On Delay Estimation and Testing for Stochastic Differential Equations

Yu. A. Kutoyants

Laboratoire de Statistique & Processus,

Université du Maine,

72085 Le Mans, Cédex 9,

FRANCE

Le Mans, January 8, 2005

Models

We consider the problems of estimation and hypotheses testing concerning the **delay parameter** ϑ by observations $X = \{X_t, 0 \leq t \leq T\}$ of the solution of stochastic differential equation

$$\mathrm{d}X_t = S(X_{t-\vartheta})\,\mathrm{d}t + \sigma(X_t)\,\mathrm{d}W_t, \quad X_0, \quad t \geq 0$$

in two quite different situations. The first one corresponds to dynamical system with small noise ($\sigma \to 0$)

$$\mathrm{d}X_t = -\gamma\, X_{t-\vartheta}\,\mathrm{d}t + \sigma\,\mathrm{d}W_t, \quad X_0, \quad t \geq 0$$

and the second to the linear ergodic SDE ($T \to \infty$)

$$\mathrm{d}X_t = -\gamma\, X_{t-\vartheta}\,\mathrm{d}t + \sigma\,\mathrm{d}W_t, \quad X_0, \quad t \geq 0.$$

Why **different**?

Regularity conditions and estimators

Let $X^n = \{X_1, \ldots, X_n\}$ be i.i.d. and X_j has density function $f(\vartheta, x)$. We have to estimate ϑ by observations X^n and describe the properties of estimators as $n \to \infty$. We have at least two situations: **regular** and **non regular**.

In **regular** situation the function $f(\vartheta, x)$ is differentiable w.r.t. ϑ and Fisher information $\mathrm{I}(\vartheta) < \infty$. Then for all estimators

$$\lim_{n\to\infty} n\mathbf{E}_\vartheta \left(\bar\vartheta_n - \vartheta\right)^2 \geq \mathrm{I}(\vartheta)^{-1}$$

Moreover the MLE $\hat\vartheta_n$ and Bayes estimators $\tilde\vartheta_n$ are consistent,

$$\sqrt{n}\left(\hat\vartheta_n - \vartheta\right) \Longrightarrow \mathcal{N}\left(0, \mathrm{I}(\vartheta)^{-1}\right), \quad \sqrt{n}\left(\tilde\vartheta_n - \vartheta\right) \Longrightarrow \mathcal{N}\left(0, \mathrm{I}(\vartheta)^{-1}\right)$$

and **both are asymptotically efficient**.

In **non regular** situation their behavior is different. Suppose that the density $f(\vartheta, x) = f(x - \vartheta)$, where the function $f(\cdot)$ has a jump $f(0+) - f(0-) \neq 0$. Then the MLE and BE have different limit distributions

$$n\left(\hat\vartheta_n - \vartheta\right) \Longrightarrow \xi, \quad n\left(\tilde\vartheta_n - \vartheta\right) \Longrightarrow \zeta$$

where ξ and ζ are some r.v. (functionals of Poisson process). Moreover, for all estimators

$$\lim_{n\to\infty} n^2 \mathbf{E}_\vartheta \left(\bar\vartheta_n - \vartheta\right)^2 \geq \mathbf{E}\zeta^2$$

and **asymptotically efficient are BE only**. We have $\mathbf{E}\xi^2 > \mathbf{E}\zeta^2$.

Question: The model

$$\mathrm{d}X_t = -\gamma\, X_{t-\vartheta}\,\mathrm{d}t + \sigma\,\mathrm{d}W_t, \quad X_0, \quad 0 \leq t \leq T$$

is **regular** or **non regular** for the problem of delay ϑ estimation?

Answer: yes regular if $\sigma \to 0$ and not non regular if $T \to \infty$.

Remind that

$$X_{t-\vartheta} = X_0 - \gamma \int_0^{t-\vartheta} X_{s-\vartheta}\,\mathrm{d}s + \sigma\, W_{t-\vartheta}.$$

Hence $X_{t-\vartheta}$ is *as smooth w.r.t.* ϑ *as Wiener process w.r.t. time*, i.e., **continuous but not differentiable**.

