Hidden Markov Processes and Adaptive Filtration.

Yury A. Kutoyants

Le Mans University, France

RMR 2024

Rouen, 19 Juin, 2024

1 Hidden Markov Processes

1.1 Model of observations.

We are given a couple of equations

$$\mathrm{d}X_t = f(\vartheta, t)\, Y_t \mathrm{d}t + \sigma(t)\,\mathrm{d}W_t, \qquad X_0, \quad 0 \le t \le T,$$

$$\mathrm{d}Y_t = a(\vartheta, t)\, Y_t \mathrm{d}t + b(\vartheta, t)\,\mathrm{d}V_t, \qquad Y_0, \quad 0 \le t \le T,$$

where $W_t, 0 \le t \le T$ and $V_t, 0 \le t \le T$ are independent Wiener processes and the observations are $X^T = (X_t, 0 \le t \le T)$. The O-U process $Y^T = (Y_t, 0 \le t \le T)$ is hidden. The functions $f(\vartheta, t), \sigma(t), a(\vartheta, t), b(\vartheta, t), t \in [0, T]$ are supposed to be known and the parameter $\vartheta \in \Theta \subset \mathcal{R}^d$ is unknown.

The conditional expectation $m(\vartheta,t)=\mathbf{E}_\vartheta\left(Y_t|X_s,0\leq s\leq t\right)$ and the error $\gamma(\vartheta,t)=\mathbf{E}_\vartheta\left(Y_t-m(\vartheta,t)\right)^2$ are solutions of Kalman-Bucy (K-B) filtration equations

$$\mathrm{d}m(\vartheta,t)=\left[a(\vartheta,t)-\frac{\gamma(\vartheta,t)f^2}{\sigma(t)^2}\right]m(\vartheta,t)\,\mathrm{d}t+\frac{\gamma(\vartheta,t)f(\vartheta,t)}{\sigma(t)^2}\mathrm{d}X_t,$$

$$\frac{\partial\gamma(\vartheta,t)}{\partial t}=2a(\vartheta,t)\gamma(\vartheta,t)-\frac{\gamma(\vartheta,t)^2f(\vartheta,t)^2}{\sigma(t)^2}+b(\vartheta,t)^2,$$

with initial values $m_0=\mathbf{E}_\vartheta\left(Y_0|X_0\right)$ and $\gamma_0=\mathbf{E}_\vartheta\left(Y_0-m(\vartheta,0)\right)^2$. Recall that $m(\vartheta,t)$ is an optimal estimator of Y_t.

Pb.: *To obtain a good recurrent approximation* $m_t^\star,0<t\leq T$ *of the process* $m(\vartheta,t),0<t\leq T$.

We need a good estimator of ϑ, which depends on $X_s,0\leq s\leq t$ to use it for construction of $m_t^\star,0<t\leq T$.

How to chose an estimator? The likelihood ratio function is:

$$L\left(\vartheta, X^T\right) = \exp\left\{\int_0^T \frac{f\left(\vartheta,t\right) m\left(\vartheta,t\right)}{\sigma\left(t\right)^2}\,\mathrm{d}X_t - \int_0^T \frac{f\left(\vartheta,t\right)^2 m\left(\vartheta,t\right)^2}{2\sigma\left(t\right)^2}\mathrm{d}t\right\}.$$

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

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

To construct the MLE and BE of the parameter ϑ we have to calculate the functions $\left\{m\left(\vartheta,t\right),\ 0\leq t\leq T\right\},\ \vartheta\in\Theta$ and $\left\{\gamma\left(\vartheta,t\right),\ 0\leq t\leq T\right\},\ \vartheta\in\Theta$ defined by the K-B equations.

But we can not put the MLE $\hat{\vartheta}_T$ or BE $\tilde{\vartheta}_T$ in $m\left(\vartheta,t\right)$ because the stochastic integral containing $\gamma(\hat{\vartheta}_T,t)$ does not defined.

Remark that the direct numerical calculations of the MLE and BE is almost impossible.

To approximate $m(\vartheta, t)$ we propose the following program:

1. *Calculate a preliminary estimator* $\bar\vartheta_\tau$ *on relatively small interval of observations* $[0, \tau]$.

2. *Using* $\bar\vartheta_\tau$ *realize One-step MLE-process* $\vartheta_t^\star, \tau < t \leq T$

3. *As approximation of* $m(\vartheta, t)$ *we propose* $m_t^\star$ *obtained with the help of K-B equations, where* ϑ *is replaced by* $\vartheta_t^\star, \tau < t \leq T$

4. *Estimate the error* $m_t^\star - m(\vartheta, t)$.

We apply this construction to 5 different models of observations.

Let $\bar{\vartheta}_\tau$ be the consistent preliminary estimator with the "bad" rate of convergence ($\tau/T \to 0$) and the Fisher information matrices $\mathbf{I}_\tau^t(\vartheta), \tau \leq t \leq T$ are known. Then the One-step MLE-process is

$$\vartheta_t^\star = \bar{\vartheta}_\tau + \mathbf{I}_\tau^t\left(\bar{\vartheta}_\tau\right)^{-1} \int_\tau^t \frac{\dot{S}\left(\bar{\vartheta}_\tau, s\right)}{\sigma(s)^2}\left[\mathrm{d}X_s - S\left(\bar{\vartheta}_\tau, s\right)\mathrm{d}s\right]$$

Here $S(\vartheta, t) = f(\vartheta, s)\, m(\vartheta, s)$. For all models

$$\varphi^{-1}\left(\vartheta_t^\star - \vartheta_0\right) \Longrightarrow \mathcal{N}\left(0, \mathbf{I}_\tau^t(\vartheta)^{-1}\right).$$

One-step can be written in the recurrent form. Denote $L(\vartheta, t) = \dot{S}\left(\bar{\vartheta}_\tau, s\right) \dot{S}\left(\bar{\vartheta}_\tau, s\right)^\top \sigma(t)^{-2}$. Then

$$\begin{aligned}\mathrm{d}\vartheta_t^\star &= \mathbf{I}_\tau^t\left(\bar{\vartheta}_\tau\right)^{-1} L\left(\bar{\vartheta}_\tau, t\right)\left[\bar{\vartheta}_\tau - \vartheta_t^\star\right]\mathrm{d}t \\ &\quad + \mathbf{I}_\tau^t\left(\bar{\vartheta}_\tau\right)^{-1} \frac{\dot{S}\left(\bar{\vartheta}_\tau, t\right)}{\sigma(t)^2}\left[\mathrm{d}X_t - S\left(\bar{\vartheta}_\tau, t\right)\mathrm{d}t\right], \qquad \vartheta_\tau^\star = \bar{\vartheta}_\tau.\end{aligned}$$

The adaptive filter $m_t^\star, \tau \leq t \leq T$ is

$$\mathrm{d}m_t^\star = \left[a\left(\vartheta_t^\star, t\right) - \frac{\gamma_t^\star f\left(\vartheta_t^\star, t\right)^2}{\sigma\left(t\right)^2}\right] m_t^\star \mathrm{d}t + \frac{\gamma_t^\star f\left(\vartheta_t^\star, t\right)}{\sigma\left(t\right)^2} \mathrm{d}X_t,$$

$$\frac{\partial \gamma_t^\star}{\partial t} = -2a\left(\vartheta_t^\star, t\right)\gamma_t^\star - \frac{\left(\gamma_t^\star\right)^2 f\left(\vartheta_t^\star, t\right)^2}{\sigma\left(t\right)^2} + b\left(\vartheta_t^\star, t\right)^2,$$

On the error of estimation. We have the minimax lower bound

$$\lim_{\nu\rightarrow 0} \underline{\lim}_{\varphi\rightarrow 0} \sup_{\left\|\vartheta-\vartheta_0\right\|\leq\nu} \varphi^{-2}\mathbf{E}_\vartheta\left(\bar m_t - m\left(\vartheta, t\right)\right)^2 \geq \mathcal{S}_t^\star\left(\vartheta_0\right)^2$$

The adaptive filter $m_t^*, \tau \leq t \leq T$ is called asymptotically efficient if for all $\vartheta_0 \in \Theta$ and $t \in (\tau, T]$

$$\lim_{\nu\rightarrow 0} \lim_{\varphi\rightarrow 0} \sup_{\left\|\vartheta-\vartheta_0\right\|\leq\nu} \varphi^{-2}\mathbf{E}_\vartheta\left(m_t^* - m\left(\vartheta, t\right)\right)^2 = \mathcal{S}_t^\star\left(\vartheta_0\right)^2$$

The quantity $\mathcal{S}_t^\star\left(\vartheta_0\right)^2$ can be calculated in all considered models.

1.2 HMP with small noises in both equations.

Consider the linear two-dimensional partially observed system

$$\mathrm{d}X_t = f(\vartheta, t)\, Y_t \mathrm{d}t + \varepsilon\sigma(t)\,\mathrm{d}W_t, \qquad X_0 = 0, \quad 0 \leq t \leq T,$$

$$\mathrm{d}Y_t = a(\vartheta, t)\, Y_t \mathrm{d}t + \varepsilon b(t)\,\mathrm{d}V_t, \quad Y_0 = y_0 \neq 0, \quad 0 \leq t \leq T,$$

Asymptotic $\varepsilon \to 0$. The preliminary estimator is $(x_t(\vartheta) = X_t|_{\varepsilon=0})$

$$\check{\vartheta}_{\tau_\varepsilon} = \arg \inf_{\vartheta\in\Theta} \int_0^{\tau_\varepsilon} |X_t - x_t(\vartheta)|^2 \mathrm{d}t, \qquad \tau_\varepsilon \to 0$$

K, (1994) Identification of Dynamical Systems with Small Noise. Kluwer Academic Publisher, Dordrecht.

K, Zhou, L. (2021) "On parameter estimation of the hidden Gaussian process in perturbed SDE", Electr. J. of Stat.,*15,211-234*

Introduce the notation: $S_* (\vartheta, \vartheta_0, t) = f (\vartheta, t)\, y_t (\vartheta, \vartheta_0)$,

$$\dot S_* (\vartheta, \vartheta_0, t) = \frac{\partial S_* (\vartheta, \vartheta_0, t)}{\partial \vartheta} = \dot f (\vartheta, t)\, y_t (\vartheta, \vartheta_0) + f (\vartheta, t)\, \dot y_t (\vartheta, \vartheta_0),$$

$$\mathbf{I} (\vartheta_0) = \int_0^T \frac{\dot S_* (\vartheta_0, \vartheta_0, t)\, \dot S_* (\vartheta_0, \vartheta_0, t)^\top}{\sigma (t)^2} \mathrm{d}t.$$

The $d \times d$ matrix $\mathbf{I} (\vartheta_0)$, $\vartheta_0 \in \Theta$ is the Fisher information. The MLE and BE are uniformly on compacts $\mathbb{K} \subset \Theta$ consistent, asymptotically normal

$$\frac{\hat\vartheta_\varepsilon - \vartheta_0}{\varepsilon} \Longrightarrow \zeta \sim \mathcal{N}\left(0, \mathbf{I} (\vartheta_0)^{-1}\right), \qquad \frac{\tilde\vartheta_\varepsilon - \vartheta_0}{\varepsilon} \Longrightarrow \zeta.$$

One-step MLE-process $\vartheta^\star_{t,\varepsilon}, \tau_\varepsilon < t \leq T$ are defined by the relations

$$\vartheta^\star_{\varepsilon} = \check\vartheta_{\tau_\varepsilon} + \mathbf{I}_{\tau_\varepsilon}^{T}\left(\check\vartheta_{\tau_\varepsilon}\right)^{-1}\int_{\tau_\varepsilon}^{T}\frac{\dot M\left(\check\vartheta_{\tau_\varepsilon},s\right)}{\sigma\left(s\right)^2}\left[{\rm d}X_s - M\left(\check\vartheta_{\tau_\varepsilon},s\right){\rm d}s\right],$$

$$\vartheta^\star_{t,\varepsilon} = \check\vartheta_{\tau_\varepsilon} + \mathbf{I}_{\tau_\varepsilon}^{t}\left(\check\vartheta_{\tau_\varepsilon}\right)^{-1}\int_{\tau_\varepsilon}^{t}\frac{\dot M\left(\check\vartheta_{\tau_\varepsilon},s\right)}{\sigma\left(s\right)^2}\left[{\rm d}X_s - M\left(\check\vartheta_{\tau_\varepsilon},s\right){\rm d}s\right].$$

Here $M\left(\vartheta,s\right) = f\left(\vartheta,s\right)m\left(\vartheta,s\right)$ and

$$\mathbf{I}_{\tau}^{t}\left(\vartheta_0\right) = \int_{\tau}^{t}\frac{\dot S_*\left(\vartheta_0,\vartheta_0,s\right)\dot S_*\left(\vartheta_0,\vartheta_0,s\right)^\top}{\sigma\left(s\right)^2}{\rm d}s$$

This estimator-process is as. normal

$$\frac{\vartheta^\star_{t,\varepsilon}-\vartheta_0}{\varepsilon} \Longrightarrow \zeta \sim \mathcal{N}\left(0,\mathbf{I}_0^t\left(\vartheta\right)^{-1}\right),$$

This estimator is used for adaptive filtration.

