Workshop on statistics of stochastic processes in honor of Yury Kutoyants' 80th birthday September 17, 2026

Risk comparison theorems in deep learning and application to diffusion coefficients

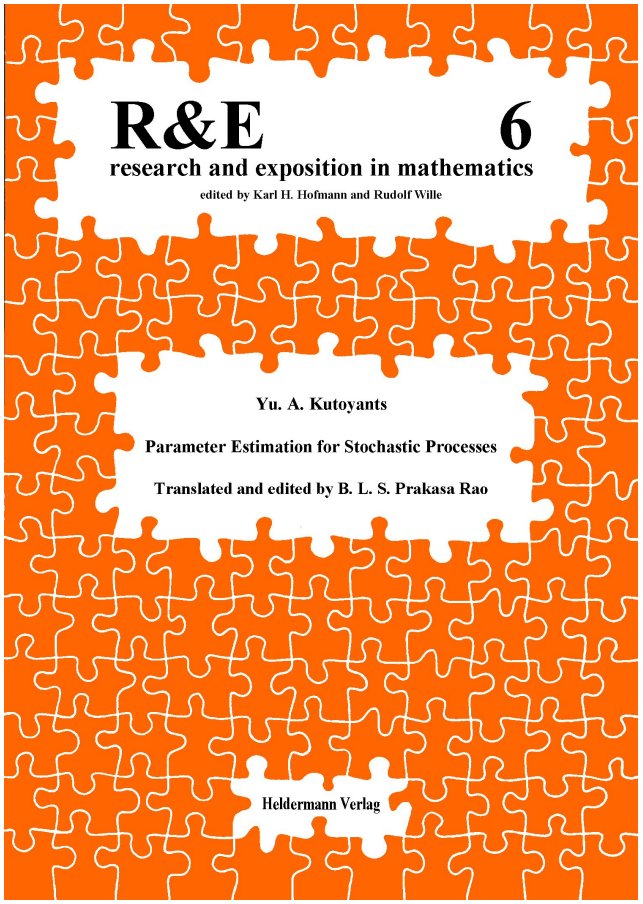
Honoring Prof. Yury A. Kutoyants on his 80th Birthday

Nakahiro Yoshida

Graduate School of Mathematical Sciences, University of Tokyo
Japan Science and Technology Agency CREST
Institute of Statistical Mathematics

My first encounter with Prof. Yury Kutoyants

- **Yu. A. Kutoyants:**
  Parameter Estimation for Stochastic Processes. Heldermann Verlag, 1984 (English version)



Ibragimov-Khasminskii-Kutoyants program

- **Beautiful theory**
  - **likelihood ratio random field** $\mathbb{Z}_n(u) = \dfrac{dP_{\theta^*+n^{-1/2}u}}{dP_{\theta^*}}$
  - **Convergence of random field** $\mathbb{Z}_n$

$$\mathbb{Z}_n \to^d \mathbb{Z} \qquad \text{in } \widehat{C} \tag{1}$$

  - (1) $\Rightarrow$ **asymptotic behavior of the estimators (MLE, BE)**

$$\hat{u}_n = \sqrt{n}\left(\hat{\theta}_n - \theta^*\right) \to^d u_\infty$$

  - **convergence of moments**

$$E\left[f\left(\hat{u}_n\right)\right] \to \mathbb{E}[f(u_\infty)] \qquad (f \in C_p)$$

under a large deviation estimate

$$P\left[\sup_{|u|\geq r} \mathbb{Z}_n(u) \geq e^{-r^\alpha}\right] \leq Ce^{-r^\alpha}$$

Kutoyants' Grand Design

- **Grand Design**

  Statistical theory for stochastic processes can be constructed on the rigid basement of the Le Cam-Hájek-Ibragimov-Khasminskii asymptotic decision theory

- **Proof and Foundational Impact: Established breakthrough methods in statistical inference for diffusion processes, Poisson processes, and hidden Markov models.**
  - **Parameter Estimation for Stochastic Processes, 1984 (English version)**
  - **Identification of Dynamical Systems with Small Noise, 1994**
  - **Statistical Inference for Spatial Poisson Processes, 1998**
  - **Statistical Inference for Ergodic Diffusion Processes, 2004**
  - **Introduction to the Statistics of Poisson Processes and Applications, 2023**
  - **Hidden Markov Processes and Adaptive Filtering, 2025**

Why convergence of moments?

- **Asymptotic decision theory. An asymptotically efficient estimator $\hat{\theta}_n$ should attain the asymptotic minimax risk bound (Hájek):**

$$\lim_{\delta \downarrow 0} \limsup_{n \to \infty} \sup_{\theta : |\theta - \theta^*| < \delta} E_\theta\big[|\sqrt{n}(\hat{\theta}_n - \theta)|^p\big] = \mathbb{E}[|\hat{u}_\infty|^p].$$

- **Information criterion. When $\hat{u}_n = \sqrt{n}(\hat{\theta}_n - \theta^*)$ for the MLE $\hat{\theta}_n$, the AIC is validated by the convergence**

$$\begin{aligned} \mathbb{E}\big[\hat{u}_n' I(\theta^*) \hat{u}_n\big] &\to E\big[\hat{u}_\infty' I(\theta^*) \hat{u}_\infty\big] \\ &= \text{the number of parameters} \end{aligned}$$

- **Asymptotic expansion of the QLA estimators ...**

Asymptotic expansion

- a process with small diffusion ($\epsilon \downarrow 0$)

$$\begin{cases} dX_t^\epsilon = V_0(X_t^\epsilon, \theta)dt + \epsilon V(X_t^\epsilon)dw_t, & t \in [0,1] \\ X_0^\epsilon = x_0 \end{cases}$$

- MLE $\hat{\theta}_\epsilon$

$$\hat{u}_\epsilon = \epsilon^{-1}(\hat{\theta}_\epsilon - \theta^*) = f_0 + \epsilon f_1 + \cdots + \epsilon^k f_k + R_\epsilon$$

with $R_\epsilon = A_\epsilon(\hat{\theta}_\epsilon - \theta^*)$ for a functional A_ϵ. (Kutoyants 1983)

- $R_\epsilon = O_{L^{\infty-}}(\epsilon^{k+1})$ because $\hat{u}_\epsilon = O_{L^{\infty-}}(1)$, $\quad L^{\infty-} = \cap_{p>1} L^p$,
by the Ibragimov-Khasminskii-Kutoyants theory (Kutoyants 1984)

- For a measurable function g, asymptotic expansion

$$E\big[g(\hat{u}_\epsilon)\big] = c_0(g) + \epsilon c_1(g) + \cdots + \epsilon^k c_k(g) + O(\epsilon^{k+1}) \quad (\epsilon \downarrow 0)$$

- Kutoyants $\Rightarrow$ statistics for SDEs $\Rightarrow$ asymptotic expansion $\Rightarrow$ option pricing

- Kutoyants $\Rightarrow$ asymptotic decision theory for stochastic processes (minimax theory – parametric / nonparametric)