The MLE $\hat{\vartheta}_T$ and BE $\tilde{\vartheta}_T$ are defined by the equations

$$\sup_{\vartheta\in\Theta} L\left(\vartheta, X^T\right) = L\left(\hat{\vartheta}_T, X^T\right), \qquad \tilde{\vartheta}_T = \frac{\int_\Theta \vartheta p\left(\vartheta\right) L\left(\vartheta, X^T\right) {\rm d}\vartheta}{\int_\Theta p\left(\vartheta\right) L\left(\vartheta, X^T\right) {\rm d}\vartheta}$$

where the log-likelihood ratio

$$\ln L\left(\vartheta, X^T\right) = -\frac{\gamma}{\sigma^2}\int_0^T X_{t-\vartheta}\, {\rm d}X_t - \frac{\gamma^2}{2\sigma^2}\int_0^T X_{t-\vartheta}^2\, {\rm d}t$$

Note that

$$\frac{\partial}{\partial\vartheta}\ln L\left(\vartheta, X^T\right)$$

does not exist.

Estimation (small noise asymptotics)

Let

$$\mathrm{d}X_t = -\gamma\, X_{t-\vartheta}\,\mathrm{d}t + \sigma\,\mathrm{d}W_t, \quad X_s = x_s, s \leq 0, \quad 0 \leq t \leq T$$

and we study the properties of estimators as $\sigma \to 0$. This case of limit corresponds to **regular** statistical model of parameter estimation.
The following quantity plays the role of Fisher information:

$$\mathrm{I}\left(\vartheta\right) = \gamma^4 \int_0^T x_{t-2\vartheta}^2\,\mathrm{d}t$$

where $\frac{\mathrm{d}x_t}{\mathrm{d}t} = -\gamma\, x_{t-\vartheta}$.

The family of measures is LAN, i.e., the likelihood ratio process

$$Z_T\left(u\right) = \frac{\mathrm{d}\mathbf{P}^{(T)}_{\vartheta+\frac{u\sigma}{\sqrt{\mathrm{I}(\vartheta)}}}}{\mathrm{d}\mathbf{P}^{(T)}_{\vartheta}}\left(X^T\right) \Longrightarrow Z\left(u\right) = e^{u\,\zeta - \frac{u^2}{2}}, \quad \zeta \sim \mathcal{N}\left(0,1\right).$$

Then for all estimators

$$\lim_{\sigma\to 0}\mathbf{E}_\vartheta\left(\frac{\bar\vartheta_\sigma-\vartheta}{\sigma}\right)^2\geq \mathrm{I}\left(\vartheta\right)^{-1}.$$

The MLE $\hat\vartheta_\sigma$ and BE $\tilde\vartheta_\sigma$ are consistent,

$$\frac{\hat\vartheta_\sigma-\vartheta}{\sigma}\Longrightarrow\mathcal{N}\left(0,\mathrm{I}\left(\vartheta\right)^{-1}\right),\qquad \frac{\tilde\vartheta_\sigma-\vartheta}{\sigma}\Longrightarrow\mathcal{N}\left(0,\mathrm{I}\left(\vartheta\right)^{-1}\right)$$

and **both estimators are asymptotically efficient**. Say,

$$\lim_{\sigma\to 0}\mathbf{E}_\vartheta\left(\frac{\hat\vartheta_\sigma-\vartheta}{\sigma}\right)^2= \mathrm{I}\left(\vartheta\right)^{-1}.$$

(K. 88, K. 94)

Asymptotic expansion

The asymptotic normality can be proved *in probability* too, i.e., the estimators can be written as

$$\hat{\vartheta}_\sigma = \vartheta + \frac{\xi}{\sqrt{\mathrm{I}\left(\vartheta\right)}}\,\sigma + o\left(\sigma\right), \qquad\qquad \xi = -\frac{\gamma^2}{\sqrt{\mathrm{I}\left(\vartheta\right)}}\int_0^T x_{t-2\vartheta}\,\mathrm{d}W_t.$$

It is interesting to see the next after Gaussian terms of these expansion by the powers of σ.

In regular problems the expansion is like (K., 94)

$$\hat{\vartheta}_\sigma = \vartheta + \frac{\xi_1}{\sqrt{\mathrm{I}\left(\vartheta\right)}}\,\sigma + \left(a_1\xi_1^2 + a_2\xi_1\xi_2 + a_3\xi_1\xi_3 + a_4\xi_4\right)\,\sigma^2 + o\left(\sigma^2\right)$$

In our case of delay estimation

$$\mathrm{d}X_t = -\gamma\, X_{t-\vartheta}\,\mathrm{d}t + \sigma\,\mathrm{d}W_t, \quad X_s = x_s, s \leq 0, \quad 0 \leq t \leq T$$

it is

$$\hat{\vartheta}_\sigma = \vartheta + \frac{\xi}{\sqrt{\mathrm{I}\left(\vartheta\right)}}\,\sigma - \frac{\gamma\sqrt{T}}{2\,\mathrm{I}\left(\vartheta\right)^{\frac{3}{4}}}\,\frac{\zeta\ \mathrm{sgn}\left(\xi\right)}{\sqrt{|\xi|}}\,\sigma^{3/2} + O\left(\sigma^2\right)$$

where $\zeta \sim \mathcal{N}\left(0,1\right)$ is independent of $\xi \sim \mathcal{N}\left(0,1\right)$.

