On asymptotically efficient estimation in partially observed systems

Vladimir Zaiats

Universitat de Vic, Spain

(joint work with Yury A. Kutoyants)

SAPS-X, Le Mans, March 20, 2015

Before we begin...

March 20th, 2015

- The day of a solar eclipse which is the first eclipse of 2015
- It is also the March Equinox

  A total solar eclipse coinciding with Northern Hemisphere's Spring Equinox and Southern Hemipshere's Autumn or Fall Equinox hasn't happened since 1662 and will not happen again until March 20, 2034.

- Super New Moon

  Only 12 hours before the beginning of the eclipse, the Moon has been at its perigee – the point closest to the Earth on its orbit around it. This makes the Moon on March 20, 2015, a Super New Moon.

March 20th, 2015

- Part of Saros Series 120

  The Saros Cycle, one of the most studied eclipse cycles, occurs every 18 years. Two solar eclipses separated by a Saros Cycle have similar features they occur at the same lunar node, with the Moon roughly at the same distance from the Earth.

  Saros Series 120 began with a Partial Solar Eclipse visible from the Southern Hemisphere on May 27, 933 CE, and will end with a Partial Solar Eclipse visible in the Northern Hemisphere on July 7, 2195.

  The previous Saros Series 120 solar eclipse occurred on March 9th, 1997, which is of course equal to $\text{SAPS–I} + \varepsilon$ where SAPS–I =January 27–28, 1997, and ϵ may be interpreted as a small noise.

Sources:

www.timeanddate.com/eclipse/10-facts-solar-eclipse-march-2015.html

http://eclipse.gsfc.nasa.gov/SEsaros/SEsaros120.html

VZ

More about March 20th, 2015

- It is Robert Liptser's birthday
- It is the last working day of SAPS X

First conclusion

All above mentioned facts enable us to clearly conclude that:

Yury Kutoyants and his team have indeed worked hard in order to, first, bring together these events and, second, to make them coincide with SAPS X!

Outline of the talk

- Setting-up
- Estimates
- Conclusions

Setting-up

Our objective is to focus on a model related to partially observed systems, where the function we would like to control is not observed directly, and to perform estimation of different functional characteristics in this model.

Assume that we observe a process $X = (X_t,\ 0 \le t \le T)$ satisfying the following system of stochastic differential equations:

$$
\begin{aligned}
\mathrm{d}X_t &= h_t Y_t \,\mathrm{d}t + \varepsilon \,\mathrm{d}W_t, \quad X_0 = 0, \\
\mathrm{d}Y_t &= g_t Y_t \,\mathrm{d}t + \varepsilon \,\mathrm{d}V_t, \quad Y_0 = y_0 \neq 0, \quad 0 \le t \le T,
\end{aligned}
$$

where W_t and V_t, $0 \le t \le T$, are two independent Wiener processes. The process $Y = (Y_t,\ 0 \le t \le T)$ **cannot be observed** directly, but it is *the one we would like to control*.

This is a Kalman filter (KF) framework.

The first equation

$$dX_t = h_t Y_t \, dt + \varepsilon \, dW_t \tag{1}$$

is the *observation equation*.

The second equation

$$dY_t = g_t Y_t \, dt + \varepsilon \, dV_t \tag{2}$$

is the *system equation* reflecting the "*state of nature*."

[Meinhold and Singpurwalla (1983) is a good reference for understanding the heuristics of the KF.]

There are many physical systems for which the “state of nature” changes over time according to a relationship prescribed by engineering or scientific principles.

This is why the ability to include a knowledge of the system behavior in the statistical model is an apparent source of attractiveness of the KF.

One of examples often quoted when KF’s are involved is the following one (assume for a while a vector-valued version of system (1)–(2)):

Imagine that we are interested in tracking a satellite orbiting the Earth.

Look at the system equation

$$\mathrm{d}Y_t = g_t Y_t \,\mathrm{d}t + \varepsilon \,\mathrm{d}V_t.$$

The unknown "state of nature" Y_t at each time instant t could be the position and speed of the satellite. These quantities cannot be measured directly. Instead, from tracking stations around the Earth, we may obtain measurements of distance to the satellite and the accompanying angles of measurement; these are the Y_t's.

Now switch to the observation equation

$$\mathrm{d}X_t = h_t Y_t \, \mathrm{d}t + \varepsilon \, \mathrm{d}W_t.$$

The principles of geometry, mapping Y into X would be incorporated in h while W would reflect the measurement error.

g would prescribe how the position and speed change in time according to the physical laws governing orbiting bodies, while V, would allow for deviations from these laws owing to such factors as non-uniformity of the Earth's gravitational field, and so on.

In this model, we consider the problem of estimation of different functions on $0 \leq t \leq T$, in the asymptotics of a *small noise*, i.e., as $\varepsilon \to 0$. We propose some kernel-type estimators for our functions and study their properties.

The first observation to be made is that the functions h_t and g_t, $0 \le t \le T$, *cannot be estimated at the same time*.