1.3 HMP with low noise observations

Consider the linear two-dimensional partially observed system

$$
\begin{aligned}
{\rm d}X_t &= f\left(\vartheta,t\right)Y_t{\rm d}t+\varepsilon\sigma\left(t\right){\rm d}W_t, \qquad X_0, \quad 0\leq t\leq T,\\
{\rm d}Y_t &= a\left(\vartheta,t\right)Y_t{\rm d}t+b\left(\vartheta,t\right){\rm d}V_t, \quad Y_0,
\end{aligned}
$$

Asymptotic $\varepsilon\rightarrow 0$. Below $\tau_\varepsilon=\varepsilon^{1/12}, t_{i+1}-t_i=\varepsilon^{1/3}, N_\varepsilon=\varepsilon^{-1/4}$

$$
\Psi_{\tau_\varepsilon,\varepsilon}=\sum_{i=0}^{N_\varepsilon-1}\left(\frac{X_{t_{i+1}+\varepsilon}-X_{t_{i+1}}}{\varepsilon}-\frac{X_{t_i+\varepsilon}-X_{t_i}}{\varepsilon}\right)^2,
$$

$$
\int_0^{\tau_\varepsilon} f\left(\check\vartheta_{\tau_\varepsilon},t\right)^2 b\left(\check\vartheta_{\tau_\varepsilon},t\right)^2{\rm d}t=\Psi_{\tau_\varepsilon,\varepsilon}
$$

K (2019) "On parameter estimation of hidden Ornstein -Uhlenbeck process", J. Multivariate Analysis. *169, 1, 248-263.*

K (2024) "Volatility estimation of hidden Markov process and adaptive filtration", Stoch. Processes Appl., *July, 173, 104381*

Fisher information matrix is

$$\mathbf{I}(\vartheta) = \int_0^T \frac{S(\vartheta,t)}{2\sigma(t)} \left[\frac{\partial}{\partial\vartheta}\ln S(\vartheta,t)\right]\left[\frac{\partial}{\partial\vartheta}\ln S(\vartheta,t)\right]^\top \mathrm{d}t.$$

Here $S(\vartheta,t) = f(\vartheta,t)\,b(\vartheta,t)$.

The MLE $\hat\vartheta_\varepsilon$ and BE $\tilde\vartheta_\varepsilon$ are consistent, and asymptotically normal, i.e.,

$$\frac{(\hat\vartheta_\varepsilon - \vartheta_0)}{\sqrt{\varepsilon}} \Longrightarrow \zeta \sim \mathcal{N}\left(0, \mathbf{I}(\vartheta_0)^{-1}\right), \qquad \frac{(\tilde\vartheta_\varepsilon - \vartheta_0)}{\sqrt{\varepsilon}} \Longrightarrow \zeta.$$

One-step MLE-process $\vartheta^\star_{t,\varepsilon}, \tau < t < T$

$$\vartheta^\star_{t,\varepsilon} = \check\vartheta_{\tau_\varepsilon} + \mathbf{I}^t_{\tau_\varepsilon}(\check\vartheta_{\tau_\varepsilon})^{-1}\int_{\tau_\varepsilon}^t \frac{\dot M(\check\vartheta_{\tau_\varepsilon},s)}{\varepsilon\sigma(s)^2}\left[\mathrm{d}X_s - M(\check\vartheta_{\tau_\varepsilon},s)\mathrm{d}s\right]$$

This estimator is used for adaptive filtration.

1.4 Hidden Telegraph process.

The observations are

$$\mathrm{d}X_t = Y_t\,\mathrm{d}t + \mathrm{d}W_t, \quad X_0 = 0, \quad 0 \le t \le T$$

where $Y_t, 0 \le t \le T$ is *telegraph process* and $W_t, 0 \le t \le T$ is an independent of $Y_t, 0 \le t \le T$ Wiener process. Recall that $Y_t, t \ge 0$ is a memory-less continuous-time stochastic Markov process that shows two distinct values $y_1 = a$ and $y_2 = b$. Parameter $\vartheta = (\lambda, \mu)$. This process can be described by the transition probabilities $P_{y_i y_j}(t) = \mathbf{P}\left(Y_t = y_j | Y_0 = y_i\right)$:

$$P_{aa}(t) = \frac{\mu}{\lambda+\mu} + \frac{\lambda}{\lambda+\mu} e^{-(\lambda+\mu)t}, \quad P_{ba}(t) = \frac{\lambda}{\lambda+\mu} - \frac{\lambda}{\lambda+\mu} e^{-(\lambda+\mu)t},$$

$$P_{ab}(t) = \frac{\mu}{\lambda+\mu} - \frac{\mu}{\lambda+\mu} e^{-(\lambda+\mu)t}, \quad P_{bb}(t) = \frac{\lambda}{\lambda+\mu} + \frac{\mu}{\lambda+\mu} e^{-(\lambda+\mu)t}.$$

The stochastic process $Y_t, t \geq 0$ admits the representation

$$\mathrm{d}X_t = m\left(\vartheta, t\right)\, \mathrm{d}t + \mathrm{d}\bar{W}_t, \qquad X_0, \qquad 0 \leq t \leq T,$$

where $m\left(\vartheta, t\right) = \mathbf{E}_\vartheta \left(Y_t | \, \mathfrak{F}_t^X\right)$ is the conditional expectation. Let us denote $\pi\left(\vartheta, t\right) = \mathbf{P}_\vartheta \left(Y_t = a | \, \mathfrak{F}_t^X\right), \mathbf{P}_\vartheta \left(Y_t = b | \, \mathfrak{F}_t^X\right) = 1 - \pi\left(\vartheta, t\right).$ Then $m\left(\vartheta, t\right) = b + \left(a - b\right)\pi\left(\vartheta, t\right).$

The random process $\pi\left(\vartheta, \cdot\right)$ satisfies the following equation

$$\begin{aligned} \mathrm{d}\pi\left(\vartheta, t\right) = & \left[\mu - \left(\lambda + \mu\right)\pi\left(\vartheta, t\right)\right. \\ & \left. + \pi\left(\vartheta, t\right)\left(1 - \pi\left(\vartheta, t\right)\right)\left(b - a\right)\left(b + \left(a - b\right)\pi\left(\vartheta, t\right)\right)\right] \mathrm{d}t \\ & + \pi\left(\vartheta, t\right)\left(1 - \pi\left(\vartheta, t\right)\right)\left(a - b\right)\, \mathrm{d}X_t. \end{aligned}$$

Chigansky, P. (2009). "Maximum likelihood estimator for hidden Markov models in continuous time". SISP. 12, 139-163.

Khasminskii, R. Z. and K (2018) "On parameter estimation of hidden telegraph process". Bernoulli, 24, 3, 2064-2090.

Having a preliminary MME $\check\vartheta_{\tau_\delta}$ we propose One-step MLE-process as follows $\left(\tau_\delta = T^\delta, \delta \in \left(\frac{1}{2}, 1\right)\right)$

$$\vartheta^\star_{t,T} = \check\vartheta_{\tau_\delta} + t^{-1}\mathbb{I}_t(\check\vartheta_{\tau_\delta})^{-1}\int_{\tau_\delta}^{t} \dot m(\check\vartheta_{\tau_\delta}, s)\,\left[{\rm d}X_s - m(\check\vartheta_{\tau_\delta}, s){\rm d}s\right].$$

Here the vector

$$\dot m(\vartheta, s) = (a-b)\,\frac{\partial \pi_\lambda\left(s,\vartheta\right)}{\partial\vartheta} = (a-b)\left(\frac{\partial\pi\left(t,\vartheta\right)}{\partial\lambda}, \frac{\partial\pi\left(t,\vartheta\right)}{\partial\mu}\right)^\top$$

and the empirical Fisher information matrix $\mathbb{I}_t(\vartheta)$ is

$$\mathbb{I}_t(\vartheta) = \frac{1}{t}\int_{\tau_\delta}^{t} \dot m(\vartheta, s)\dot m(\vartheta, s)^\top {\rm d}s \longrightarrow \mathbb{I}(\vartheta)$$

as $t\to\infty$ by the law of large numbers. Here $\mathbb{I}(\vartheta)$ is the Fisher information matrix

$$\mathbb{I}(\vartheta) = (a-b)^2\,\mathbf{E}_\vartheta\,\frac{\partial\pi\left(s,\vartheta\right)}{\partial\vartheta}\,\frac{\partial\pi\left(s,\vartheta\right)}{\partial\vartheta}^\top.$$

1.5 Hidden AR process

Consider the model of partially observed time series

$$
\begin{aligned}
X_t &= f\,Y_{t-1} + \sigma\,w_t, \qquad && X_0, \qquad && t = 1, 2, \ldots, \\
Y_t &= a\,Y_{t-1} + b\,v_t, \qquad && Y_0,
\end{aligned}
$$

where $X^T = (X_0, X_1, \ldots, X_T)$ are observations and $Y_t, t \geq 0$ is a hidden AR process. Here $w_t, t \geq 1$ and $v_t, t \geq 1$ are independent standard Gaussian random variables, i.e., $w_t \sim \mathcal{N}(0,1)$, $v_t \sim \mathcal{N}(0,1)$. The initial values are $X_0 \sim \mathcal{N}\left(0, d_x^2\right)$ and $Y_0 \sim \mathcal{N}\left(0, d_y^2\right)$. The system is defined by the parameters $a^2 \in [0,1), b, f, \sigma^2$. We suppose that some of these parameters are unknown and have to be estimated by observations X^T.

Kalman filter for $m_t(\vartheta) = \mathbf{E}_\vartheta(Y_t | X_0, \ldots, X_t)$ is

$$m_t(\vartheta) = a\, m_{t-1}(\vartheta) + \frac{a f \gamma_*(\vartheta)}{\sigma^2 + f^2 \gamma_*(\vartheta)} \left[X_t - f m_{t-1}(\vartheta)\right], \quad t \geq 1$$

where

$$\gamma_*(\vartheta) = \frac{f^2 b^2 - \sigma^2\left(1 - a^2\right)}{2 f^2} + \frac{1}{2}\left[\left(\frac{\sigma^2\left(1 - a^2\right)}{f^2} - b^2\right)^2 + \frac{4 b^2 \sigma^2}{f^2}\right]^{1/2}$$

We propose MMEs for a, b, f, σ^2 and consider the problems of estimation one, two and three-dimensional parameters like $\vartheta = b$, $\vartheta = (f, a)$, $\vartheta = \left(f, a, \sigma^2\right)$. Describe the properties of MLE and BE, study One-step MLE-processes for these parameters and introduce adaptive Kalman filter.