The similar expansion can be found for Bayesian estimator as well, but the final expression is more cumbersome.

Generalizations

Scale estimation

The similar results we have for the model of *scale-type* delay:

$$\mathrm{d}X_t = -\gamma \, X_{\vartheta t} \,\mathrm{d}t + \sigma \,\mathrm{d}W_t, \quad X_0 = x_0, \quad 0 \leq t \leq T$$

where $\vartheta \in (0,1)$. The Fisher information is

$$\mathrm{I}\,(\vartheta) = \gamma^2 \int_0^T t^2 \; x_{\vartheta^2 t}^2 \,\mathrm{d}t$$

and the estimators have the similar asymptotic normality.

All proofs can be found in *Prob. Theory & Appl.* 1988, 33, 1, 175-179 or in Kutoyants, *Identification of dynamical systems with small noise*, Kluwer, 1994.

Nonlinear system

Let

$$\mathrm{d}X_t = S\left(X_{t-\vartheta}, t\right)\,\mathrm{d}t + \sigma\, b\left(X_t, t\right)\mathrm{d}W_t, \qquad 0 \leq t \leq T$$

where $S(\cdot)$ is smooth function, then we have for the same estimators

$$\frac{\hat\vartheta_\sigma - \vartheta}{\sigma} \Longrightarrow \mathcal{N}\left(0, \mathrm{I}\left(\vartheta\right)^{-1}\right), \qquad \frac{\tilde\vartheta_\sigma - \vartheta}{\sigma} \Longrightarrow \mathcal{N}\left(0, \mathrm{I}\left(\vartheta\right)^{-1}\right)$$

where

$$\mathrm{I}\left(\vartheta\right) = \int_0^T \frac{S\left(x_{t-2\vartheta}, t-\vartheta\right)^2 \dot S\left(x_{t-\vartheta}, t\right)^2}{b\left(x_t, t\right)^2}\mathrm{d}t$$

and

$$\frac{\mathrm{d}x_t}{\mathrm{d}t} = S\left(x_{t-\vartheta}\right), \qquad x_s,\, s \leq 0, \quad 0 \leq t \leq T.$$

Apoyan, *Scient. Notes of Yerevan State University*, 1986, 33-62.

Multiple delays

Let

$$\mathrm{d}X_t = \sum_{j=1}^{k} \lambda_j \; X_{t-\tau_j} \,\mathrm{d}t + \sigma \,\mathrm{d}W_t, \quad X_s = x_s, s \leq 0, \quad 0 \leq t \leq T$$

where $\boldsymbol{\vartheta} = (\lambda_1, \ldots, \lambda_k, \tau_1, \ldots, \tau_k)$ is unknown parameter.

It is shown that the MLE $\hat{\boldsymbol{\vartheta}}_T$ is consistent, asymptotically normal and asymptotically efficient estimator. K.-Mourid-Bosq, *Ann. Inst. Henri Poincarre*, 1992, 28, 1, 95-106.

The properties of a minimum distance estimator of $\boldsymbol{\vartheta}$ are described by K.-Mourid, *Publ. Inst. Stat. Univ. Paris*, 1994, , 2, 3-18.

Estimation (large samples asymptotics)

Let

$$\mathrm{d}X_t = -\gamma\, X_{t-\vartheta}\, \mathrm{d}t + \sigma\, \mathrm{d}W_t, \quad X_s = x_s, s \leq 0, \quad 0 \leq t \leq T$$

and we have to study the properties of estimators as $T \rightarrow \infty$.

Suppose that $\vartheta \in \left(0, \frac{\pi}{2\gamma}\right)$, then the process X_t has ergodic properties.

The problem of parameter ϑ estimation became **non regular**.

