow, P.-L., Khasminskii, R. & Liptser, R.Sh. (1997). Tracking of signal and its derivatives in Gaussian white noise, *Stochastic Processes Appl.*, 69, 2 259-273.

Chow, P.-L., Khasminskii, R. & Liptser, R.Sh. (2001). On estimation of time dependent spatial signal in Gaussian white noise, *Stochastic Processes Appl.*, 96, 161-175.

Ibragimov, I. & Khasminskii, R. (1980a). Estimation of a signal, its derivatives, and point of maximum for Gaussian observations, *Theory Probability Appl.*, 25, 703-720.

Ibragimov, I. & Khasminskii, R. (1980b). On nonparametric estimation of regression, *Soviet Math. Dokl.*, 21, 810-814.

Ibragimov, I. & Khasminskii, R. (1981). *Statistical estimation: Asymptotic Theory.* Springer Verlag (Russian ed. 1979).

Khasminskii, R. & Liptser, R. (2002). Tracking of a smooth regression function, *Theory Probability Appl.*, 47, 3, 567-575.

Stone, C. (1980). Optimal rates of convergence for nonparametric estimators, *Ann. Statist.*, 8, 1348-1360.

Resum

Es presenta un recull d'alguns resultats recents en estimació no-paramètrica en línia. El primer resultat tracta d'una estimació en línia per a un senyal suau S(t) en el model clàssic "senyal més soroll blanc Gaussià (GWN)". Aleshores es descriu i justifica un estimador en línia anàleg pel problema d'estimació de regressió amb disseny equidistant. Finalment, es descriuen alguns resultats preliminars en relació a l'estimació en línia pel procés de difusió observada.

MSC: 62G05, 62G08, 62M05.

Paraules clau: soroll blanc Gaussià, estimació en línia, filtres no linials