Deep Learning of diffusion matrix

A. Gloter - Y

Diffusion matrix learning

- **m-dimensional stochastic regression model**

$$dY_t \;=\; b(X_t)dt + \sigma(X_t)dw_t \quad (t \in \mathbb{R}_+) \tag{2}$$

where $w = (w_t)_{t\in\mathbb{R}_+}$ is an r-dimensional Wiener process. $S := \sigma\sigma^\star$. S^* the true value of S

- **$X = (X_t)_{t\in\mathbb{R}_+}$: α-mixing Itô process. $Z = (X, Y)$**
  With $\Delta_j Y = Y_{t_j} - Y_{t_{j-1}}$, $h = t_j - t_{j-1}$ and $\mathbb{I}_n = \{1, ..., n\}$. the contrast function

$$\Phi_n(S, Z) \;=\; \frac{1}{2n}\sum_{j\in\mathbb{I}_n}\left\{h^{-1}S(X_{t_{j-1}})^{-1}[(\Delta_j Y)^{\otimes 2}] + \log\det S(X_{t_{j-1}})\right\}\mathcal{W}(X_{t_{j-1}}). \tag{3}$$

We (nearly) optimize Ψ_n to obtain $\widehat{S}_n$. $\widehat{S}_n$ is $\sigma[(X_{t_j}, Y_{t_j})_{j\in\mathbb{I}_{0,n}}]$-measurable. $\mathbb{I}_{0,n} = \{0, 1, ..., n\}$

- **$(\overline{X}, \overline{Y})$: an independent copy of (X, Y)**
- **The generalization error $\overline{\mathrm{R}}_n = \overline{\mathrm{R}}_n(\widehat{S}_n, S^*)$ is defined as**

$$\overline{\mathrm{R}}_n \;=\; \frac{1}{2n}E\Big[\sum_{j\in\mathbb{I}_n}\Big\{(\widehat{S}_n(\overline{X}_{t_{j-1}})^{-1} - S^*(\overline{X}_{t_{j-1}})^{-1})\big[h^{-1}(\Delta_j\overline{Y})^{\otimes 2}\big] + \log\frac{\det\widehat{S}_n(\overline{X}_{t_{j-1}})}{\det S^*(\overline{X}_{t_{j-1}})}\Big\}\mathcal{W}(\overline{X}_{t_{j-1}})\Big]$$

- **$t_j = t_j^n = jh$, $h = h_n$, $nh \to \infty$ and $h \to 0$ as $n \to \infty$.**
- **$U(x, S) = \frac{1}{2}\Big\{\big(S(x)^{-1} - S^*(x)^{-1}\big)[S^*(x)] + \log\frac{\det S(x)}{\det S^*(x)}\Big\}\mathcal{W}(x) \quad (x \in \mathcal{X})$.**
- **Aim. An oracle inequality:**

$$\overline{\mathrm{R}}_n \;\le\; C_1\Big(\Delta_n + \widehat{\Delta}_n + \inf_{S\in\mathfrak{F}_n} E\big[\overline{U}_n(S)\big] + n^{-1}(\log n)\log\mathcal{N}_n + \delta_n\log n + h_n^2\Big) \tag{4}$$

Setting

- $(\Omega, \mathcal{F}, \mathbf{F}, P)$ a stochastic basis, $\mathbf{F} = (\mathcal{F}_t)_{t\in\mathbb{R}_+}$
- Let $\mathbb{R}_+ = \mathbb{R}_{\geq 0}$, $\mathbb{Z}_+ = \mathbb{Z}_{\geq 0}$ and $\mathbb{N} = \mathbb{Z}_{\geq 1}$.
- $Y = (Y_t)_{t\in\mathbb{R}_+}$ an $\mathbf{m}$-dimensional adapted process the stochastic regression model satisfying

$$dY_t \;=\; b(X_t)dt + \sigma(X_t)dw_t \quad (t \in \mathbb{R}_+) \tag{5}$$

- $b : \mathbb{R}^{\mathbf{d}} \to \mathbb{R}^{\mathbf{m}}$, $\sigma : \mathbb{R}^{\mathbf{d}} \to \mathbb{R}^{\mathbf{m}} \otimes \mathbb{R}^{\mathbf{r}}$
- $w = (w_t)_{t\in\mathbb{R}_+}$ is an $\mathbf{r}$-dimensional Wiener process
- $X = (X_t)_{t\in\mathbb{R}_+}$ a mixing, progressively measurable, stationary process
- $X = (X_t)_{t\in\mathbb{R}_+}$ may have jumps.
- $n \in \mathbb{Z}_{\geq 2}$
- $\mathbb{I}_{0,n} = \{0, 1, ..., n\}$
- $\mathbb{I}_n = \{1, ..., n\}$
- $t_j = t_j^n = jh$, $h = h_n$
- $nh \to \infty$ and $h \to 0$ as $n \to \infty$.

Remark

- observations $(X_{t_j}, Y_{t_j})_{t \in \mathbb{I}_{0,n}}$, $\mathbb{I}_{0,n} = \{0, 1, ..., n\}$
- $\mathcal{B}_I^X = \sigma[X_t, ;\ t \in I]$ for $I \subset \mathbb{R}_+$
- The α-mixing coefficient is given by

$$\alpha^X(r) = \sup_{j \in \mathbb{N}} \sup \left\{ \left| P[\mathtt{A} \cap \mathtt{B}] - P[\mathtt{A}]P[\mathtt{B}] \right|;\ \mathtt{A} \in \mathcal{B}^X_{[0,t]}, \mathtt{B} \in \mathcal{B}^X_{[t+r,\infty)} \right\}$$

- We assume that there exists a positive constant γ such that $\alpha^X(r) \leq \gamma^{-1} e^{-\gamma r}$ for all $r > 0$.
- Remark.
  - $\mathcal{B}_I^{X,dY} = \sigma[X_t, Y_t - Y_s;\ s, t \in I]$ for $I \subset \mathbb{R}_+$,

$$\alpha^{X,dY}(r) = \sup_{t \in \mathbb{R}_+} \sup \left\{ \left| P[\mathtt{A} \cap \mathtt{B}] - P[\mathtt{A}]P[\mathtt{B}] \right|;\ \mathtt{A} \in \mathcal{B}^{X,dY}_{[0,t]}, \mathtt{B} \in \mathcal{B}^{X,dY}_{[t+r,\infty)} \right\}.$$

  - Then $\alpha^{X,dY}(r) \geq \alpha^X(r)$.
  - The conventional approach would only give an error bound $T^{-\kappa}$ for $T = nh$ for some constant $\kappa < 1$.
  - This bound is suboptimal since $T \ll n$. We want the bound $n^{-\kappa}$.
  - The Risk Comparison Theorem overcomes this difficulty.