Three statistics are introduced

$$S_{1,T}\left(X^T\right) = \frac{1}{T}\sum_{t=2}^{T}\left(X_t - X_{t-1}\right)^2,$$

$$S_{2,T}\left(X^T\right) = \frac{1}{T}\sum_{t=3}^{T}\left(X_t - X_{t-1}\right)\left(X_{t-1} - X_{t-2}\right),$$

$$S_{3,T}\left(X^T\right) = \frac{1}{T}\sum_{t=4}^{T}\left(X_t - X_{t-1}\right)\left(X_{t-2} - X_{t-3}\right)$$

Denote

$$\Phi_1\left(\vartheta\right) = \frac{2f^2b^2}{1+a} + 2\sigma^2, \qquad \Phi_2\left(\vartheta\right) = \frac{f^2b^2a\left(1-a\right)}{\left(1+a\right)} - \sigma^2,$$

$$\Phi_3\left(\vartheta\right) = \frac{f^2b^2a\left(a-1\right)}{\left(1+a\right)}.$$

Preliminary MME

$$S_{1,T}\left(X^T\right)=\frac{2f^2b^2}{1+a}+2\sigma^2,\qquad S_{2,T}\left(X^T\right)=\frac{f^2b^2\left(a-1\right)}{1+a}-\sigma^2,$$
$$S_{3,T}\left(X^T\right)=\frac{f^2b^2a\left(a-1\right)}{\left(1+a\right)}.$$

The MME $\vartheta_T^*=\left(a_T^*,f_T^*,\sigma_T^{2*}\right)$ is the following solution of this system:

$$a_T^*=\frac{2S_{3,T}\left(X^T\right)}{S_{1,T}\left(X^T\right)+2S_{2,T}\left(X^T\right)}+1,\qquad f_T^*=\frac{S_{3,T}\left(X^T\right)\left(1+a_T^*\right)}{b^2a_T^*\left(a_T^*-1\right)},$$
$$\sigma_T^{2*}=\frac{1}{2}S_{1,T}\left(X^T\right)-\frac{\left(f_T^*\right)^2b^2}{1+a_T^*}.$$

Fisher informations.

Parameter $\vartheta = b$:

$$\mathrm{I}_b\left(\vartheta_0\right) = \frac{\dot{P}_b\left(\vartheta_0\right)^2\left[P\left(\vartheta_0\right)^2 + a^2\sigma^4\right]}{2P\left(\vartheta_0\right)^2\left[P\left(\vartheta_0\right)^2 - a^2\sigma^4\right]}.$$

Here $P\left(\vartheta\right) = \sigma^2 + f^2\gamma_*\left(\vartheta\right)$

Parameter $\vartheta = \sigma^2$:

$$\mathrm{I}_{\sigma^2}\left(\vartheta_0\right) = \frac{a^2\left[\vartheta_0\dot{\Gamma}_{\sigma^2}\left(\vartheta_0\right) - \Gamma\left(\vartheta_0\right)\right]^2}{P\left(\vartheta_0\right)^4\left(1 - A\left(\vartheta_0\right)^2\right)} + \frac{\dot{P}_{\sigma^2}\left(\vartheta_0\right)^2}{2P\left(\vartheta_0\right)^2}.$$

Here

$$\Gamma\left(\vartheta\right) = f^2\gamma_*\left(\vartheta\right), \qquad A\left(\vartheta\right) = \frac{a\vartheta}{P\left(\vartheta\right)}$$

Parameter $\vartheta = a$. **Parameter** $\vartheta = \left(f, a, \sigma^2\right)$.

Unknown parameter $\vartheta = \left(f, \sigma^2\right)$

The unknown parameter $\vartheta = \left(f, \sigma^2\right) = (\theta_1, \theta_2)$ and the partially observed system is

$$
\begin{aligned}
X_t &= \theta_1 \, Y_{t-1} + \sqrt{\theta_2}\, w_t, \qquad X_0, \qquad t = 1, 2, \ldots, \\
Y_t &= a \, Y_{t-1} + b \, v_t, \qquad\qquad Y_0.
\end{aligned}
$$

For the Fisher information matrix we have the representation

$$
\mathbf{I}\left(\vartheta_0\right) = \begin{pmatrix} I_{11}\left(\vartheta_0\right), & I_{12}\left(\vartheta_0\right) \\ I_{21}\left(\vartheta_0\right), & I_{22}\left(\vartheta_0\right) \end{pmatrix},
$$

where

$$I_{11}(\vartheta_0) = \frac{\dot{B}_f(\vartheta_0, \vartheta_0)^2}{P(\vartheta_0)\left(1 - A(\vartheta_0)^2\right)} + \frac{\dot{P}_f^2}{2P(\vartheta_0)^2},$$

$$I_{12}(\vartheta_0) = \frac{\dot{B}_f(\vartheta_0, \vartheta_0)\,\dot{B}_{\sigma^2}(\vartheta_0, \vartheta_0)}{P(\vartheta_0)\left(1 - A(\vartheta_0)^2\right)} + \frac{\dot{P}_f\,\dot{P}_{\sigma^2}}{2P(\vartheta_0)^2},$$

$$I_{22}(\vartheta_0) = \frac{\dot{B}_{\sigma^2}(\vartheta_0, \vartheta_0)^2}{P(\vartheta_0)\left(1 - A(\vartheta_0)^2\right)} + \frac{\dot{P}_{\sigma^2}^2}{2P(\vartheta_0)^2}.$$

and always

$$\mathrm{Det}\left(\mathbf{I}(\vartheta_0)\right) \neq 0.$$

One-step MLE-process. Consider the construction of a one-step MLE-process in the case of unknown parameter $\vartheta = b$. Fix a learning interval $X^{\tau_T} = \left(X_0, X_1, \ldots, X_{\tau_T}\right)$, where $\tau_T = \left[T^\delta\right], \delta \in \left(\frac{1}{2}, 1\right)$. The preliminary estimator is $\vartheta^*_{\tau_T} = b^*_{\tau_T}$

$$\vartheta^\star_{t,T} = \vartheta^*_{\tau_T} + \frac{\mathrm{I}_b(\vartheta^*_{\tau_T})^{-1}}{\left(t - \tau_T\right)} \sum_{s=\tau_T+1}^{t} \left[\frac{\left[X_s - fm_{s-1}(\vartheta^*_{\tau_T})\right]}{P(\vartheta^*_{\tau_T})} f\dot{m}_{s-1}(\vartheta^*_{\tau_T}) \right.$$

$$\left. + \left(\left[X_s - fm_{s-1}(\vartheta^*_{\tau_T})\right]^2 - P(\vartheta^*_{\tau_T})\right) \frac{\dot{P}(\vartheta^*_{\tau_T})}{2P(\vartheta^*_{\tau_T})^2} \right].$$

Here $t \in \left[\tau_T + 2, T\right]$.

For any fixed $v \in (0,1)$ and $t = vT$ uniformly on $\vartheta_0 \in \mathbb{K}$

$$\sqrt{t}\left(\vartheta^\star_{t,T} - \vartheta_0\right) \Longrightarrow \mathcal{N}\left(0, \mathrm{I}_b(\vartheta_0)^{-1}\right) \qquad t\mathbf{E}_{\vartheta_0}\left(\vartheta^\star_{t,T} - \vartheta_0\right)^2 \longrightarrow \mathrm{I}_b(\vartheta_0)^{-1}.$$

Let us write this estimator in recurrent form too:

$$
\begin{aligned}
\vartheta^{\star}_{t,T} = \frac{\vartheta^*_{\tau_T}}{t-\tau_T} + \left(1 - \frac{1}{t-\tau_T}\right)\ \vartheta^{\star}_{t-1,T} \\
+ \frac{\Big[X_t - fm_{t-1}(\vartheta^*_{\tau_T})\Big]\, f\dot{m}_{t-1}(\vartheta^*_{\tau_T})}{\mathrm{I}_b(\vartheta^*_{\tau_T})\,\big(t-\tau_T\big)\,P(\vartheta^*_{\tau_T})} \\
+ \frac{\left(\Big[X_t - fm_{t-1}(\vartheta^*_{\tau_T})\Big]^2 - P(\vartheta^*_{\tau_T})\right)\dot{P}(\vartheta^*_{\tau_T})}{2\mathrm{I}_b(\vartheta^*_{\tau_T})\,\big(t-\tau_T\big)\,P(\vartheta^*_{\tau_T})^2}, \\
t \in \big[\tau_T + 1, T\big]\,.
\end{aligned}
$$

The random sequences $m_{s-1}(\vartheta^*_{\tau_T})$ and $\dot m_{s-1}(\vartheta^*_{\tau_T})$ are

$$
\begin{aligned}
m_{s-1}(\vartheta^*_{\tau_T}) &= P(\vartheta^*_{\tau_T})^{-1}\left[a\sigma^2 m_{s-2}(\vartheta^*_{\tau_T}) + af\gamma_*(\vartheta^*_{\tau_T})X_{s-1}\right],\\
\dot m_{s-1}(\vartheta^*_{\tau_T}) &= P(\vartheta^*_{\tau_T})^{-2}a\sigma^2\left[P(\vartheta^*_{\tau_T})\dot m_{s-2}(\vartheta^*_{\tau_T}, s-2)\right.\\
&\qquad\left. - f^2\dot\gamma_*(\vartheta^*_{\tau_T})\, m_{s-2}(\vartheta^*_{\tau_T})\right] + \frac{af\sigma^2\dot\gamma_*(\vartheta^*_{\tau_T})}{P(\vartheta^*_{\tau_T})^2}\,X_{s-1}
\end{aligned}
$$

Adaptive filter.

$$
m^\star_{t,T} = P(\vartheta^\star_{t-1,T})^{-1}\left[a\sigma^2 m^\star_{t-1,T} + af\gamma_*(\vartheta^\star_{t-1,T})X_t\right],\ t\in\left[\tau_T+1, T\right].
$$

Theorem 1. *Let* $t = [vT], v \in (0,1],\ k = t - \tau_T + 2$ *and* $T \to \infty$. *Then the following relations hold*

$$\sqrt{t}\left[m^{\star}_{t,T} - m_t(\vartheta_0)\right] \approx \dot{B}^*(\vartheta_0,\vartheta_0)\sum_{m=0}^{k} A(\vartheta_0)^m\,\eta_{t-m-1,T}\;\zeta_{t-m}(\vartheta_0),$$

$$t\mathbf{E}_{\vartheta_0}\left[m^{\star}_{t,T} - m_t(\vartheta_0)\right]^2 \longrightarrow S^{\star}_b(\vartheta_0)^2 \equiv \frac{\dot{B}^*(\vartheta_0,\vartheta_0)^2}{\mathrm{I}_b(\vartheta_0)\left(1 - A(\vartheta_0)^2\right)}.$$

Notation

$$A(\vartheta_0) = \frac{a\sigma^2}{\sigma^2 + f^2\gamma_*(\vartheta_0)}, \qquad B^*(\vartheta_0,\vartheta_0) = \frac{a\sigma^2 f\dot{\gamma}_*(\vartheta_0)}{\left[\sigma^2 + f^2\gamma_*(\vartheta_0)\right]^{3/2}},$$

$$S^{\star}_b(\vartheta)^2 = \frac{\dot{B}^*(\vartheta,\vartheta)^2}{\mathrm{I}(\vartheta)\left(1 - A(\vartheta)^2\right)}.$$

Asymptotic efficiency

The lower bound is $(t = vT, v \in (0,1], T \to \infty)$

$$\lim_{\nu\to 0}\varliminf_{t\to\infty}\sup_{\|\vartheta-\vartheta_0\|\leq\nu} t\mathbf{E}_\vartheta\left(\bar m_t - m\left(\vartheta,t\right)\right)^2 \geq \mathcal{S}_b^\star\left(\vartheta_0\right)^2$$

The adaptive filter $m^*_{t,T}, \tau_T \leq t \leq T$ is called asymptotically efficient if for all $\vartheta_0 \in \Theta$ and $t \in (\tau_T, T]$

$$\lim_{\nu\to 0}\lim_{t\to\infty}\sup_{\|\vartheta-\vartheta_0\|\leq\nu} t\mathbf{E}_\vartheta\left(m^*_{t,T} - m\left(\vartheta,t\right)\right)^2 = \mathcal{S}_b^\star\left(\vartheta_0\right)^2$$

The adaptive filter $m^\star_{t,T}$ is asymptotically efficient.

K (2024) "Hidden AR process and adaptive Kalman filter". To appear in AISM.