The likelihood ratio process

$$Z_T\left(u\right)=\frac{\mathrm{d}\mathbf{P}^{(T)}_{\vartheta+\frac{u}{T\gamma}}}{\mathrm{d}\mathbf{P}^{(T)}_{\vartheta}}\left(X^T\right)\Longrightarrow Z(u)=\exp\left\{W(u)-\frac{1}{2}\left|u\right|\right\}.$$

The same limit as in the case of discontinuous trend, say, (K. 2003)

$$\mathrm{d}X_t=-\mathrm{sgn}(X_t-\vartheta)\,\mathrm{d}t+\sigma\,\mathrm{d}W_t,\quad X_0,\quad 0\leq t\leq T.$$

Introduce the random variables $\hat u$ and $\tilde u$ by equations

$$Z_\tau\left(\hat u\right)=\sup_{u\in\mathbf{R}}Z(u),\qquad \tilde u=\frac{\int_{\mathbf{R}}u\,Z(u)\,\mathrm{d}u}{\int_{\mathbf{R}}Z(u)\,\mathrm{d}u}.$$

We have $\mathbf{E}\hat u^2=26$ (Terent'ev, 1968) and $\mathbf{E}\tilde u^2=\sigma_0^2$ where the value $\sigma_0^2=16\,\zeta\left(3\right)=19,23...$ (Rubin and Song, 1995). Another representation for $\mathbf{E}\tilde u^2$ was proposed by Golubev (1979).

For all estimators we have

$$\lim_{T\to\infty} T^2 \mathbf{E}_\vartheta \left(\bar\vartheta_T - \vartheta\right)^2 \geq \frac{\sigma_0^2}{\gamma^2}$$

and the MLE $\hat\vartheta_T$ and BE $\tilde\vartheta_T$ are consistent,

$$T\left(\hat\vartheta_T - \vartheta\right) \Longrightarrow \frac{\hat u}{\gamma}, \qquad T\left(\tilde\vartheta_T - \vartheta\right) \Longrightarrow \frac{\tilde u}{\gamma}$$

Moreover BE is asymptotically efficient.

Küchler-K., *Scand. J. Statist*, 2000, 3, 405-414.

Let

$$\mathrm{d}X_t = -\gamma\, X_{t-\tau}\,\mathrm{d}t + \sigma\,\mathrm{d}W_t, \quad X_s = x_s, s \leq 0, \quad 0 \leq t \leq T$$

where $\boldsymbol{\vartheta} = (\gamma, \tau)$ is unknown parameter.

Using the similar arguments it is shown that the MLE $\hat{\boldsymbol{\vartheta}}_T = (\hat{\gamma}_T, \hat{\tau}_T)$ and BE $\tilde{\boldsymbol{\vartheta}}_T = (\tilde{\gamma}_T, \tilde{\tau}_T)$ are consistent,

$$\sqrt{T}\,(\hat{\gamma}_T - \gamma) \Longrightarrow \eta, \qquad T\,(\hat{\tau}_T - \tau) \Longrightarrow \frac{\hat{u}}{\gamma}$$

and

$$\sqrt{T}\,(\tilde{\gamma}_T - \gamma) \Longrightarrow \eta, \qquad T\,(\tilde{\tau}_T - \tau) \Longrightarrow \frac{\tilde{u}}{\gamma}.$$

Here $\eta \sim \mathcal{N}\left(0, r\left(\boldsymbol{\vartheta}\right)^2\right)$ is independent of $\hat{u}$ and $\tilde{u}$ and

$$r\left(\boldsymbol{\vartheta}\right)^2 = \int_0^{\infty} x_t^2 \,\mathrm{d}t.$$

The function x_t is solution of the fundamental equation

$$\frac{\mathrm{d}x_t}{\mathrm{d}t} = -\gamma\, x_{t-\tau}, \quad x_s = 0, s < 0, \quad x_0 = 1.$$

Hypotheses Testing (small noise asymptotics)

The observed process is

$$\mathrm{d}X_t = -\gamma\, X_{t-\vartheta}\,\mathrm{d}t + \sigma\,\mathrm{d}W_t, \quad X_s = x_s, s \leq 0 \quad 0 \leq t \leq T$$

and we have to test the following two hypotheses

$$\begin{aligned} \mathcal{H}_0 : &\qquad \vartheta = 0, \qquad \text{(no delay)} \\ \mathcal{H}_1 : &\qquad \vartheta > 0, \end{aligned}$$

The contigual alternatives correspond to $\vartheta = u\sigma$:

$$\begin{aligned} \mathcal{H}_0 : &\qquad u = 0, \qquad \text{(no delay)} \\ \mathcal{H}_1 : &\qquad u > 0, \end{aligned}$$