The **reason** is as follows: Even in the situation where there is no noise we have

$$\begin{aligned} \mathrm{d}x_t &= h_t y_t \,\mathrm{d}t, \quad x_0 = 0, \\ \mathrm{d}y_t &= g_t y_t \,\mathrm{d}t, \quad y_0 \neq 0, \quad 0 \le t \le T. \end{aligned}$$

The second equation gives $y_t = y_0 \exp\left\{\int_0^t g_s \mathrm{d}s\right\}$ and therefore

$$\frac{\mathrm{d}x_t}{\mathrm{d}t} = f_t, \quad x_0 = 0, \qquad \text{where} \quad f_t = h_t y_t = h_t\, y_0 \exp\left\{\int_0^t g_s \mathrm{d}s\right\}.$$

Hence, we can only estimate f_t, the latter being a mixture of g_t and h_t. This is why we have to consider estimation of the functions f_t, g_t, and h_t separately, one by one. **We assume that all unknown functions are at least continuous.**

Estimation of f_t

Suppose that $f_t \in \mathcal{F}$, where $\mathcal{F}$ is the set of uniformly continuous functions bounded by a constant.

Introduce the estimator

$$\hat{f}_{t,\varepsilon} = \frac{1}{\varphi_\varepsilon} \int_0^T K\left(\frac{s-t}{\varphi_\varepsilon}\right) \mathrm{d}X_s$$

Heuristics: Since in the unperturbed model we have $f_t = \dot{x}_t$, we use a "*slow derivative*" of X_t as an estimator of f_t. This "slow derivative" is just the above kernel-type estimator.

Here and in the sequel, the kernel $K(\cdot)$ is bounded and satisfies the "*standard conditions*": K has a compact support $[A, B]$ with $A < 0$, $B > 0$, and

$$\int_A^B K(u)\, \mathrm{d}u = 1.$$

The first result is the following:

Proposition: *Let* $\varphi_\varepsilon \to 0$ *and* $\varepsilon^2\varphi_\varepsilon^{-1} \to 0$ *as* $\varepsilon \to 0$. *Then for any* $0 < a < b < T$ *we have*

$$\lim_{\varepsilon\to 0} \sup_{f\in\mathcal{F}} \sup_{a\le t\le b} \mathbf{E}_f \left(\hat{f}_{t,\varepsilon} - f_t\right)^2 = 0.$$

Remark. As usual, this convergence can be made uniform on $[0, T]$ by a special choice of one-sided kernels.

Suppose that the functions h_t and g_t are such that the function f_t is k-times continuously differentiable and the k-th derivative satisfies the Hölder condition of order $\alpha \in (0, 1]$:

$$\left| f_t^{(k)} - f_s^{(k)} \right| \leq L \left| t - s \right|^{\alpha}.$$

We denote by $\mathcal{F}_\beta$, where $\beta = k + \alpha$, the class of such functions. The kernel $K(\cdot)$ satisfies the following standard conditions in addition to those already stated:

$$\int_A^B K(u)\, u^l \mathrm{d}u = 0, \qquad l = 1, \ldots, k.$$

We take now $\varphi_\varepsilon = \varepsilon^{\frac{2}{2\beta+1}}$.

Proposition: *There exists a constant $C > 0$ such that*

$$\lim_{\varepsilon \to 0} \sup_{f \in \mathcal{F}_\beta} \sup_{a \leq t \leq b} \varepsilon^{-\frac{4\beta}{2\beta+1}} \mathbf{E}_f \left(\hat{f}_{t,\varepsilon} - f_t \right)^2 \leq C.$$

This rate of convergence is optimal in the following sense:

Proposition: *There exists a constant* $c > 0$ *such that*

$$\varliminf_{\varepsilon \to 0} \inf_{\bar{f}_{t,\varepsilon}} \sup_{f \in \mathcal{F}_\beta} \varepsilon^{-\frac{4\beta}{2\beta+1}} \mathbf{E}_f \left(\bar{f}_{t,\varepsilon} - f_t\right)^2 \geq c.$$

Here, $\bar{f}_{t,\varepsilon}$ is an arbitrary estimator of f_t. Therefore an estimator whose rate of convergence would be better than that of $\hat{f}_{t,\varepsilon}$ **does not exist**.

The proof of this bound requires the Kalman-Bucy filter for writing the likelihood ratio.

Estimation of h_t

Let us come back to our model

$$\begin{aligned} \mathrm{d}X_t &= h_t Y_t \,\mathrm{d}t + \varepsilon \,\mathrm{d}W_t, \quad X_0 = 0, \\ \mathrm{d}Y_t &= g_t Y_t \,\mathrm{d}t + \varepsilon \,\mathrm{d}V_t, \quad Y_0 = y_0 \neq 0, \quad 0 \leq t \leq T. \end{aligned}$$

Suppose now that g_t is a known bounded function and we would like to estimate h_t which is unknown. Then the estimator is as follows:

$$\hat{h}_{t,\varepsilon} = \frac{1}{\varphi_\varepsilon} \int_0^T K\left(\frac{s-t}{\varphi_\varepsilon}\right) y_s^{-1} \,\mathrm{d}X_s.$$

Heuristics: In the unperturbed model we have $h_t = \dot{x}_t / y_t$. This hints us to use the "slow derivative" of X_t is the form of kernel-type estimator, and y_s appears in the denominator.