1.6 Hidden O-U process, Ergodic case

We are given a linear system with constant coefficients

$$\mathrm{d}X_t = f(\vartheta)\, Y_t\, \mathrm{d}t + \sigma\, \mathrm{d}W_t, \qquad X_0, \qquad 0 \le t \le T,$$
$$\mathrm{d}Y_t = -a(\vartheta)\, Y_t\, \mathrm{d}t + b(\vartheta)\, \mathrm{d}V_t, \qquad Y_0.$$

Here $f(\vartheta) \neq 0$, $b(\vartheta) \neq 0$, $\sigma^2 > 0$ and $a(\vartheta) > 0$. Asymptotic $T \to \infty$. Examples: $(\mathbf{F}) : f(\vartheta) = \vartheta, \quad a(\vartheta) = a, \quad b(\vartheta) = b,$ $(\mathbf{A}) : f(\vartheta) = f, \quad a(\vartheta) = \vartheta, \quad b(\vartheta) = b$, $(\mathbf{F}, \mathbf{B})$ forbidden, $(\mathbf{F}, \mathbf{A}) : f(\vartheta) = \theta_1, \quad a(\vartheta) = \theta_2, \quad b(\vartheta) = b, i.e.\ \vartheta = (\theta_1, \theta_2).$

K (2019) "On parameter estimation of hidden ergodic Ornstein-Uhlenbeck process" Electr. J. of Stat., *13, 4508-4526.*

K (2024) "Hidden ergodic Ornstein-Uhlenbeck process and adaptive filtration" submitted.

We are given a linear system

$$\mathrm{d}X_t = f\,Y_t\,\mathrm{d}t + \sigma\,\mathrm{d}W_t, \qquad X_0, \qquad 0 \leq t \leq T,$$
$$\mathrm{d}Y_t = -a\,Y_t\,\mathrm{d}t + b\,\mathrm{d}V_t, \qquad Y_0, \qquad t \geq 0.$$

For instant we denote $\vartheta = (f, a, b)$.

The equations of Kalman-Bucy filtration are

$$\mathrm{d}m\,(\vartheta, t) = -\left[a + \frac{\gamma\,(\vartheta, t)\,f^2}{\sigma^2}\right] m\,(\vartheta, t)\,\mathrm{d}t + \frac{\gamma\,(\vartheta, t)\,f}{\sigma^2}\mathrm{d}X_t,$$
$$\frac{\partial \gamma\,(\vartheta, t)}{\partial t} = -2a\gamma\,(\vartheta, t) - \frac{\gamma\,(\vartheta, t)^2\,f^2}{\sigma^2} + b^2,$$

with the initial values $m_0 = \mathbf{E}_\vartheta\,(Y_0|X_0)$, $\gamma_0 = \mathbf{E}_\vartheta\,(Y_0 - m\,(\vartheta, 0))^2$.

Recall that Riccati equation has an explicit solution

$$\gamma(\vartheta,t)=e^{-2r(\vartheta)t}\left[\frac{1}{\gamma(\vartheta,0)-\gamma_*(\vartheta)}+\frac{f^2\left(1-e^{-2r(\vartheta)t}\right)}{2r(\vartheta)\,\sigma^2}\right]^{-1}+\gamma_*(\vartheta),$$

where

$$r(\vartheta)=\left(a^2+\frac{f^2b^2}{\sigma^2}\right)^{1/2},\qquad\qquad \gamma_*(\vartheta)=\frac{\sigma^2\left[r(\vartheta)-a\right]}{f^2},$$

and we suppose that $\gamma(\vartheta,0)\neq\gamma_*(\vartheta)$. Note that there exists a constant $C>0$ such that

$$\left|\gamma(\vartheta,t)-\gamma_*(\vartheta)\right|\leq C\,e^{-2r(\vartheta)t}\longrightarrow 0,\qquad \text{as}\quad t\longrightarrow\infty.$$

This exponential convergence essentially simplifies the calculations. We suppose that $\gamma(\vartheta,0)=\gamma_*(\vartheta)$ and therefore $\gamma(\vartheta,t)\equiv\gamma_*(\vartheta)$ for all $t\geq 0$.

1.7 Method of moments estimators.

At the beginning we suppose that $\vartheta = (f, a, b) \in \Theta$. The set Θ is a bounded, convex, closed subset of $\mathcal{R}^3$ and the values of a are positive and separated from zero. For simplicity of exposition we suppose that T is integer.

Introduce the notation

$$R_{1,T} = \frac{1}{T}\sum_{k=1}^{T}\left[X_k - X_{k-1}\right]^2, \qquad \Phi_1\left(\vartheta\right) = \frac{f^2b^2}{a^3}\left[e^{-a} - 1 + a\right] + \sigma^2,$$

$$R_{2,T} = \frac{1}{T}\sum_{k=2}^{T}\left[X_k - X_{k-1}\right]\left[X_{k-1} - X_{k-2}\right], \ \Phi_2\left(\vartheta\right) = \frac{f^2b^2}{2a^3}\left[1 - e^{-a}\right]^2,$$

$$R_T = \left(R_{1,T}, R_{2,T}\right)^\top, \qquad \Phi\left(\vartheta\right) = \left(\Phi_1\left(\vartheta\right), \Phi_2\left(\vartheta\right)\right)^\top$$

and $\xi_* = (\xi_1, \xi_2)^\top \sim \mathcal{N}(0, \mathbf{K}(\vartheta_0))$, where

$$\mathbf{K}(\vartheta) = \begin{pmatrix} K_{1,12}(\vartheta) & K_{1,2}(\vartheta) \\ K_{2,1}(\vartheta) & K_{2,2}(\vartheta) \end{pmatrix},$$

$$K_{1,1}(\vartheta) = \frac{2f^4b^4}{a^6}\left[e^{-a} - 1 + a\right]^2 + \frac{f^4b^4e^{4a}\left[1 - e^{-a}\right]^3}{a^6\left[1 + e^{-a}\right]} + \frac{4f^2\sigma^2b^2}{a^3}\left[e^{-a} - 1 + a\right] + 2\sigma^4$$

$$K_{1,2}(\vartheta) = \ldots,$$

$$K_{2,2}(\vartheta) = \ldots$$

Proposition 1. *The statistic R_T has the following properties:*

1. *Uniformly on Θ it converges to $\Phi(\vartheta)$, i.e., for any $\nu > 0$*

$$\lim_{T\to\infty} \sup_{\vartheta\in\Theta} \mathbf{P}_\vartheta \left(\left\|R_T - \Phi(\vartheta)\right\| \geq \nu\right) = 0.$$

2. *Is uniformly on compacts $\mathbb{K} \subset \Theta$ asymptotically normal*

$$\sqrt{T}\left(R_T - \Phi(\vartheta)\right) \Longrightarrow \xi_*,$$

3. *The moments converge: for any $p > 0$ uniformly on $\mathbb{K}$*

$$\lim_{T\to\infty} T^{p/2}\mathbf{E}_\vartheta \left\|R_T - \Phi(\vartheta)\right\|^p = \mathbf{E}_\vartheta \left|\xi_*\right|^p,$$

and there exists a constant $C > 0$ such that

$$\sup_{\vartheta\in\mathbb{K}} T^{p/2}\mathbf{E}_\vartheta \left\|R_T - \Phi(\vartheta)\right\|^p \leq C.$$

1.7.1 Estimation of the parameter f.

Suppose that $\vartheta = f \in \Theta = (\alpha, \beta)$, $\alpha > 0$ and the values of $a > 0$ and $b \neq 0$ are known. Note that in this case

$$\vartheta = \left(\frac{\left(\Phi_1(\vartheta) - \sigma^2\right) a^3}{b^2 \left[e^{-a} - 1 + a\right]} \right)^{1/2}.$$

Hence the MME ϑ_T^* can be defined as follows. Let us denote

$$\bar{\vartheta}_T = \left(\frac{a^3}{b^2 \left[e^{-a} - 1 + a\right] T} \sum_{k=1}^{T} \left(\left[X_k - X_{k-1}\right]^2 - \sigma^2 \right) \right)^{1/2}$$

and set

$$\vartheta_T^* = \alpha \mathbb{1}_{\{R_{1,T} \leq \Phi_1(\alpha)\}} + \bar{\vartheta}_T \mathbb{1}_{\{\Phi_1(\alpha) \leq R_{1,T} \leq \Phi_1(\beta)\}} + \beta \mathbb{1}_{\{R_{1,T} \geq \Phi_1(\beta)\}}.$$

The true value is denoted ϑ_0 and we set

$$D_f\left(\vartheta\right)^2=\frac{a^3\;K_{1,1}\left(\vartheta\right)}{4b^2\left(e^{-a}-1+a\right)\left(\Phi_1\left(\vartheta\right)-\sigma^2\right)}.$$

Proposition 2. *The MME ϑ^*_T is uniformly on compacts $\mathbb{K}\subset\Theta$ consistent, asymptotically normal*

$$\sqrt{T}\left(\vartheta^*_T-\vartheta_0\right)\Longrightarrow\xi_f\sim\mathcal{N}\left(0,D_f\left(\vartheta_0\right)^2\right),$$

and the polynomial moments converge: for any $p\geq 2$

$$T^{p/2}\mathbf{E}_{\vartheta_0}\left|\vartheta^*_T-\vartheta_0\right|^p\longrightarrow\mathbf{E}_{\vartheta_0}\left|\xi_1\right|^p.$$

1.7.2 Estimation of the parameter b.

Suppose that $\vartheta = b \in \Theta = (\alpha, \beta)\,, \alpha > 0$ and the values of $a > 0$ and $f \neq 0$ are known. In this case

$$\vartheta = \left(\frac{\left(\Phi_1(\vartheta) - \sigma^2\right) a^3}{f^2\left[e^{-a} - 1 + a\right]}\right)^{1/2},$$

$$\bar\vartheta_T = \left(\frac{a^3}{f^2\left[e^{-a} - 1 + a\right] T} \sum_{k=1}^{T} \left(\left[X_k - X_{k-1}\right]^2 - \sigma^2\right)\right)^{1/2}.$$

Proposition 3. *The MME ϑ_T^* of the parameter $\vartheta = b$ is uniformly on compacts $\mathbb{K} \subset \Theta$ consistent, asymptotically normal*

$$\sqrt{T}\left(\vartheta_T^* - \vartheta_0\right) \Longrightarrow \xi_b \sim \mathcal{N}\left(0, D_b(\vartheta_0)^2\right),$$

and the polynomial moments converge: for any $p \geq 2$

$$T^{p/2} \mathbf{E}_{\vartheta_0} \left|\vartheta_T^* - \vartheta_0\right|^p \longrightarrow \mathbf{E}_{\vartheta_0} \left|\xi_b\right|^p.$$

1.7.3 Estimation of the parameter a.

Suppose that $\vartheta = a \in \Theta = (\alpha, \beta)\,, \alpha > 0$ and the values of $f \neq 0$ and $b \neq 0$ are known.

Let us verify that the function

$$\Phi_1(\vartheta) = \frac{f^2 b^2}{\vartheta^3}\left[e^{-\vartheta} - 1 + \vartheta\right] + \sigma^2, \qquad \vartheta \in \Theta$$

is strictly decreasing.

Introduce the function $h(x) = x^{-3}\left(x - 1 + e^{-x}\right), x > 0$. Then for its derivative we have the expression
$h'(x) = x^{-4}\left[3 - 2x - (3 + x)\, e^{-x}\right]$. Remark, that the function $g(x) = 3 - 2x - (3 + x)\, e^{-x}$ at $x = 0$ is $g(0) = 0$ and has derivative $g'(x) = -2 + (2 + x)\, e^{-x}$ with $g'(0) = 0$. At last $g''(x) = -(1 + x)\, e^{-x} < 0$. Hence $g'(x) < 0$, $g(x) < 0$ and $h'(x) < 0$ for $x > 0$.