Estimator (First stage)

- $\mathcal{W}: \mathbb{R}^{\mathbf{d}} \to \mathbb{R}_+ = [0, \infty)$, $\mathcal{W}^{1/2} \in C_b^{\ell, \beta-\ell}(\mathbb{R}^{\mathbf{d}})$, $\Delta_j Y = Y_{t_j} - Y_{t_{j-1}}$, $Z = (X, Y)$
- The contrast function

$$\Phi_n(S, Z) \;=\; \frac{1}{2n} \sum_{j \in \mathbb{I}_n} \left\{ h^{-1} S(X_{t_{j-1}})^{-1} [(\Delta_j Y)^{\otimes 2}] + \log \det S(X_{t_{j-1}}) \right\} \mathcal{W}(X_{t_{j-1}})$$

- $\mathbb{S}$: a family of measurable functions $S : \mathcal{X} \to \mathcal{S}_+$ ($\mathbf{m} \times \mathbf{m}$ nonnegative-definite symmetric matrices) such that

$$0 < \inf_{x \in \mathcal{X}, S \in \mathbb{S}} \lambda_{\min}(S(x)) \leq \sup_{x \in \mathcal{X}, S \in \mathbb{S}} \lambda_{\max}(S(x)) < \infty, \tag{6}$$

- (Deep Learning) $\mathfrak{F}_n$ is a class of functions in $\mathbb{S}$:

$$\mathfrak{F}_n \subset \mathbb{S} \tag{7}$$

- $\widehat{S}_n^0$: an estimator that asymptotically minimizes the contrast function $\Phi_n(S, Z)$. in $S \in \mathfrak{F}_n$. $\widehat{S}_n^0$ is $\sigma[(X_{t_j}, Y_{t_j})_{j \in \mathbb{I}_{0,n}}]$-measurable, and

$$\widehat{S}_n^0 \;\in\; \mathfrak{F}_n \quad a.s. \tag{8}$$

- The error $\Delta_n = E\left[\Phi_n(\widehat{S}_n^0; Z) - \inf_{S \in \mathfrak{F}_n} \Phi_n(S; Z) \right]$

Estimator (Second stage)

- $|S| = \left(\sum_{i,j} S_{i,j}^2\right)^{1/2}$ for a matrix $S = (S_{i,j})$
- notation.

$$|S|_n \;=\; \left(n^{-1}\sum_{j\in\mathbb{I}_n} \left|S(X_{t_{j-1}})\right|^2 \mathcal{W}(X_{t_{j-1}})\right)^{1/2}$$

  for a function $S : \mathcal{X} \to \mathbb{R}^{\mathbf{m}} \otimes \mathbb{R}^{\mathbf{m}}$. $|S|_n$ is well defined; $|S(x)|\mathcal{W}(x)$ should read 0 when $\mathcal{W}(x) = 0$.

- a set $\mathbb{B} \subset C(\mathcal{X}; \mathcal{S}_+) \cap \mathbb{S}$ such that $\{S|_{\mathcal{X}^o};\ S \in \mathbb{B}\}$ is a bounded set in $C_b^{\ell,\beta-\ell}(\mathcal{X}^o; \mathbb{R}^{\mathbf{m}} \otimes \mathbb{R}^{\mathbf{m}})$
  $\mathcal{X}^o = \mathrm{Int}\ \mathcal{X}$. The set $C_b^{\ell,\beta-\ell}(\mathcal{X}^o; \mathbb{R}^{\mathbf{m}} \otimes \mathbb{R}^{\mathbf{m}})$ is identified with $\left(C_b^{\ell,\beta-\ell}(\mathcal{X}^o)\right)^{\mathbf{m}^2}$.
- We consider the second-stage estimator $\widehat{S}_n$: it is a $\mathbb{B}$-valued measurable function:

$$\widehat{S}_n \in \mathbb{B} \quad a.s.$$

  $\widehat{S}_n$ is a projection of $\widehat{S}_n^0$ to $\mathbb{B}$. (smoothing)

Estimator (Second stage)

- **Recall**

$$\left|\boldsymbol{S}_1 - \boldsymbol{S}_0\right|_n^2 \;=\; n^{-1} \sum_{j \in \mathbb{I}_n} \left|\boldsymbol{S}_1(\boldsymbol{X}_{t_{j-1}}) - \boldsymbol{S}_0(\boldsymbol{X}_{t_{j-1}})\right|^2 \boldsymbol{\mathcal{W}}(\boldsymbol{X}_{t_{j-1}})$$

  for matrices $\boldsymbol{S}_0$ and $\boldsymbol{S}_1$ of the same size.

- **The error in the projection**

$$\widehat{\boldsymbol{\Delta}}_n \;=\; \left|\widehat{\boldsymbol{S}}_n - \widehat{\boldsymbol{S}}_n^0\right|_n^2 - \inf_{\boldsymbol{S} \in \mathbb{B}} \left|\boldsymbol{S} - \widehat{\boldsymbol{S}}_n^0\right|_n^2 \tag{9}$$

Generalization error

- $\overline{\boldsymbol{Z}} = (\overline{\boldsymbol{X}}, \overline{\boldsymbol{Y}})$ an independent copy of $\boldsymbol{Z} = (\boldsymbol{X}, \boldsymbol{Y})$. $\boldsymbol{S}^*|_{\mathcal{X}} \in \mathbb{B}$
- The generalization error $\overline{\mathrm{R}}_n = \overline{\mathrm{R}}_n(\widehat{\boldsymbol{S}}_n) = \overline{\mathrm{R}}_n(\widehat{\boldsymbol{S}}_n, \boldsymbol{S}^*)$ for $\widehat{\boldsymbol{S}}_n$ is defined as $\overline{\mathrm{R}}_n = \boldsymbol{E}\big[\boldsymbol{\Phi}_n(\widehat{\boldsymbol{S}}_n, \overline{\boldsymbol{Z}})\big] - \boldsymbol{E}\big[\boldsymbol{\Phi}_n(\boldsymbol{S}^*, \overline{\boldsymbol{Z}})\big]$, that is,

$$\overline{\mathrm{R}}_n \;=\; \frac{1}{2n}\boldsymbol{E}\Big[\sum_{j\in\mathbb{I}_n}\Big\{(\widehat{\boldsymbol{S}}_n(\overline{\boldsymbol{X}}_{t_{j-1}})^{-1} - \boldsymbol{S}^*(\overline{\boldsymbol{X}}_{t_{j-1}})^{-1})\big[h^{-1}(\boldsymbol{\Delta}_j\overline{\boldsymbol{Y}})^{\otimes 2}\big] + \log\frac{\det\widehat{\boldsymbol{S}}_n(\overline{\boldsymbol{X}}_{t_{j-1}})}{\det\boldsymbol{S}^*(\overline{\boldsymbol{X}}_{t_{j-1}})}\Big\}\boldsymbol{\mathcal{W}}(\overline{\boldsymbol{X}}_{t_{j-1}})\Big]$$

  We do not assume the stationarity of $(\boldsymbol{X}, \boldsymbol{Y})$.