Let us fix some $\varepsilon \in (0,1)$ and denote by $\mathcal{K}_\varepsilon$ the class of tests of asymptotic level $1-\varepsilon$. The limit $(\sigma \to 0)$ of the power function of Neyman-Pearson test $\phi_\sigma \in \mathcal{K}_\varepsilon$ is

$$\beta\left(u\right) = \mathbf{P}\left\{\zeta > z_\varepsilon - u\,\mathrm{I}_0^{1/2}\right\},$$

where z_ε is $1-\varepsilon$ quantil of $\mathcal{N}\left(0,1\right)$ and the Fisher information $\mathrm{I}_0 = \mathrm{I}\left(0\right)$ is

$$\mathrm{I}\left(0\right) = \frac{x_0^2\,\gamma^3}{2}\left(1 - e^{-2\gamma T}\right).$$

Let us introduce the *score-function test*

$$\phi_\sigma^*\left(X^\sigma\right)=\chi_{\left\{\Delta_\sigma>z_\varepsilon\right\}},$$

where

$$\Delta_\sigma=-\frac{\gamma^2}{\sigma\sqrt{\mathrm{I}_0}}\int_0^T x_{t-2\vartheta}\;\left[\mathrm{d}X_t+\gamma X_t\,\mathrm{d}t\right].$$

Note that under $\mathcal{H}_0$ the statistic $\Delta_\sigma\sim\mathcal{N}\left(0,1\right)$ and its power function $\beta_\sigma^*\left(u\right)=\mathbf{E}_u\phi_\sigma^*$ converges to the power function $\beta\left(u\right)$ uniformly on compacts $0\leq u\leq K$ for any $K>0$.

Therefore the test $\phi_\sigma^*\left(X^\sigma\right)$ is *asymptotically uniformly most powerful* in the class $\mathcal{K}_\varepsilon$.

The same properties have the likelihood ratio test

$$\phi_\sigma\left(X^\sigma\right)=\chi_{\left\{\delta_\sigma>b_\varepsilon\right\}},\qquad \delta_\sigma\left(X^\sigma\right)=\sup_{\vartheta>0}L\left(\vartheta,X^\sigma\right),$$

where $b_\varepsilon=\exp\left\{z_\varepsilon^2/2\mathrm{I}_0\right\}$ and the test based on the MLE

$$\hat\phi_\sigma\left(X^\sigma\right)=\chi_{\left\{\hat\delta_\sigma>z_\varepsilon \mathrm{I}_0^{-1/2}\right\}},\qquad \hat\delta_\sigma\left(X^\sigma\right)=\frac{\hat\vartheta_\sigma}{\sigma}$$

respectively.

Hypotheses Testing (large samples asymptotics)

The observed process is

$$\mathrm{d}X_t = -\gamma\, X_{t-\vartheta}\,\mathrm{d}t + \sigma\,\mathrm{d}W_t, \quad X_s = x_s, s \leq 0 \quad 0 \leq t \leq T$$

and we have to test the following two hypotheses

$$\begin{aligned} &\mathcal{H}_0 : \qquad && \vartheta = 0, \qquad && \text{(no delay)} \\ &\mathcal{H}_1 : \qquad && \vartheta > 0, && \end{aligned}$$

The contigual alternatives correspond to $\vartheta = u/\gamma T$:

$$\begin{aligned} &\mathcal{H}_0 : \qquad && u = 0, \qquad && \text{(no delay)} \\ &\mathcal{H}_1 : \qquad && u > 0, && \end{aligned}$$

We have no asymptotically uniformly most powerful test.

The likelihood ratio test

$$\phi_T\left(X^T\right) = \chi_{\left\{\delta_T > \frac{1}{\varepsilon}\right\}}, \qquad \delta_T\left(X^T\right) = \sup_{\vartheta>0} L\left(\vartheta, X^T\right),$$

belongs to $\mathcal{K}_\varepsilon$ and its power function

$$\beta_T\left(u, \hat{\phi}_T\right) = \mathbf{P}\left\{\zeta + \max\left[\eta, \xi\right] > \ln\frac{1}{\varepsilon} - \frac{u}{2}\right\} + o\left(1\right),$$

where the random variables ζ, η and ξ are independent and

$$\zeta \sim \mathcal{N}\left(0, u\right), \qquad F_\xi\left(x\right) = 1 - \mathrm{e}^{-x},$$

$$F_\eta\left(x\right) = \Phi\left(\frac{x}{\sqrt{u}} + \frac{\sqrt{u}}{2}\right) + \mathrm{e}^{-x}\left[1 - \Phi\left(\frac{x}{\sqrt{u}} - \frac{\sqrt{u}}{2}\right)\right],$$