As the function $\Phi_1(\vartheta), \vartheta \in \Theta$ is strictly decreasing, the equation

$$R_{1,T} = \Phi_1(\vartheta_T^*), \qquad \vartheta_T^* \in \Theta$$

has a unique solution. Introduce the function $H(y)$ inverse to the function $h(x)$, i.e., $H(h(x)) = x, x > 0$. Then we can write

$$\vartheta_T^* = H\left(\frac{R_{1,T} - \sigma^2}{f^2 b^2}\right) = \vartheta_0 + \frac{R_{1,T} - \Phi_1(\vartheta_0)}{h'(\vartheta_0) f^2 b^2} + O\left(T^{-1}\right).$$

Hence

$$\sqrt{T}\left(\vartheta_T^* - \vartheta_0\right) \Longrightarrow \xi_a \sim \mathcal{N}\left(0, D_a(\vartheta_0)^2\right),$$

where

$$D_a(\vartheta_0)^2 = \frac{K_{1,1}(\vartheta_0)}{h'(\vartheta_0)^2 f^4 b^4}.$$

1.7.4 More general model

All considered cases of estimation of one-dimensional parameters are particular cases of a slightly more general model of observations

$$
\begin{aligned}
\mathrm{d}X_t &= f(\vartheta)\, Y_t \mathrm{d}t + \sigma \mathrm{d}W_t, && X_0, \qquad t \geq 0, \\
\mathrm{d}Y_t &= -a(\vartheta)\, Y_t \mathrm{d}t + b(\vartheta)\, \mathrm{d}V_t, && Y_0,
\end{aligned}
$$

where $f(\vartheta), a(\vartheta), b(\vartheta), \vartheta \in \Theta = (\alpha, \beta)$ are known smooth functions. To estimate the parameter ϑ by observations $X^T = (X_t, 0 \leq t \leq T)$ we can use the MME

$$
\vartheta_T^* = \arg\inf_{\vartheta \in \Theta} \left| R_{1,T} - \Psi(\vartheta) \right|, \qquad R_{1,T} = \frac{1}{T} \sum_{k=1}^{T} \left[X_k - X_{k-1} \right]^2,
$$

$$
\Psi(\vartheta) = \frac{f(\vartheta)^2\, b(\vartheta)^2}{a(\vartheta)^2} \left[e^{-a(\vartheta)} - 1 + a(\vartheta) \right],
$$

Proposition 4. *Let the functions* $f(\vartheta), a(\vartheta), b(\vartheta), \vartheta \in [\alpha, \beta]$ *have two continuous derivatives, and*

$$\inf_{\vartheta\in\Theta} \left|\dot\Psi(\vartheta)\right| > 0.$$

Then the MME $\check\vartheta_T$ *is uniformly consistent and for any* $p \geq 2$

$$\sup_{\vartheta_0\in\mathbb{K}} T^{p/2}\mathbf{E}_{\vartheta_0}\left|\check\vartheta_T - \vartheta_0\right|^p \leq C.$$

1.7.5 Two-dimensional parameter $\vartheta = (a, f)$

Suppose that

$$\vartheta = (\theta_1, \theta_2)^\top = (a, f)^\top \in \Theta = (\alpha_a, \beta_a) \times (\alpha_f, \beta_f), \quad \alpha_a > 0 \quad \alpha_f > 0.$$

The MME $\vartheta^*_T = (\theta^*_{1,T}, \theta^*_{2,T})^\top = (a^*_T, f^*_T)^\top$ is defined as solution $\vartheta = \vartheta^*_T$ of the system of equations

$$\Phi_1(\vartheta) = R_{1,T}, \qquad \Phi_2(\vartheta) = R_{2,T}, \qquad \vartheta \in \Theta$$

Recall that

$$\Phi_1(\vartheta) = \frac{f^2 b^2}{a^3}\left[e^{-a} - 1 + a\right] + \sigma^2, \qquad \Phi_2(\vartheta) = \frac{f^2 b^2}{2a^3}\left[1 - e^{-a}\right]^2.$$

This system can be solved in two steps as follows: the equation

$$\frac{e^{-a_T^*} - 1 + a_T^*}{\left(1 - e^{-a_T^*}\right)^2} = \frac{R_{1,T} - \sigma^2}{2R_{2,T}},$$

can be solved the first and then having solution a_T^* of this equation we define the second estimator

$$f_T^* = \frac{2\left(a_T^*\right)^{3/2} R_{2,T}^{1/2}}{b^2 \left(1 - e^{-a_T^*}\right)^2}.$$

Let us denote $\xi^* = (\xi_3, \xi_4)^\top$ the Gaussian vector with the corresponding covariance matrix $\mathbf{Q}(\vartheta_0)$.

Proposition 5. *The MME ϑ_T^* is uniformly consistent, uniformly on compacts $\mathbb{K} \subset \Theta$ asymptotically normal*

$$\sqrt{T}\left(\vartheta_T^* - \vartheta_0\right) \Longrightarrow \xi^* \sim \mathcal{N}\left(0, \mathbf{Q}\left(\vartheta_0\right)\right),$$

the polynomial moments converge and the estimate

$$\sup_{\vartheta_0 \in \mathbb{K}} T^{p/2} \mathbf{E}_{\vartheta_0} \left\| \vartheta_T^* - \vartheta_0 \right\|^p \leq C$$

holds.

1.8 MLE and BE

We are given a two-dimensional diffusion process $\{X_t, Y_t,\ t \geq 0\}$

$$
\begin{aligned}
\mathrm{d}Y_t &= -\,a\,(\vartheta)\,Y_t\,\mathrm{d}t + b\,(\vartheta)\ \mathrm{d}V_t, \qquad && Y_0,\\
\mathrm{d}X_t &= \quad f\,(\vartheta)\,Y_t\,\mathrm{d}t + \ \sigma\,\mathrm{d}W_t, \qquad && X_0 = 0,
\end{aligned}
$$

This model of observations allows us to study estimators in the five different situations:

F. $f\,(\vartheta) = \vartheta, \quad a\,(\vartheta) = a, \quad b\,(\vartheta) = b,$

A. $f\,(\vartheta) = f, \quad a\,(\vartheta) = \vartheta, \quad b\,(\vartheta) = b,$

B. $f\,(\vartheta) = f, \quad a\,(\vartheta) = a, \quad b\,(\vartheta) = \vartheta,$

AB. $f\,(\vartheta) = f, \quad a\,(\vartheta) = \theta_1, \quad b\,(\vartheta) = \theta_2$, i.e., $\vartheta = (\theta_1, \theta_2)^\top$.

AF. $f\,(\vartheta) = \theta_2, \quad a\,(\vartheta) = \theta_1, \quad b\,(\vartheta) = b.$

The likelihood ratio function has the following form:

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

The MLE $\hat\vartheta_T$ and BE $\tilde\vartheta_T$ are defined by the relations

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

Here $M\left(\vartheta,t\right) = f\left(\vartheta\right) m\left(\vartheta,t\right)$ and $p\left(\cdot\right)$ is continuous positive density function on Θ.

To construct the MLE of the parameter ϑ we have to calculate the function $\left\{M\left(\vartheta,t\right),\ 0\leq t\leq T\right\},\ \vartheta\in\Theta$ defined by the equation

$${\rm d}M\left(\vartheta,t\right) = -r\left(\vartheta\right) M\left(\vartheta,t\right)\,{\rm d}t + \gamma\left(\vartheta\right){\rm d}X_t, \qquad\qquad M\left(\vartheta,0\right).$$

1.8.1 One-dimensional parameter

Suppose that the parameter ϑ is one-dimensional and $\Theta = (\alpha, \beta)$. Introduce the quantity

$$\mathrm{I}\left(\vartheta\right) = \frac{\dot a\left(\vartheta\right)^2}{2a\left(\vartheta\right)} - \frac{2\dot a\left(\vartheta\right)\ \dot r\left(\vartheta\right)}{r\left(\vartheta\right)+a\left(\vartheta\right)} + \frac{\dot r\left(\vartheta\right)^2}{2r\left(\vartheta\right)},$$

$$r\left(\vartheta\right) = \left(a\left(\vartheta\right)^2 + \sigma^{-2} f\left(\vartheta\right)^2 b\left(\vartheta\right)^2\right)^{1/2}$$

which will play the role of Fisher information.

Under regularity conditions given in the theorem below we have the lower bound on the mean square risks of all estimators (Hajek-Le Cam's bound)

$$\lim_{\delta\to 0} \varliminf_{T\to\infty} \sup_{\left|\vartheta-\vartheta_0\right|\le\delta} T\mathbf{E}_\vartheta \left|\bar\vartheta_T - \vartheta\right|^2 \ge \mathrm{I}\left(\vartheta_0\right)^{-1}.$$

The estimator $\bar\vartheta_T$ is as. efficient if we have equality here.

Theorem 2. *Suppose that the following conditions hold.*

1. *The functions* $a(\vartheta)$, $b(\vartheta)$, $f(\vartheta)$, $\vartheta \in [\alpha,\beta]$ *twice continuously differentiable on* $\bar\Theta$.

2. *For all* $\nu > 0$ *we have*

$$\inf_{\vartheta_0\in\Theta} \inf_{|\vartheta-\vartheta_0|>\nu} \Bigl(|a(\vartheta) - a(\vartheta_0)| + |r(\vartheta) - r(\vartheta_0)|\Bigr) > 0.$$

3. *The following condition holds* $\inf_{\vartheta\in\Theta}\Bigl(|\dot a(\vartheta)| + |\dot r(\vartheta)|\Bigr) > 0$.

Then the MLE $\hat\vartheta_T$ *and BE* $\tilde\vartheta_T$ *are uniformly consistent, uniformly on compacts* $\mathbb{K}\subset\Theta$ *asymptotically normal*

$$\sqrt{T}\left(\hat\vartheta_T - \vartheta_0\right) \Longrightarrow \zeta \sim \mathcal{N}\left(0, \mathrm{I}(\vartheta_0)^{-1}\right), \quad \sqrt{T}\left(\tilde\vartheta_T - \vartheta_0\right) \Longrightarrow \zeta,$$

polynomial moments converge and the both estimators are asymptotically efficient.

Proof. Ibragimov-Khasminskii program. Denote

$$Z_T(u) = \frac{L(\vartheta_0 + \varphi_T u, X^T)}{L(\vartheta_0, X^T)}, \quad u \in \mathbb{U}_T = \left(\sqrt{T}(\alpha - \vartheta_0), \sqrt{T}(\beta - \vartheta_0)\right)$$

and verify the properties

1. $\ln Z_T(u) = u\Delta_T(\vartheta_0) - \frac{u^2}{2}\mathrm{I}(\vartheta_0) + r_T, \quad \Delta_T(\vartheta_0) \Rightarrow \Delta(\vartheta_0),$
   $r_T \to 0.$
2. $\mathbf{E}_{\vartheta_0}\left|Z_T(u_1)^{\frac{1}{2}} - Z_T(u_2)^{\frac{1}{2}}\right|^2 \leq C\left|u_1 - u_2\right|^2$
3. $\mathbf{E}_{\vartheta_0} Z_T(u)^{\frac{1}{2}} \leq \frac{C}{|u|^N}$

Then

$$Z_T(\cdot) \Longrightarrow Z(\cdot) \qquad Z(u) = e^{u\zeta - \frac{u^2}{2}\mathrm{I}(\vartheta_0)}, \quad \text{in} \quad \mathcal{C}(R)$$

and

$$\sqrt{T}\left(\hat\vartheta_T - \vartheta_0\right) \Longrightarrow \arg\sup_{u \in R} Z(u) = \frac{\zeta}{\mathrm{I}(\vartheta_0)} \sim \mathcal{N}\left(0, \mathrm{I}(\vartheta_0)^{-1}\right)$$

1.8.2 Case F, $i.e., \vartheta = f$.

Consider the particular case $\vartheta = f$. Then $r(\vartheta) = \left(a^2 + b^2\vartheta^2\sigma^{-2}\right)^{1/2}$ and the Fisher information is

$$\mathrm{I}(\vartheta) = \frac{b^4\vartheta^2}{2\sigma^4 r(\vartheta)^3}.$$

Hence by Theorem 2 the MLE and BE are consistent, asymptotically normal

$$\sqrt{T}\left(\hat{\vartheta}_T - \vartheta_0\right) \Longrightarrow \mathcal{N}\left(0, \frac{2\sigma^4 r(\vartheta_0)^3}{b^4\vartheta_0^2}\right)$$

and asymptotically efficient.