- The discrepancy between the matrices $\boldsymbol{S}(\boldsymbol{x})$ and $\boldsymbol{S}^*(\boldsymbol{x})$ is evaluated by using the loss function

$$\boldsymbol{U}(\boldsymbol{x}, \boldsymbol{S}) \;=\; \frac{1}{2}\Big\{(\boldsymbol{S}(\boldsymbol{x})^{-1} - \boldsymbol{S}^*(\boldsymbol{x})^{-1})[\boldsymbol{S}^*(\boldsymbol{x})] + \log\frac{\det\boldsymbol{S}(\boldsymbol{x})}{\det\boldsymbol{S}^*(\boldsymbol{x})}\Big\}\boldsymbol{\mathcal{W}}(\boldsymbol{x})$$

- The predictable generalization error $\overline{\mathrm{R}}^p_n = \overline{\mathrm{R}}^p_n(\widehat{\boldsymbol{S}}_n) = \overline{\mathrm{R}}^p_n(\widehat{\boldsymbol{S}}_n, \boldsymbol{S}^*)$ for $\widehat{\boldsymbol{S}}_n$ defined by

$$\overline{\mathrm{R}}^p_n \;=\; n^{-1}\boldsymbol{E}\Big[\sum_{j\in\mathbb{I}_n}\boldsymbol{U}(\overline{\boldsymbol{X}}_{t_{j-1}}, \widehat{\boldsymbol{S}}_n)\Big],$$

  equivalently,

$$\overline{\mathrm{R}}^p_n \;=\; \frac{1}{2n}\boldsymbol{E}\Big[\sum_{j\in\mathbb{I}_n}\Big\{(\widehat{\boldsymbol{S}}_n(\overline{\boldsymbol{X}}_{t_{j-1}})^{-1} - \boldsymbol{S}^*(\overline{\boldsymbol{X}}_{t_{j-1}})^{-1})[\boldsymbol{S}^*(\overline{\boldsymbol{X}}_{t_{j-1}})] + \log\frac{\det\widehat{\boldsymbol{S}}_n(\overline{\boldsymbol{X}}_{t_{j-1}})}{\det\boldsymbol{S}^*(\overline{\boldsymbol{X}}_{t_{j-1}})}\Big\}\boldsymbol{\mathcal{W}}(\overline{\boldsymbol{X}}_{t_{j-1}})\Big].$$

  The predictable generalization error is always nonnegative.

Oracle inequality

- Denote by $\{S_k\}_{k=1,\dots,\mathcal{N}_n}$ a set of functions $S_k \in \mathbb{S}$ such that

$$\mathfrak{F}_n \;=\; \bigcup_{k=1}^{\mathcal{N}_n} \left\{ S \in \mathfrak{F}_n;\; \sup_{x \in \operatorname{supp} \mathcal{W}} \left\| S(x) - S_k(x) \right\| < \delta_n \right\}$$

  for $\delta_n > 0$. Here $\|S\| = \max_{i,j} |S_{i,j}|$ for a matrix $S = (S_{i,j})$.

- The number $\mathcal{N}_n$ is the covering number of $\mathfrak{F}_n$ by the δ_n-balls with respect to the sup-norm.

- $\overline{U}_n(S) = \frac{1}{n} \sum_{j \in \mathbb{I}_n} U(X_{t_{j-1}}, S)$.

- Oracle inequality

  Theorem 1. There exists a constant C_1 such that

$$\overline{\mathrm{R}}_n \vee \overline{\mathrm{R}}_n^p \;\le\; C_1 \left(\Delta_n + \widehat{\Delta}_n + \inf_{S \in \mathfrak{F}_n} E\left[\overline{U}_n(S) \right] + n^{-1} (\log n) \log \mathcal{N}_n + \delta_n \log n + h_n^2 \right)$$

  for all $n \in \mathbb{N}$.

Remarks

Remark 2. (i) Denote by the subscript log (like $\sim_{\log}$) the relation modulo a (positive or negative) power of $\log n$. For example, in the case where

$$n^{-1} \log \mathcal{N}_n \ \sim_{\log} \ \gamma_n \sim_{\log} n^{-\frac{2\beta}{2\beta+\mathrm{d}}}, \tag{10}$$

then Condition (35) is rephrased as

$$hn^{\frac{2\beta}{2\beta+\mathrm{d}}} \ \to_{\log} \ \infty. \tag{11}$$

Then h must satisfy $h \gg_{\log} n^{-\frac{2\beta}{2\beta+\mathrm{d}}}$.
(ii) On the other hand, for the error bound provided by Theorems 5 to make sense in comparison with the discretization error, it should hold that $h^2 \leq \gamma_n$ as well as $h \gg_{\log} n^{-\frac{2\beta}{2\beta+\mathrm{d}}}$. Thus, it is reasonable to consider the case where

$$n^{-\frac{2\beta}{2\beta+\mathrm{d}}} \ll_{\log} \ h \ \ll_{\log} \ n^{-\frac{\beta}{2\beta+\mathrm{d}}}. \tag{12}$$

If making $\beta \uparrow \infty$ just formally in (12), we would have the conditions

$$T = nh \to \infty \ \text{ and } \ nh^2 \to 0. \tag{13}$$

The conditions in (13) are natural by analogy to the parametric case since the second one is the basic balance condition between n and h.
(iii) The best rate attained by the bound (34) is typically given by the balance between the terms $\inf_{S \in \mathfrak{F}_n} E[\overline{U}_n(S)]$ and $n^{-1}(\log n) \log \mathcal{N}_n$.

Strategy of the proof: Empirical risks

- The predictable empirical risk of $\widehat{S}_n$ is define by

$$\begin{aligned}\mathcal{E}_n^p &= \frac{1}{2n}\sum_{j\in\mathbb{I}_n}\left\{(\widehat{S}_n(X_{t_{j-1}})^{-1} - S^*(X_{t_{j-1}})^{-1})\big[S^*(X_{t_{j-1}})\big] + \log\frac{\det \widehat{S}_n(X_{t_{j-1}})}{\det S^*(X_{t_{j-1}})}\right\}\mathcal{W}(X_{t_{j-1}}) \\ &= n^{-1}\sum_{j\in\mathbb{I}_n} U(X_{t_{j-1}}, \widehat{S}_n). \end{aligned} \tag{14}$$

Let $\mathrm{R}_n^p = E[\mathcal{E}_n^p]$.

- The empirical risk of $\widehat{S}_n^0$ is define by

$$\begin{aligned}\mathcal{E}_n^0 &= \Phi_n(\widehat{S}_n^0, Z) - \Phi_n(S^*, Z) \\ &= \frac{1}{2n}\sum_{j\in\mathbb{I}_n}\left\{(\widehat{S}_n^0(X_{t_{j-1}})^{-1} - S^*(X_{t_{j-1}})^{-1})\big[h^{-1}(\Delta_j Y)^{\otimes 2}\big] + \log\frac{\det \widehat{S}_n^0(X_{t_{j-1}})}{\det S^*(X_{t_{j-1}})}\right\}\mathcal{W}(X_{t_{j-1}}) \end{aligned} \tag{15}$$

Let $\mathrm{R}_n^0 = E[\mathcal{E}_n^0]$.