The estimator $\hat{h}_{t,\varepsilon}$ has properties similar to those of $\hat{f}_{t,\varepsilon}$. In particular, if $\varphi_\varepsilon = \varepsilon^{\frac{2}{2\beta+1}}$, then

$$\lim_{\varepsilon \to 0} \sup_{h \in \mathcal{F}_\beta} \sup_{a \le t \le b} \varepsilon^{-\frac{4\beta}{2\beta+1}} \mathbf{E}_h \left(\hat{h}_{t,\varepsilon} - h_t \right)^2 \le C.$$

Estimation of y_t

Suppose that h_t is a known function bounded away from zero. We would like to estimate $y_t = y_0 \exp\left\{\int_0^t g_s \mathrm{d}s\right\}$, where g_t is unknown. Then it can be shown that the estimator

$$\hat{y}_{t,\varepsilon} = \frac{1}{\varphi_\varepsilon} \int_0^T K\left(\frac{s-t}{\varphi_\varepsilon}\right) h_s^{-1} \, \mathrm{d}X_s$$

has similar properties to those of $\hat{f}_{t,\varepsilon}$, i.e., it is uniformly consistent and has asymptotically optimal rate of convergence under a smoothness condition ($g_t \in \mathcal{F}_{\beta-1}$): If $\varphi_\varepsilon = \varepsilon^{\frac{2}{2\beta+1}}$, then

$$\lim_{\varepsilon \to 0} \sup_{g \in \mathcal{F}_{\beta-1}} \sup_{a \le t \le b} \varepsilon^{-\frac{4\beta}{2\beta+1}} \mathbf{E}_g \left(\hat{y}_{t,\varepsilon} - y_t\right)^2 \le C.$$

Estimation of g_t

Suppose now that h_t is a known function bounded away from zero. We would like to estimate the function g_t from observations $X = (X_t,\ 0 \le t \le T)$. If we consider the unperturbed model, we will see that

$$g_t = \frac{y_t'}{y_t}.$$

Therefore if we construct good estimators $\hat{y}_{t,\varepsilon}'$ and $\hat{y}_{t,\varepsilon}$ of y_t' and y_t, respectively, then their ratio can be a consistent estimator of g_t. Since the convergence of moments is required, we introduce truncation and put

$$\hat{g}_{t,\varepsilon} = \frac{\hat{y}_{t,\varepsilon}'}{\hat{y}_{t,\varepsilon}} \ \mathbb{1}_{\{\hat{y}_{t,\varepsilon} > \kappa\}}$$

where the constant $\kappa > 0$ is sufficiently small.

Suppose that $g_t \in \mathcal{F}_{\beta-1}$ and introduce the estimator of the derivative

$$\hat{y}'_{t,\varepsilon} = \frac{1}{\varphi_\varepsilon^2} \int_0^T Q\left(\frac{s-t}{\varphi_\varepsilon}\right) h_s^{-1}\, \mathrm{d}X_s$$

where the kernel $Q(\cdot)$ has compact support $[A_*, B_*]$, $A_* < 0$, $B_* > 0$, and satisfies the conditions

$$\int_{A_*}^{B_*} Q(u)\, u\, \mathrm{d}u = 1, \quad \int_{A_*}^{B_*} Q(u)\, u^l \mathrm{d}u = 0, \quad l = 0, 2, \ldots, k.$$

Then it can be shown that

$$\sup_{a \le t \le b} \varepsilon^{-\frac{4\beta-4}{2\beta+1}} \mathbf{E}_g \left(\hat{y}'_{t,\varepsilon} - y'_t\right)^2 \le C,$$

where we put $\varphi_\varepsilon = \varepsilon^{2/(2\beta+1)}$.

Suppose that $y_0 > 0$ and put

$$\kappa = \frac{1}{2} \inf_{g \in \mathcal{F}_{\beta-1}} \inf_{0 \le t \le T} y_t .$$

The rate of convergence of the estimator $\hat{g}_{t,\varepsilon}$ is the same:

Proposition: *There exists a constant $C > 0$ such that*

$$\sup_{a \le t \le b} \varepsilon^{-\frac{4\beta-4}{2\beta+1}} \mathbf{E}_g \left(\hat{g}_{t,\varepsilon} - g_t \right)^2 \le C.$$

It can be shown that this rate is optimal in the minimax sense.

Note that if $|g_t| \le D$, then we can take

$$\kappa = \frac{y_0}{2} e^{-DT} .$$

Conclusions

We have shown different possibilities of estimation in partially observed linear models.

For all estimators, the rate of convergence is established, and this rate is shown to be the optimal one.

Thank you!