1.8.3 Case **A**, $i.e., \vartheta = a$.

We have $r(\vartheta) = \left(\vartheta^2 + b^2 f^2 \sigma^{-2}\right)^{1/2}$ and the Fisher information is

$$\mathrm{I}(\vartheta) = \frac{\gamma(\vartheta)^2 \left[\left(r(\vartheta) + \vartheta\right)^2 + r(\vartheta)\,\vartheta\right]}{2\vartheta r(\vartheta)^3 \left[r(\vartheta) + \vartheta\right]},$$

the both estimators are asymptotically normal and asymptotically efficient. Say,

$$\sqrt{T}\left(\hat{\vartheta}_T - \vartheta_0\right) \Longrightarrow \mathcal{N}\left(0, \mathrm{I}(\vartheta_0)^{-1}\right).$$

1.8.4 Case B, $i.e., \vartheta = b$.

The Fisher information is

$$\mathrm{I}(\vartheta) = \frac{\vartheta^2 f^4}{2r(\vartheta)^3 \sigma^4}$$

and the MLE and BE have the corresponding asymptotic properties.

1.8.5 Case AB, i.e., $\vartheta = (\theta_1, \theta_2,)^\top, \theta_{1,0} = a_0, \theta_{2,0} = b_0$.

As before, we suppose that $\vartheta \in \Theta = (\alpha_a, \beta_a) \times (\alpha_b, \beta_b)$, where $\alpha_a > 0, \alpha_b > 0$. The system is

$$\begin{aligned} \mathrm{d}X_t &= fY_t \mathrm{d}t + \sigma \mathrm{d}W_t, \qquad X_0 = 0, \qquad t \geq 0 \\ \mathrm{d}Y_t &= \theta_1 Y_t \mathrm{d}t + \theta_2 \mathrm{d}V_t, \qquad Y_0, \end{aligned}$$

and the observations are $X^T = (X_t, 0 \leq t \leq T)$.

To calculate Fisher information matrix $\mathbf{I}(\vartheta)$ we first note that the values $\mathrm{I}_{1,1}(\vartheta)$ and $\mathrm{I}_{2,2}(\vartheta)$ were already calculated above

Therefore Fisher information matrix is

$$\mathbf{I}(\vartheta_0) = \begin{pmatrix} \frac{\gamma(\vartheta_0)^2\left[r(\vartheta_0)\theta_{1,0}+(r(\vartheta_0)+\theta_{1,0})^2\right]}{2\theta_{1,0}r(\vartheta_0)^3[r(\vartheta_0)+\theta_{1,0}]}, & \frac{f^2\theta_{2,0}\left(\theta_{1,0}^2+r(\vartheta_0)\theta_{1,0}-2r(\vartheta_0)\right)}{2r(\vartheta_0)^3(r(\vartheta_0)+\theta_{1,0})\sigma^2} \\ \frac{f^2\theta_{2,0}\left(\theta_{1,0}^2+r(\vartheta_0)\theta_{1,0}-2r(\vartheta_0)\right)}{2r(\vartheta_0)^3(r(\vartheta_0)+\theta_{1,0})\sigma^2}, & \frac{\theta_{2,0}^2f^4}{2r(\vartheta_0)^3\sigma^4} \end{pmatrix}$$

The determinant of the Fisher information matrix with a slightly simplified but obvious notation can be calculated as follows

$$\mathrm{Det}\left(\mathbf{I}(\vartheta_0)\right) = \frac{b^2f^4\left(r-a\right)^2}{4ar^3\left(r+a\right)^2\sigma^4} = \frac{\theta_{2,0}^2f^4\gamma\left(\vartheta_0\right)^2}{4\theta_{1,0}r\left(\vartheta_0\right)^3\left(r\left(\vartheta_0\right)+\theta_{1,0}\right)^2\sigma^4}.$$

Therefore

$$\inf_{\vartheta_0\in\Theta}\mathrm{Det}\Big(\mathbf{I}\left(\vartheta_0\right)\Big) > 0$$

Theorem 3. *The MLE* $\hat{\vartheta}_T$ *and BE* $\tilde{\vartheta}_T$ *are uniformly consistent, uniformly on compacts* $\mathbb{K} \subset \Theta$ *asymptotically normal*

$$\sqrt{T}\left(\hat{\vartheta}_T - \vartheta_0\right) \Longrightarrow \zeta, \quad \sqrt{T}\left(\tilde{\vartheta}_T - \vartheta_0\right) \Longrightarrow \zeta,$$

polynomial moments converge: for any $p \geq 2$

$$\lim_{T\to\infty} T^{p/2} \mathbf{E}_{\vartheta_0} \left\|\hat{\vartheta}_T - \vartheta_0\right\|^p = \mathbf{E}_{\vartheta_0} \left\|\zeta\right\|^p,$$

$$\lim_{T\to\infty} T^{p/2} \mathbf{E}_{\vartheta_0} \left\|\tilde{\vartheta}_T - \vartheta_0\right\|^p = \mathbf{E}_{\vartheta_0} \left\|\zeta\right\|^p,$$

and the both estimators are asymptotically efficient.

1.9 One-Step MLE-process.

Consider once more the partially observed system

$$
\begin{aligned}
{\rm d}X_t &= f\left(\vartheta\right) Y_t {\rm d}t + \sigma\, {\rm d}W_t, \qquad X_0 = 0, \qquad 0 \leq t \leq T,\\
{\rm d}Y_t &= -a\left(\vartheta\right) Y_t {\rm d}t + b\left(\vartheta\right) {\rm d}V_t, \qquad Y_0.
\end{aligned}
$$

The unknown parameter is one-dimensional, $\vartheta \in \Theta = (\alpha, \beta)$ and we have to estimate it by observations $X^T = (X_t, 0 \leq t \leq T)$.

K-B filter is

$$
{\rm d}M\left(\vartheta, t\right) = -r\left(\vartheta\right) M\left(\vartheta, t\right) {\rm d}t + \gamma\left(\vartheta\right) {\rm d}X_t,
$$

where $\gamma\left(\vartheta\right) = r\left(\vartheta\right) - a\left(\vartheta\right)$, $r\left(\vartheta\right) = \left(a\left(\vartheta\right)^2 + b(\vartheta)^2 f(\vartheta)^2 \sigma^{-2}\right)^{-1/2}$.

The MLE and BE described above have nice asymptotic properties and in particular are asymptotically efficient, but the calculation of these estimators are computationally rather difficult problems because just for one value of LR

$$L\left(\vartheta, X^T\right) = \exp\left(\int_0^T \frac{M(\vartheta,t)}{\sigma^2}\mathrm{d}X_t - \int_0^T \frac{M(\vartheta,t)^2}{2\sigma^2}\mathrm{d}t\right), \quad \vartheta\in\Theta$$

we need solutions of Kalman-Bucy equations for all or many $\vartheta\in\Theta$.

The One-step MLE-process will be constructed in two steps. First: by observations X^{τ_T} on the relatively small *learning* interval $\left[0,\tau_T\right], \tau_T = \left[T^\delta\right], \delta\in\left(\frac{1}{2},1\right)$ we construct a MME $\vartheta^*_{\tau_T}$ and then using this estimator as preliminary we propose One-step MLE-process. Here $[A]$ means the integer part of A.

Introduce the notations

$$R_{1,\tau_T} = \frac{1}{\tau_T}\sum_{k=1}^{\tau_T}\left[X_k - X_{k-1}\right]^2, \quad \vartheta^*_{\tau_T} = \arg\inf_{\vartheta\in\Theta}\left|R_{1,\tau_T} - \Psi\left(\vartheta\right)\right|,$$

$$\Psi\left(\vartheta\right) = \frac{f(\vartheta)^2 b(\vartheta)^2}{a(\vartheta)^2}\Big(e^{a(\vartheta)} - 1 + a(\vartheta)\Big) + \sigma^2,$$

$$\mathrm{I}\left(\vartheta\right) = \frac{\dot a\left(\vartheta\right)^2}{2a\left(\vartheta\right)} - \frac{2\dot a\left(\vartheta\right)\dot r\left(\vartheta\right)}{a\left(\vartheta\right) + r\left(\vartheta\right)} + \frac{\dot r\left(\vartheta\right)^2}{2r\left(\vartheta\right)}$$

and the estimator-process $\vartheta^\star_{t,T}, \tau_T < t \leq T$

$$\vartheta^\star_{t,T} = \vartheta^*_{\tau_T} + \frac{1}{\mathrm{I}\big(\vartheta^*_{\tau_T}\big)\left(t-\tau_T\right)}\int_{\tau_T}^{t}\frac{\dot M(\vartheta^*_{\tau_T},s)}{\sigma^2}\Big[\mathrm{d}X_s - M(\vartheta^*_{\tau_T},s)\mathrm{d}s\Big].$$

The random processes $M(\vartheta^*_{\tau_T},t), \dot M(\vartheta^*_{\tau_T},t), \tau_T \leq t \leq T$ satisfy the equations

$$\mathrm{d}M(\vartheta^*_{\tau_T},t) = -r(\vartheta^*_{\tau_T})M(\vartheta^*_{\tau_T},t)\mathrm{d}t + \gamma(\vartheta^*_{\tau_T})\mathrm{d}X_t,$$
$$\mathrm{d}\dot M(\vartheta^*_{\tau_T},t) = -r(\vartheta^*_{\tau_T})\dot M(\vartheta^*_{\tau_T},t)\mathrm{d}t - \dot r(\vartheta^*_{\tau_T})M(\vartheta^*_{\tau_T},t)\mathrm{d}t + \dot\gamma(\vartheta^*_{\tau_T})\mathrm{d}X_t.$$

The initial values $M(\vartheta^*_{\tau_T},\tau_T), \dot M(\vartheta^*_{\tau_T},\tau_T)$ are defined as follows.

$$\begin{aligned} M(\vartheta,\tau_T) &= M\left(\vartheta,0\right)e^{-r(\vartheta)\tau_T} + \gamma(\vartheta)e^{-r(\vartheta)\tau_T}\int_0^{\tau_T} e^{r(\vartheta)s}\mathrm{d}X_s \\ &= M\left(\vartheta,0\right)e^{-r(\vartheta)\tau_T} + \gamma(\vartheta)\left[X_{\tau_T} - r(\vartheta)\int_0^{\tau_T} e^{-r(\vartheta)\left(\tau_T - s\right)}X_s\mathrm{d}s\right], \end{aligned}$$

$$\dot M(\vartheta,\tau_T) = \ldots$$

Introduce conditions $\mathcal{A}$

1. *The functions* $f(\vartheta), a(\vartheta), b(\vartheta), \vartheta \in [\alpha, \beta]$ *are positive and twice continuously differentiable.*

2. $\inf_{\vartheta\in\Theta} |\dot\Psi(\vartheta)| > 0.$

3. $\inf_{\vartheta\in\Theta} \left(|\dot a(\vartheta)| + |\dot r(\vartheta)|\right) > 0.$

Change the variables:
$t = vT, v \in \left[\varepsilon_T, 1\right], \varepsilon_T = \tau_T T^{-1} = T^{-1+\delta} \to 0$ and denote
$\vartheta^\star_T(v) = \vartheta^\star_{vT,T}, \varepsilon_T < v \leq 1.$

Theorem 4. *Let the conditions $\mathcal{A}$ hold. Then One-step MLE-process $\vartheta^\star_T(v), \varepsilon_T < v \leq 1$ with $\delta \in (1/2, 1)$ is uniformly on compacts $\mathbb{K} \subset \Theta$ consistent: for any $\nu > 0$ and any $\varepsilon_0 \in (0, 1]$*

$$\lim_{T\to\infty} - \sup_{\vartheta_0\in\mathbb{K}} \mathbf{P}_{\vartheta_0}\left(\sup_{\varepsilon_0\leq v\leq 1}\left|\vartheta^\star_T(v) - \vartheta_0\right| > \nu\right) = 0,$$

asymptotically normal

$$\sqrt{T}\left(\vartheta^\star_T(v) - \vartheta_0\right) \Longrightarrow \zeta_v \sim \mathcal{N}\left(0, v^{-1}\mathrm{I}(\vartheta_0)^{-1}\right),$$

and the moments converge: for any $p \geq 2$

$$T^{p/2}\mathbf{E}_{\vartheta_0}\left|\vartheta^\star_T(v) - \vartheta_0\right|^p \longrightarrow \mathbf{E}_{\vartheta_0}\left|\zeta_v\right|^p.$$

The random process $\eta_T(v) = v\sqrt{T\mathrm{I}(\vartheta_0)}\left(\vartheta^\star_T(v) - \vartheta_0\right), \varepsilon_0 \leq v \leq 1$ converges in distribution in the measurable space $(\mathcal{C}[\varepsilon_0, 1], \mathfrak{B})$ to the Wiener process $W(v), \varepsilon_0 \leq v \leq 1$.