- The predictable empirical risk of $\widehat{S}_n^0$ is define by

$$\mathcal{E}_n^{0,p} = \frac{1}{2n}\sum_{j\in\mathbb{I}_n}\left\{(\widehat{S}_n^0(X_{t_{j-1}})^{-1} - S^*(X_{t_{j-1}})^{-1})\big[S^*(X_{t_{j-1}})\big] + \log\frac{\det \widehat{S}_n^0(X_{t_{j-1}})}{\det S^*(X_{t_{j-1}})}\right\}\mathcal{W}(X_{t_{j-1}}) \tag{16}$$

Let $\mathrm{R}_n^{0,p} = E[\mathcal{E}_n^{0,p}]$.

Strategy of the proof

- **Strategy.**

  (i) $\mathbf{R}_n^p$ **is estimated by** $\mathbf{R}_n^{0,p}$

  (ii) Estimate the statistical error $\mathbf{R}_n^{0,p} - \mathbf{R}_n^0$ **with the Brownian large deviation**

  (iii) $\mathbf{R}_n^0$ **is controlled due to optimization**

  (iv) Obtain an estimate of $\mathbf{R}_n^p$

  (v) $\overline{\mathbf{R}}_n^p$ **is estimated by** $\mathbf{R}_n^p$ **with the comparison theorem and the** X**-mixing large deviation**

  (vi) Estimate $\overline{\mathbf{R}}_n$ **by** $\overline{\mathbf{R}}_n^p$**.**

$$\overline{\mathbf{R}}_n \lesssim \overline{\mathbf{R}}_n^p \overset{RCT+XMLD}{\lesssim} \mathbf{R}_n^p \lesssim \mathbf{R}_n^{0,p} \overset{\text{BLD}}{\lesssim} \mathbf{R}_n^0$$

RCT: Risk Comparison Theorem

XMLD: X-Mixing Large Deviation

BLD: Brownian Large Deviation

Risk comparison theorem

- Suppose that $\mathcal{X}$ is a compact set in $\mathbb{R}^{\mathbf{d}}$ and $\hat{\mathcal{X}}$ is an open set of $\mathbb{R}^{\mathbf{d}}$ such that $\mathcal{X} \subset \hat{\mathcal{X}}$.
- $m \in \mathbb{N}$.
- $\xi_n = (\xi_{n,j})_{j=1,\dots,n}$ and $\xi'_n = (\xi'_{n,j})_{j=1,\dots,n}$ $(n \in \mathbb{N})$ are collections of $\mathbb{R}^{\mathbf{d}}$-valued random variables; these variables may be correlated with each other.
- The sequence of positive numbers $(d_n)_{n\in\mathbb{N}}$ satisfies $\lim_{n\to\infty} d_n = 0$.
- $\mathtt{B}$ is a bounded set in $C_b^{\ell,\beta-\ell}(\mathbb{R}^{\mathbf{d}})$.
- $D_n = (D_n^{(i)})_{i\in[\mathbf{d}_D]}$ is a collection of $\mathtt{B}$-valued random fields $D_n^{(i)}$ satisfying that supp $D_n^{(i)} \in \mathcal{X}$ a.s. for every $n \in \mathbb{N}$ and $i \in [\mathbf{d}_D] = \{1, \dots, \mathbf{d}_D\}$.
- $\varphi_{0,d_n/(4m)}$ is a specific kernel function.

Risk comparison theorem

- Define $\mathbb{R}_n(\boldsymbol{\xi}_n)$ as

$$\mathbb{R}_n(\boldsymbol{\xi}_n) \;=\; \boldsymbol{E}\Big[\boldsymbol{n}^{-1}\sum_{j=1}^{n}\big|\boldsymbol{D}_n(\boldsymbol{\xi}_{n,j})\big|^{2m}\boldsymbol{\psi}(\boldsymbol{\xi}_{n,j})\Big].$$

- For $\boldsymbol{\epsilon} > \mathbf{0}$, let

$$\begin{aligned}\boldsymbol{Q}_{1,n}(\boldsymbol{a}) \;=\; \boldsymbol{P}\Big[\Big|\boldsymbol{n}^{-1}\sum_{j=1}^{n}\Big\{\boldsymbol{\varphi}_{0,d_n/(4m)}(\boldsymbol{\xi}_{n,j}-\boldsymbol{a})\boldsymbol{\psi}(\boldsymbol{\xi}_{n,j}) - \boldsymbol{E}\big[\boldsymbol{\varphi}_{0,d_n/(4m)}(\boldsymbol{\xi}_{n,j}-\boldsymbol{a})\boldsymbol{\psi}(\boldsymbol{\xi}_{n,j})\big]\Big\}\Big| \\ \geq \frac{\boldsymbol{\epsilon}}{\mathbf{2}(\mathbf{1}+\boldsymbol{\epsilon})}\boldsymbol{n}^{-1}\sum_{j=1}^{n}\boldsymbol{E}\big[\boldsymbol{\varphi}_{0,d_n/(4m)}(\boldsymbol{\xi}_{n,j}-\boldsymbol{a})\boldsymbol{\psi}(\boldsymbol{\xi}_{n,j})\big]\Big]. \qquad (17)\end{aligned}$$

Theorem 3. (Gloter-Y 2026) Suppose that $\boldsymbol{\psi} : \mathbb{R}^{\mathbf{d}} \to \mathbb{R}$ is a nonnegative bounded measurable function. Suppose that $\boldsymbol{\xi}_n =^d \boldsymbol{\xi}'_n$ for every $\boldsymbol{n} \in \mathbb{N}$. Then there exists a constant $\boldsymbol{C}$ such that

$$\mathbb{R}_n(\boldsymbol{\xi}'_n) \;\leq\; \boldsymbol{C}\Big[\mathbb{R}_n(\boldsymbol{\xi}_n) + \boldsymbol{d}_n^{2m\beta} + \boldsymbol{d}_n^{-\mathbf{d}}\sup_{\boldsymbol{a}\in\hat{\boldsymbol{\mathcal{X}}}}\boldsymbol{Q}_{1,n}(\boldsymbol{a})\Big] \qquad (18)$$

for all $\boldsymbol{n} \in \mathbb{N}$.

- We use $\boldsymbol{D}_n(\boldsymbol{x})$ for $\big(\widehat{\boldsymbol{S}}_n(\boldsymbol{x}) - \boldsymbol{S}^*(\boldsymbol{x})\big)\boldsymbol{\mathcal{W}}(\boldsymbol{x})^{1/2}$.

Application to Deep Learning of Diffusion Matrix

- For $\mathrm{v} = (v_1, ..., v_d)^\star$, the shifted ReLU activation function $\boldsymbol{\sigma}_v : \mathbb{R}^d \to \mathbb{R}^d$ as

$$\boldsymbol{\sigma}_v(x) \;=\; \big((x_1 - v_1)_+, ..., (x_d - v_d)_+\big)^\star \qquad (x = (x_1, ..., x_d)^\star \in \mathbb{R}^d)$$

  where $u_+ = \max\{u, 0\}$ for $u \in \mathbb{R}$.