Remark 1. The One-step MLE-process can be given in the recurrent form as follows. For the values $t \in (\tau_T, T]$ we have the equality

$$\left(t-\tau_T\right)\vartheta^\star_{t,T} = \left(t-\tau_T\right)\vartheta^*_{\tau_T} + \frac{1}{\mathrm{I}\left(\vartheta^*_{\tau_T}\right)}\int_{\tau_T}^{t}\frac{\dot M(\vartheta^*_{\tau_T},s)}{\sigma^2}\left[\mathrm{d}X_s - M(\vartheta^*_{\tau_T},s)\mathrm{d}s\right].$$

Hence

$$\mathrm{d}\vartheta^\star_{t,T} = \frac{\vartheta^*_{\tau_T}-\vartheta^\star_{t,T}}{t-\tau_T}\mathrm{d}t + \frac{\dot M(\vartheta^*_{\tau_T},t)}{\mathrm{I}\left(\vartheta^*_{\tau_T}\right)\sigma^2\left(t-\tau_T\right)}\left[\mathrm{d}X_t - M(\vartheta^*_{\tau_T},t)\mathrm{d}t\right].$$

To avoid the singularities like 0/0 at the beginning $t=\tau_T$ in numerical simulations the value $t-\tau_T$ in this formula can be replaced by $t-\tau_T+\varepsilon^*$, where $\varepsilon^*>0$ is some small value. This modification, of course, will not change the asympt. results of $\vartheta^\star_{t,T}$.

1.9.1 Estimation of the parameter f.

If $f(\vartheta)=\vartheta\in\Theta=(\alpha_f,\beta_f),\alpha_f>0$, $a(\vartheta)=a>0$ and $b(\vartheta)=b>0$, then the preliminary estimator is

$$\vartheta^*_{\tau_T}=\left(\frac{a^3}{b^2\left[e^{-a}-1+a\right]\tau_T}\sum_{k=1}^{\tau_T}\left(\left[X_k-X_{k-1}\right]^2-\sigma^2\right)\right)^{1/2}.$$

The One-step MLE-process $\vartheta^\star_{t,T},\tau_T<t\leq T$ is

$$\vartheta^\star_{t,T}=\vartheta^*_{\tau_T}+\frac{2\sigma^3 r(\vartheta^*_{\tau_T})^3}{b^4\vartheta^{*2}_{\tau_T}\left(t-\tau_T\right)}\int_{\tau_T}^t\dot M(\vartheta^*_{\tau_T},s)\left[\mathrm{d}X_s-M(\vartheta^*_{\tau_T},s)\mathrm{d}s\right].$$

According to Theorem this estimator has the properties mentioned in this theorem and in particular we have the asymptotic normality

$$\sqrt{T}\left(\vartheta^\star_{\tau_T}(v)-\vartheta_0\right)\Longrightarrow\mathcal{N}\left(0,\frac{2\sigma^4 r\left(\vartheta_0\right)^3}{v\,b^4\vartheta_0^2}\right).$$

1.9.2 Estimation of the parameter b.

If $b(\vartheta) = \vartheta \in \Theta = (\alpha_b, \beta_b)\,, \alpha_b > 0,\ f(\vartheta) = f > 0$ and $a(\vartheta) = a > 0$, then the preliminary estimator and One-step MLE-process are

$$\vartheta^*_{\tau_T} = \left(\frac{a^3}{f^2\left[e^{-a} - 1 + a\right]\tau_T} \sum_{k=1}^{\tau_T} \left(\left[X_k - X_{k-1}\right]^2 - \sigma^2 \right) \right)^{1/2},$$

$$\vartheta^\star_{t,T} = \vartheta^*_{\tau_T} + \frac{2\sigma^3 r(\vartheta^*_{\tau_T})^3}{f^4\,\vartheta^{*2}_{\tau_T}\left(t - \tau_T\right)} \int_{\tau_T}^{t} \dot M(\vartheta^*_{\tau_T}, s)\left[\mathrm{d}X_s - M(\vartheta^*_{\tau_T}, s)\mathrm{d}s \right],$$

with the corresponding processes $M(\vartheta, t)\,, t \geq 0$ and $\dot M(\vartheta, t)\,, t \geq 0$.

We have as well

$$\sqrt{T}\left(\vartheta^\star_T(v) - \vartheta_0\right) \Longrightarrow \mathcal{N}\left(0, \frac{2\sigma^4 r(\vartheta_0)^3}{v\,f^4\vartheta_0^2}\right).$$

1.9.3 Estimation of the parameter a.

If $a(\vartheta)=\vartheta\in\Theta=(\alpha_a,\beta_a)\,,\alpha_a>0,\ f(\vartheta)=f>0, a(\vartheta)=a>0.$

$$\vartheta^*_{\tau_T}=\arg\inf_{\vartheta\in\Theta}\left|\frac{1}{\tau_T}\sum_{k=1}^{\tau_T}\left[X_k-X_{k-1}\right]^2-\Psi(\vartheta)\right|,$$

$$\Psi(\vartheta)=\frac{f^2b^2}{\vartheta^3}\left[e^{-\vartheta}-1+\vartheta\right]+\sigma^2.$$

The Fisher information $\mathrm{I}(\vartheta)$ was defined above

$$\mathrm{I}(\vartheta_0)=\frac{\gamma(\vartheta_0)\left[r(\vartheta_0)+\vartheta_0+r(\vartheta_0)\vartheta_0\gamma(\vartheta_0)\right]}{2\vartheta_0 r(\vartheta_0)^3\left[r(\vartheta_0)+\vartheta_0\right]},$$

and the One-step MLE-process is asymptotically normal

$$\sqrt{T}\left(\vartheta^\star_T(v)-\vartheta_0\right)\Longrightarrow\mathcal{N}\left(0,v^{-1}\mathrm{I}(\vartheta_0)^{-1}\right).$$

1.9.4 Estimation of the two-dimensional parameter (a, f).

The partially observed system is

$$\mathrm{d}X_t = \theta_2 Y_t \mathrm{d}t + \sigma \mathrm{d}W_t, \qquad X_0, \qquad 0 \le t \le T,$$
$$\mathrm{d}Y_t = -\theta_1 Y_t \mathrm{d}t + b\mathrm{d}V_t, \qquad Y_0, \qquad t \ge 0,$$

where $b \neq 0$ is known and we have to estimate
$\vartheta = (a, f) = (\theta_1, \theta_2) \in \Theta = (\alpha_a, \beta_a) \times (\alpha_f, \beta_f)$, $\alpha_a > 0, \alpha_f > 0$.

Recall that the preliminary MME $\vartheta^*_{\tau_T} = (\theta^*_{1,\tau_T}, \theta^*_{2,\tau_T}) = (a^*_{\tau_T}, f^*_{\tau_T})$ can be calculated as follows

$$\frac{e^{-a^*_{\tau_T}} - 1 + a^*_{\tau_T}}{(1 - e^{-a^*_{\tau_T}})^2} = \frac{R_{1,\tau_T} - \sigma^2}{2R_{2,\tau_T}}, \qquad R_{1,\tau_T} = \frac{1}{\tau_T} \sum_{k=1}^{\tau_T} \left[X_k - X_{k-1}\right]^2,$$

$$f^*_{\tau_T} = \frac{2(a^*_{\tau_T})^{3/2} R^{1/2}_{2,\tau_T}}{b^2(1 - e^{-a^*_{\tau_T}})^2}, \quad R_{2,\tau_T} = \frac{1}{\tau_T} \sum_{k=2}^{\tau_T} \left[X_k - X_{k-1}\right]\left[X_{k-1} - X_{k-2}\right].$$

Fisher information matrix was given above and it is non degenerate. One-step MLE-process is the vector process $\vartheta^\star_{t,T}, \tau_T < t \leq T$ defined by the expression

$$\vartheta^\star_{t,T} = \vartheta^*_{\tau_T} + \frac{\mathbf{I}(\vartheta^*_{\tau_T})^{-1}}{\sigma\left(t-\tau_T\right)} \int_{\tau_T}^{t} \dot M(\vartheta^*_{\tau_T}, s) \left[\mathrm{d}X_s - M(\vartheta^*_{\tau_T}, s)\mathrm{d}s\right].$$

Proposition 6. *One-step MLE-process $\vartheta^\star_{t,T}, \tau_T < t \leq T$ is uniformly consistent, asymptotically normal*

$$\sqrt{vT}\left(\vartheta^\star_{vT,T} - \vartheta_0\right) \Longrightarrow \zeta \sim \mathcal{N}\left(0, \mathbf{I}(\vartheta_0)^{-1}\right)$$

and the polynomial moments converge.

1.10 Adaptive filtration

Consider once more the same partially observed linear system

$$\mathrm{d}X_t = f(\vartheta)\, Y_t \mathrm{d}t + \sigma \mathrm{d}W_t, \qquad X_0, \qquad 0 \le t \le T,$$

$$\mathrm{d}Y_t = -a(\vartheta)\, Y_t \mathrm{d}t + b(\vartheta)\, \mathrm{d}V_t, \qquad Y_0, \qquad t \ge 0,$$

where $W_t, t \ge 0$ and $V_t, t \ge 0$ are independent Wiener processes. Here the functions $f(\vartheta), a(\vartheta), b(\vartheta), \vartheta \in \Theta$ are known, positive and smooth. The value of the parameter $\vartheta \in \Theta = (\alpha, \beta)$ is unknown.

The conditional expectation $M(\vartheta, t) = f(\vartheta)\, \mathbf{E}_\vartheta (Y_t | X_s, 0 \le s \le t)$ satisfies the Kalman-Bucy equations

$$\mathrm{d}M(\vartheta, t) = -\left[a(\vartheta) + \frac{\gamma(\vartheta, t)\, f(\vartheta)^2}{\sigma^2}\right] M(\vartheta, t)\, \mathrm{d}t + \frac{\gamma(\vartheta, t)\, f(\vartheta)^2}{\sigma^2} \mathrm{d}X_t,$$

The Riccati equation is

$$\frac{\partial \gamma\left(\vartheta,t\right)}{\partial t}=-2a\left(\vartheta\right)\gamma\left(\vartheta,t\right)-\frac{\gamma\left(\vartheta,t\right)^{2}f\left(\vartheta\right)^{2}}{\sigma^{2}}+b\left(\vartheta\right)^{2},\qquad \gamma\left(\vartheta,0\right),$$

The main idea is to use the One-step MLE-process on the time interval $t\in(\tau_{T},T]$

$$\vartheta_{t,T}^{\star}=\vartheta_{\tau_{T}}^{*}+\frac{\mathrm{I}(\vartheta_{\tau_{T}}^{*})^{-1}}{\left(t-\tau_{T}\right)\sigma^{2}}\int_{\tau_{T}}^{t}\dot M(\vartheta_{\tau_{T}}^{*},s)\left[{\rm d}X_{s}-M(\vartheta_{\tau_{T}}^{*},s){\rm d}s\right],$$

where $\vartheta_{\tau_{T}}^{*}$ is a MME constructed by observations $X^{\tau_{T}}=\left(X_{s},0\leq s\leq\tau_{T}\right)$, $\tau_{T}=T^{\delta},\delta\in(1/2,1)$, and to replace ϑ by $\vartheta_{t,T}^{\star}$ in the equations of filtration.

There are at least three possibilities of such approximation of the random process $m(\vartheta, t), 0 \leq t \leq T$.