- Define the map $f\big(\cdot; (\boldsymbol{W}_L, ..., \boldsymbol{W}_1, \boldsymbol{W}_0), (\mathrm{v}_L, ..., \mathrm{v}_1)\big) : \mathbb{R}^{\mathbf{p}(0)} \to \mathbb{R}^{\mathbf{p}(L+1)}$ by

$$f\big(x; (\boldsymbol{W}_L, ..., \boldsymbol{W}_1, \boldsymbol{W}_0), (\mathrm{v}_L, ..., \mathrm{v}_1)\big) \;=\; \boldsymbol{W}_L \boldsymbol{\sigma}_{\mathrm{v}_L} \boldsymbol{W}_{L-1} \boldsymbol{\sigma}_{\mathrm{v}_{L-1}} \cdots \boldsymbol{W}_1 \boldsymbol{\sigma}_{\mathrm{v}_1} \boldsymbol{W}_0 x \quad (x \in \mathbb{R}^{\mathbf{p}_0}). \tag{19}$$

  for weight matrices $\boldsymbol{W}_i \in \mathbb{R}^{\mathbf{p}(i+1)} \otimes \mathbb{R}^{\mathbf{p}(i)}$ $(i = 0, 1, ..., L)$ and shift vectors $\mathrm{v}_i \in \mathbb{R}^{\mathbf{p}(i)}$ $(i = 1, ..., L)$.

- $\mathcal{D}$: the set of functions f that of the form (19) for some L, $(\boldsymbol{W}_L, ..., \boldsymbol{W}_1, \boldsymbol{W}_0)$ and $(\mathrm{v}_L, ..., \mathrm{v}_1)$.

- The leaning machine $\mathfrak{F}_n$ specified by

$$\begin{aligned} \mathfrak{F}_n \;=\; \Big\{ & f \in \mathcal{D} \text{ of the form (19)}; \max_{\ell=0,...,L_n, j=1,...,L_n} \big(\|\boldsymbol{W}_\ell\|_\infty \vee \|\mathrm{v}_j\|_\infty\big) \le 1, \\ & \sum_{\ell=0}^{L_n} \|\boldsymbol{W}_\ell\|_0 + \sum_{j=1}^{L_n} \|\mathrm{v}_j\|_0 \le s_n,\ \|f\|_\infty \le \check{K} \Big\} \end{aligned}$$

  for some $\check{K} > 0$ and s_n, $L_n, \mathbf{p}_n(0) = \mathbf{d}, \mathbf{p}_n(1), ...\mathbf{p}_n(L_n) \in \mathbb{N}, \mathbf{p}_n(L_n + 1) = \mathbf{d}^2$. Here $\mathbf{d}$ is the dimension of the covariate process $\boldsymbol{X}$, and the efficient dimension is $\mathbf{d}(\mathbf{d}+1)/2$ for $\mathbf{p}_n(L_n + 1)$ by symmetry. We write $\mathbf{p}_n(i)$ for $\mathbf{p}(i)$ to specify the dependency of $\mathbf{p}(i)$ on n.

  The ∞-norm $\|\cdot\|_\infty$ denotes the maximum-entry norm of the object and the 0-norm $\|\cdot\|_0$ denotes the number of non-zero entries of the object.

Application to Deep Learning of Diffusion Matrix

- For a rectangle D and a number K, let

$$B^{\beta}(D,K) \;=\; \left\{ f \in C(D);\ \|f\|_{C_b^{\ell,\beta-\ell}(\mathrm{Int}(D))} \le K \right\}.$$

- Suppose that the true function S^* admits a representation

$$S^* \;=\; \mathtt{S}_q \circ \cdots \circ \mathtt{S}_0,$$

where $\mathtt{S}_i = (\mathtt{S}_{i,j})_{j\in[d_{i+1}]} : [a_i,b_i]^{d_i} \to [a_{i+1},b_{i+1}]^{d_{i+1}}$ satisfies $\mathtt{S}_{i,j} \in B^{\beta_i}([a_i,b_i]^{\mathtt{q}_i}, K)$ $(j \in [d_{i+1}])$ for $\mathtt{q}_i \in \mathbb{N}$, and a_i, b_i $(|a_i|, |b_i| \le K)$. $d_0 = \mathtt{d}$ and $d_q = \mathtt{d}^2$ である.

- Suppose $\beta_i \ge 2$. The efficient smoothness is $\beta_i^* \equiv \beta_i \prod_{j=i+1}^{q} (\beta_j \wedge 1) = \beta_i$, and the key convergence rate is

$$\phi_n \;=\; \max_{i=0,\ldots,q} n^{-\frac{2\beta_i}{2\beta_i+\mathtt{q}_i}}.$$

Let $\beta = \min_{i=0,\ldots,q} \beta_i$.

Application to Deep Learning of Diffusion Matrix

Theorem 4 (Gloter and Y). If the number of layers L_n and the sparsity are controlled suitably, and if

$$\Delta_n + \widehat{\Delta}_n \leq C'\big(\phi_n L_n (\log n)^2 + h^2\big) \quad (n \in \mathbb{N}) \tag{20}$$

for some constant C', then there exists a constant C such that

$$\overline{\mathrm{R}}_n \vee \overline{\mathrm{R}}_n^p \leq C\big(\phi_n L_n (\log n)^2 + h^2\big) \quad (n \in \mathbb{N}). \tag{21}$$

- **In particular, if we take $\sum_{i=0}^{q} \log_2 4(\mathbf{q}_i \vee \beta_i) \log_2 n \approx L_n$, then there exists a constant C such that**

$$\overline{\mathrm{R}}_n \vee \overline{\mathrm{R}}_n^p \leq C\big(\phi_n (\log n)^3 + h^2\big) \quad (n \in \mathbb{N}). \tag{22}$$

- **This bound is optimal.**

^o^/

Happy 80th Birthday, Yury!

With deepest gratitude for your monumental contributions to our field.

You are my torch forever.



Generalization error

- If $\boldsymbol{X}$ is stationary, then the modified generalization error $\mathbb{R}_n^\circ$ of the estimator $\widehat{\boldsymbol{S}}_n$ is expressed as

$$\mathbb{R}_n^\circ = \mathbb{R}_n^\circ(\widehat{\boldsymbol{S}}_n, \boldsymbol{S}^*) \;=\; \boldsymbol{E}\big[\boldsymbol{U}(\overline{\boldsymbol{X}}_0; \widehat{\boldsymbol{S}}_n)\big]$$

- $|\boldsymbol{S}| = \big(\sum_{i,j} \boldsymbol{S}_{i,j}^2\big)^{1/2}$ for a matrix $\boldsymbol{S} = (\boldsymbol{S}_{i,j})$.
- Let $\widehat{\boldsymbol{S}}_n^0$ be a DL estimator with the contrast function $\Phi_n(\boldsymbol{S}, \boldsymbol{Z})$. $\widehat{\boldsymbol{S}}_n^0$ is $\boldsymbol{\sigma}[(\boldsymbol{X}_{t_j}, \boldsymbol{Y}_{t_j})_{j \in \mathbb{I}_{0,n}}]$-measurable. We assume that

$$\widehat{\boldsymbol{S}}_n^0 \;\in\; \mathfrak{F}_n \quad \boldsymbol{a.s.} \tag{23}$$

for a family $\mathfrak{F}_n$ satisfying

$$\mathfrak{F}_n \subset \big\{\boldsymbol{S}|_{\mathcal{X}};\; \boldsymbol{S} \in \mathbb{S}\big\} \quad (\boldsymbol{n} \in \mathbb{N}). \tag{24}$$

A family $\mathfrak{F}_n$ satisfying (24) is easily constructed, for example, as the image of a bounded set of $\mathbf{m} \times \mathbf{m}$ symmetric matrices by the exponential map.

Generalization error

- We assume that $\boldsymbol{\sigma} \in C_p^2(\mathbb{R}^{\mathbf{d}}; \mathbb{R}^{\mathbf{m}} \otimes \mathbb{R}^{\mathbf{r}})$ and $\boldsymbol{b} \in C_p^2(\mathbb{R}^{\mathbf{d}}; \mathbb{R}^{\mathbf{m}})$. The subscript $\boldsymbol{p}$ means that the any order (≤ 2) of derivative of each element is of at most polynomial growth.
- Let $\mathcal{D}$ is an open set in $\mathbb{R}^{\mathbf{d}}$. Let $\beta > 0$ and ℓ the maximum integer satisfying $\ell < \beta$. Denote by $C_b^{\ell,\beta-\ell}(\mathcal{D})$ the set of ℓ-times continuously differentiable function f on $\mathbb{R}^{\mathbf{d}}$ such that $\|f\|_{C_b^{\beta}(\mathcal{D})} < \infty$, where

$$\|f\|_{C_b^{\ell,\beta-\ell}(\mathcal{D})} = \sum_{i:0\leq i\leq \ell} \sup_{x\in\mathcal{D}} |\partial^i f(x)| + \sup_{x,y\in\mathcal{D},x\neq y} \frac{|\partial^\ell f(x) - \partial^\ell f(y)|}{|x-y|^{\beta-\ell}}. \quad (25)$$

  In (25), $\partial^i f^{(i)}$ denotes the i-th derivative (tensor) of f. The space $C_b^{\ell,\beta-\ell}(\mathcal{D})$ is regarded as a measurable space equipped with its Borel σ-field generated by the topology associated with the norm $\|f\|_{C_b^{\ell,\beta-\ell}(\mathcal{D})}$.
- Assume $\mathcal{W} \in C_b^{\ell,\beta-\ell}(\mathbb{R}^{\mathbf{d}})$.
- We assume

$$\boldsymbol{\sigma}\big|_{\{x\in\mathbb{R}^{\mathbf{d}};\mathcal{W}(x)>0\}} \in C^{\ell,\beta-\ell}(\{x \in \mathbb{R}^{\mathbf{d}}; \mathcal{W}(x) > 0\}; \mathbb{R}^{\mathbf{m}} \otimes \mathbb{R}^{\mathbf{r}}).$$

  In particular, $\beta \geq 2$.

Generalization error

- Suppose that

$$\limsup_{n\to\infty} \frac{\log\log n}{\log \mathcal{N}_n} < \infty. \tag{26}$$

- Write

$$\lambda_n = \{n^{-1}(\log n)\log \mathcal{N}_n\}^{1/2}.$$

- Assume

$$\limsup_{n\to\infty} \lambda_n \ < \ \infty. \tag{27}$$

  We are assuming that $n \in \mathbb{Z}_{\geq 2}$.

- We assumed that

$$\lim_{n\to\infty} \big(h \log n\big) \ = \ 0. \tag{28}$$

- Assume that

$$\lim_{n\to\infty} \frac{nh}{\log n} \ = \ \infty. \tag{29}$$

Generalization error

- We will use the following notation:

$$|S|_n = \left(n^{-1} \sum_{j \in \mathbb{I}_n} |S(X_{t_{j-1}})|^2 \mathcal{W}(X_{t_{j-1}}) \right)^{1/2}$$

for a function $S : \mathcal{X} \to \mathbb{R}^{\mathbf{m}} \otimes \mathbb{R}^{\mathbf{m}}$. $|S|_n$ is well defined; $|S(x)|\mathcal{W}(x)$ should read 0 when $\mathcal{W}(x) = 0$.

- Accordingly, for $|S_1 - S_0|_n$, we have

$$|S_1 - S_0|_n^2 = n^{-1} \sum_{j \in \mathbb{I}_n} |S_1(X_{t_{j-1}}) - S_0(X_{t_{j-1}})|^2 \mathcal{W}(X_{t_{j-1}})$$

for matrices S_0 and S_1 of the same size.

- Let $\mathcal{X}$ be an open set of $\mathbb{R}^{\mathbf{d}}$ such that supp $\mathcal{W} \subset \mathcal{X}$, where supp $\mathcal{W} = \overline{\{x \in \mathbb{R}^{\mathbf{d}}; \mathcal{W}(x) > 0\}}$.

Suppose that

$$\mathbb{B} \subset C(\mathcal{X}; \mathcal{S}_+) \cap C_b^{\ell,\beta}(\mathcal{X}; \mathbb{R}^{\mathbf{m}} \otimes \mathbb{R}^{\mathbf{m}}).$$

The set $C_b^{\ell,\beta}(\mathcal{X}; \mathbb{R}^{\mathbf{m}} \otimes \mathbb{R}^{\mathbf{m}})$ is identified with $\left(C_b^{\ell,\beta}(\mathcal{X})\right)^{\mathbf{m}^2}$. Since $C_b^{\ell,\beta}(\mathcal{X}; \mathbb{R}^{\mathbf{m}} \otimes \mathbb{R}^{\mathbf{m}})$ has a topology associated with the norm $\| \cdot \|_{C^{\ell,\beta-\ell}(\mathcal{X})}$, it is regarded as a measurable space.

Generalization error

- Assume that $\mathbb{B}$ is bounded in $C_b^{\ell,\beta}(\mathcal{X};\mathbb{R}^{\mathsf{m}}\otimes\mathbb{R}^{\mathsf{m}})$.
- Suppose that $S^*|_{\mathcal{X}} \in \mathbb{B}$.
- Denote by $\mathbb{S}_{\mathsf{m}}$ the set of $\mathsf{m}\times\mathsf{m}$ symmetric matrices. Let $\widehat{S}_n$ be a $\mathbb{B}$-valued measurable map satisfying

$$\widehat{S}_n \in \mathbb{B} \quad a.s.$$

Define $\widehat{\Delta}_n$ as

$$\widehat{\Delta}_n = E\left[\left|\widehat{S}_n - \widehat{S}_n^0\right|_n^2 - \inf_{S\in\mathbb{B}}\left|S - \widehat{S}_n^0\right|_n^2\right].$$

Generalization error

- The predictable empirical risk of $\widehat{S}_n$ is define by

$$\begin{aligned}\boldsymbol{\mathcal{E}}_n^p &= \frac{1}{2n}\sum_{j\in\mathbb{I}_n}\left\{(\widehat{S}_n(X_{t_{j-1}})^{-1} - S^*(X_{t_{j-1}})^{-1})\big[S^*(X_{t_{j-1}})\big] + \log\frac{\det \widehat{S}_n(X_{t_{j-1}})}{\det S^*(X_{t_{j-1}})}\right\}\mathcal{W}(X_{t_{j-1}}) \\ &= n^{-1}\sum_{j\in\mathbb{I}_n} U(X_{t_{j-1}}, \widehat{S}_n). \end{aligned} \tag{30}$$

  Let $\mathrm{R}_n^p = \boldsymbol{E}[\boldsymbol{\mathcal{E}}_n^p]$.