Consider the random process $\hat{m}_{t,T}, \tau_T < t \leq T$

$$\mathrm{d}\hat{m}_{t,T} = -a\left(\vartheta^{\star}_{t,T}\right)\hat{m}_{t,T}\mathrm{d}t + \frac{\hat{\gamma}_{t,T} f\left(\vartheta^{\star}_{t,T}\right)}{\sigma^2}\left[\mathrm{d}X_t - f\left(\vartheta^{\star}_{t,T}\right)\hat{m}_{t,T}\mathrm{d}t\right],$$

with $\hat{m}_{\tau_T,T}$ and

$$\frac{\partial\hat{\gamma}_{t,T}}{\partial t} = -2a\left(\vartheta^{\star}_{t,T}\right)\hat{\gamma}_{t,T} - \frac{\hat{\gamma}^2_{t,T} f\left(\vartheta^{\star}_{t,T}\right)^2}{\sigma^2} + b\left(\vartheta^{\star}_{t,T}\right)^2,$$

with $\hat{\gamma}_{\tau_T,T} = \gamma(\vartheta^{*}_{\tau_T}, \tau_T)$.

Another possibility is to use the explicit expression for the solution of the equation

$$\gamma\left(\vartheta,t\right)=\frac{e^{-2r(\vartheta)t}}{\frac{1}{\gamma(\vartheta,0)-\gamma_*(\vartheta)}+\frac{f(\vartheta)^2\left(1-e^{-2r(\vartheta)t}\right)}{2r(\vartheta)\sigma^2}}+\gamma_*\left(\vartheta\right),$$

where

$$r\left(\vartheta\right)=\left(a\left(\vartheta\right)^2+\frac{f\left(\vartheta\right)^2b\left(\vartheta\right)^2}{\sigma^2}\right)^{1/2},\qquad \gamma_*\left(\vartheta\right)=\frac{\sigma^2\left[r\left(\vartheta\right)-a\left(\vartheta\right)\right]}{f\left(\vartheta\right)^2}.$$

The approximation of $m\left(\vartheta,t\right)$ can be realized by the random process $\hat m_T\left(t\right),\tau_T<t\leq T$

$$\begin{aligned}\mathrm{d}\hat m_T\left(t\right)&=-a\left(\vartheta^\star_{t,T}\right)\hat m_T\left(t\right)\mathrm{d}t\\&\qquad+\frac{\gamma\left(\vartheta^\star_{t,T},t\right)f\left(\vartheta^\star_{t,T}\right)}{\sigma^2}\left[\mathrm{d}X_t-f\left(\vartheta^\star_{t,T}\right)\hat m_T\left(t\right)\mathrm{d}t\right],\end{aligned}$$

with the initial value $m_T\left(\tau_T\right)$.

The third opportunity is to use the exponential convergence of the solution $\gamma(\vartheta,t)$ of the Riccati to the steady-state value $\gamma_*(\vartheta)$ and the approximation of $m(\vartheta,t)$ can be done with the help of the solution $m^\star_{t,T}, \tau_T < t \leq T$ of the equation

$$\mathrm{d}m^\star_{t,T} = -r\left(\vartheta^\star_{t,T}\right) m^\star_{t,T}\mathrm{d}t + B\left(\vartheta^\star_{t,T}\right)\,\mathrm{d}X_t, \qquad m^\star_{\tau_T,T}$$

where we denoted $B(\vartheta) = [r(\vartheta) - a(\vartheta)]\, f(\vartheta)^{-1}$.

Introduce the constants

$$K(\vartheta_0) = -\left[\dot a(\vartheta_0) + \frac{\dot f(\vartheta_0)}{f(\vartheta_0)}\gamma(\vartheta_0)\right]\frac{\sigma}{f(\vartheta_0)},$$

$$N(\vartheta_0) = \frac{\dot B(\vartheta_0)\,\sigma}{\sqrt{2r(\vartheta_0)}}, \quad \eta_T(v) = \sqrt{T}\left(\vartheta^\star_{t,T} - \vartheta_0\right),$$

$$L(\vartheta_0) = \frac{r(\vartheta_0)\gamma(\vartheta_0) + a(\vartheta_0)r(\vartheta_0) - 4a(\vartheta_0)\gamma(\vartheta_0)}{2a(\vartheta_0)\gamma(\vartheta_0)r(\vartheta_0)}$$

and random variables

$$\xi_T(v) = \sqrt{2r(\vartheta_0)\,T}\int_{v-\nu_T}^{v} e^{-r(\vartheta_0)(v-u)T}\mathrm{d}W_T(u), \qquad v \in [\varepsilon_0, 1].$$

Here $\varepsilon_0 \in (0,1)$, $\nu_T = T^{-\kappa}, \kappa \in (0,1)$ and note that

$$\mathbf{E}_{\vartheta_0}\xi_T(v)^2 = 1 + o(1).$$

If $v_1 \neq v_2$, then

$$\mathbf{E}_{\vartheta_0}\xi_T(v_1)\,\xi_T(v_2) = 2r(\vartheta_0)\,T\int_{(v_1-\nu_T)\vee(v_2-\nu_T)}^{v_1\wedge v_2} e^{-2r(\vartheta_0)(v-u)T}\mathrm{d}u = 0$$

for $|v_2 - v_1| > \nu_T \to 0$.

Theorem 5. *Suppose that the conditions of Theorem 4 hold. Then the random process* $m^\star_{t,T} - m(\vartheta_0, t)$ *after re-parametrization* $m^\star_{vT,T} - m(\vartheta_0, vT), v \in [\varepsilon_0, 1]$ *admits the representation*

$$\sqrt{T}\left(m^\star_{vT,T} - m(\vartheta_0, vT)\right) = N(\vartheta_0)\,\frac{\eta_T(v)\,\xi_T(v)}{v} + o(1)$$

$$+\ \frac{K(\vartheta_0)\,\eta_T(v)}{v}\sqrt{T}\int_{-\infty}^{v}\left[e^{-a(\vartheta_0)(v-u)T} - e^{-\gamma(\vartheta_0)(v-u)T}\right]\mathrm{d}W_T(u),$$

where $\eta_T(\cdot) \Rightarrow W(\cdot)$, *The r. v.'s* $\xi_T(v) \Longrightarrow \xi(v) \sim \mathcal{N}(0,1)$, $v \in [\varepsilon_0, 1]$, *where* $W(\cdot)$ *and* $\xi(v)$ *are independent.*

The mean square error is

$$\mathbf{E}_{\vartheta_0}\left(m^\star_{vT,T} - m(\vartheta_0, vT)\right)^2 = \frac{N(\vartheta_0)^2 + K(\vartheta_0)^2 L(\vartheta_0)}{v\,\mathrm{I}(\vartheta_0)}\,\frac{1}{T}\,(1 + o(1))\,.$$

Remark 2. Suppose that the unknown parameter is $\vartheta = b$ and write the adaptive filter for the system

$$\begin{aligned} \mathrm{d}X_t &= f\,Y_t\,\mathrm{d}t + \sigma\,\mathrm{d}W_t, \qquad X_0, \qquad 0 \le t \le T, \\ \mathrm{d}Y_t &= -a\,Y_t\,\mathrm{d}t + \vartheta\,\mathrm{d}V_t, \qquad Y_0, \qquad t \ge 0, \end{aligned}$$

in recurrent form. Recall that $\tau_T = T^\delta, \delta \in (\frac{1}{2}, 1)$, the preliminary estimator and One-step MLE process are

$$\vartheta^*_{\tau_T} = \left(\frac{a^3}{f^2\left[e^{-a} - 1 + a\right]\tau_T} \sum_{k=1}^{\tau_T} \Big(\left[X_k - X_{k-1}\right]^2 - \sigma^2 \Big) \right)^{1/2},$$

$$\mathrm{d}\vartheta^\star_{t,T} = \frac{\vartheta^*_{\tau_T} - \vartheta^\star_{t,T}}{t - \tau_T}\,\mathrm{d}t + \frac{2\sigma^3\,r(\vartheta^*_{\tau_T})^3\,\dot{M}(\vartheta^*_{\tau_T}, t)}{f^4\,\vartheta^{*2}_{\tau_T}\,\left(t - \tau_T\right)}\,\Big[\mathrm{d}X_t - M(\vartheta^*_{\tau_T}, t)\mathrm{d}t\Big].$$

The equations for $M(\vartheta^*_{\tau_T},t),\tau_T<t\leq T$ and $\dot M(\vartheta^*_{\tau_T},t),\tau_T<t\leq T$ are

$$\mathrm{d}M(\vartheta^*_{\tau_T},t)=-r(\vartheta^*_{\tau_T})M(\vartheta^*_{\tau_T},t)\mathrm{d}t+\gamma(\vartheta^*_{\tau_T})\mathrm{d}X_t,$$
$$\mathrm{d}\dot M(\vartheta^*_{\tau_T},t)=-r(\vartheta^*_{\tau_T})\dot M(\vartheta^*_{\tau_T},t)\mathrm{d}t+\dot r(\vartheta^*_{\tau_T})M(\vartheta^*_{\tau_T},t)\mathrm{d}t+\dot\gamma(\vartheta^*_{\tau_T})\mathrm{d}X_t.$$

where

$$r\left(\vartheta\right)=\left(a^2+\frac{f^2\vartheta^2}{\sigma^2}\right)^{1/2},\qquad\qquad \gamma\left(\vartheta\right)=\sigma^2\left[r\left(\vartheta\right)-a\right],$$

and the initial values $M(\vartheta^*_{\tau_T},\tau_T),\dot M(\vartheta^*_{\tau_T},\tau_T)$ are defined as before.

The adaptive K-B filter is

$$\mathrm{d}m^\star_{t,T}=-r\left(\vartheta^\star_{t,T}\right)m^\star_{t,T}\mathrm{d}t+B\left(\vartheta^\star_{t,T}\right)\,\mathrm{d}X_t,\quad m^\star_{\tau_T,T},\quad \tau_T<t\leq T.$$

1.11 Asymptotic efficiency

Theorem 6. *We have the following lower minimax bound for any estimator* $\bar{m}_{t,T}$ *of* $m(\vartheta,t)$ *(below* $t = vT$*)*

$$\lim_{\nu\to 0} \varliminf_{T\to\infty} \sup_{|\vartheta-\vartheta_0|\leq\nu} t\mathbf{E}_\vartheta \left|\bar{m}_{t,T} - m(\vartheta,t)\right|^2 \geq \mathbf{E}_{\vartheta_0}\left(\dot{m}(\vartheta_0)^2\,\eta(\vartheta_0)^2\right).$$

Here $\eta_{t,T}(\vartheta_0) = \sqrt{t}\left(\vartheta^\star_{t,T} - \vartheta_0\right) \Rightarrow \eta(\vartheta_0)$

$$\mathbf{E}_{\vartheta_0}\left(\dot{m}(\vartheta_0)^2\,\eta(\vartheta_0)^2\right) = \lim_{T\to\infty} \mathbf{E}_{\vartheta_0}\left(\dot{m}(\vartheta_0,t)^2\,\eta_{t,T}(\vartheta_0)^2\right).$$

We call the estimator $\check{m}_{t,T}$ (adaptive filter) *asymptotically efficient* if for all $\vartheta_0 \in \Theta$

$$\lim_{\nu\to 0} \lim_{T\to\infty} \sup_{|\vartheta-\vartheta_0|\leq\nu} t\mathbf{E}_\vartheta \left|\check{m}_{t,T} - m(\vartheta)\right|^2 = \mathbf{E}_{\vartheta_0}\left(\dot{m}(\vartheta_0)^2\,\eta(\vartheta_0)^2\right).$$

Theorem 7. *The asymptotically efficient adaptive filter is* $\check{m}_{t,T} = m\left(\vartheta^{\star}_{t,T}, t\right), \tau_T < t \leq T$, *where*

$$m\left(\vartheta, t\right) = m\left(\vartheta, \tau_T\right) e^{-r(\vartheta)\left(t-\tau_T\right)} + X_t - e^{-r(\vartheta)\left(t-\tau_T\right)} X_{\tau_T} - r\left(\vartheta\right) \int_{\tau_T}^{t} e^{-r(\vartheta)(t-s)} X_s \mathrm{d}s, \qquad \tau_T < t \leq T$$