- The empirical risk of $\widehat{S}_n^0$ is define by

$$\begin{aligned}\boldsymbol{\mathcal{E}}_n^0 &= \Phi_n(\widehat{S}_n^0, Z) - \Phi_n(S^*, Z) \\ &= \frac{1}{2n}\sum_{j\in\mathbb{I}_n}\left\{(\widehat{S}_n^0(X_{t_{j-1}})^{-1} - S^*(X_{t_{j-1}})^{-1})\big[h^{-1}(\Delta_j Y)^{\otimes 2}\big] + \log\frac{\det \widehat{S}_n^0(X_{t_{j-1}})}{\det S^*(X_{t_{j-1}})}\right\}\mathcal{W}(X_{t_{j-1}})\end{aligned} \tag{31}$$

  Let $\mathrm{R}_n^0 = \boldsymbol{E}[\boldsymbol{\mathcal{E}}_n^0]$.

- The predictable empirical risk of $\widehat{S}_n^0$ is define by

$$\boldsymbol{\mathcal{E}}_n^{0,p} = \frac{1}{2n}\sum_{j\in\mathbb{I}_n}\left\{(\widehat{S}_n^0(X_{t_{j-1}})^{-1} - S^*(X_{t_{j-1}})^{-1})\big[S^*(X_{t_{j-1}})\big] + \log\frac{\det \widehat{S}_n^0(X_{t_{j-1}})}{\det S^*(X_{t_{j-1}})}\right\}\mathcal{W}(X_{t_{j-1}}) \tag{32}$$

  Let $\mathrm{R}_n^{0,p} = \boldsymbol{E}[\boldsymbol{\mathcal{E}}_n^{0,p}]$.

Generalization error

- **Strategy.**

  (i) $\mathbf{R}_n^p$ **is estimated by** $\mathbf{R}_n^{0,p}$

  (ii) Estimate the statistical error $\mathbf{R}_n^{0,p} - \mathbf{R}_n^0$ **with a large deviation argument**

  (iii) $\mathbf{R}_n^0$ **is controlled due to optimization**

  (iv) Obtain an estimate of $\mathbf{R}_n^p$

  (v) $\overline{\mathbf{R}}_n^p$ **is estimated by** $\mathbf{R}_n^p$ **with the comparison theorem**

  (vi) Estimate $\overline{\mathbf{R}}_n - \overline{\mathbf{R}}_n^p$ **by large deviation argument.**

$$\overline{\mathbf{R}}_n \lesssim \overline{\mathbf{R}}_n^p \lesssim \mathbf{R}_n^p \lesssim \mathbf{R}_n^{0,p} \lesssim \mathbf{R}_n^0$$

Generalization error

- Define $\boldsymbol{\Delta}_n$ as

$$\boldsymbol{\Delta}_n \;=\; \boldsymbol{E}\left[\boldsymbol{\Phi}_n(\widehat{\boldsymbol{S}}_n^0;\boldsymbol{Z}) - \inf_{\boldsymbol{S}\in\mathfrak{F}_n} \boldsymbol{\Phi}_n(\boldsymbol{S};\boldsymbol{Z})\right] \tag{33}$$

- Recall

$$\widehat{\boldsymbol{\Delta}}_n \;=\; \boldsymbol{E}\left[\left|\widehat{\boldsymbol{S}}_n - \widehat{\boldsymbol{S}}_n^0\right|_n^2 - \inf_{\boldsymbol{S}\in\mathbb{B}} \left|\boldsymbol{S} - \widehat{\boldsymbol{S}}_n^0\right|_n^2\right].$$

Theorem 5. There exists a constant C such that

$$\overline{\mathrm{R}}_n \vee \overline{\mathrm{R}}_n^p \;\le\; \boldsymbol{C}\Bigg(\boldsymbol{\Delta}_n + \widehat{\boldsymbol{\Delta}}_n + \inf_{\boldsymbol{S}\in\mathfrak{F}_n} \boldsymbol{E}\left[\boldsymbol{n}^{-1}\sum_{\boldsymbol{j}\in\mathbb{I}_n} \boldsymbol{U}(\boldsymbol{X}_{\boldsymbol{t}_{j-1}},\boldsymbol{S})\right] + \boldsymbol{n}^{-1}(\log \boldsymbol{n})\log \boldsymbol{\mathcal{N}}_n + \boldsymbol{\delta}_n \log \boldsymbol{n} + \boldsymbol{h}_n^2\Bigg). \tag{34}$$

Generalization error

- **We request**

$$\lim_{n\to\infty} T\gamma_n^{\frac{\mathbf{d}}{2\beta}}(\log n)^{-3} \;=\; \infty. \tag{35}$$

- **Denote by $\sim_{\log}$ the equivalence under neglecting $\log n$ factors.**
- **For example, in the case where**

$$\log \mathcal{N}_n \;\sim_{\log}\; \gamma_n \sim_{\log} n^{-\frac{2\beta}{2\beta+\mathbf{d}}}, \tag{36}$$

then Condition (35) is rephrased as

$$hn^{\frac{2\beta}{2\beta+\mathbf{d}}} \;\to_{\log}\; \infty. \tag{37}$$

In this case, it must hold that $2\beta > \mathbf{d}$. Then h must satisfy $h \gg n^{-\frac{2\beta}{2\beta+\mathbf{d}}}$.

Neoclassical/Metamodern description

- **High-dimensional paradigm v.s. Fixed-dimensional paradigm**
- **Asymptotic theory for non-identifiable models**
  - **Neural networks are non-identifiable models. The true values set Θ^*.**
  - **ex. A counting process N_t having an intensity process**

$$\lambda_t = g + \sum_{i=1}^{\mathsf{a}} \alpha_i \exp(\beta_i X_t^i) \quad (g \geq 0,\ \alpha_i \geq 0,\ \beta_i \in \mathbb{R})$$

  - **The regularized estimator converges to a point θ^* in Θ^*.**
  - **The asymptotic distribution of the error of the estimator can be specified. (J. Yoshida and Y 2024)**
  - **The rate of convergence is the traditional one, e.g., $\sqrt{n}$.**
  - **The prediction error can be captured more sharply than the high-dimensional framework.**

Ibragimov-Khasminskii theory for quasi-likelihood analysis

Ibragimov-Khasminskii theory for QLA: a formulation

- Ω: a sample space
- Θ: a bounded open set in $\mathbb{R}^{\mathbf{p}}$, the parameter space
- $n \in \mathbb{N}$
- $\mathbb{H}_n : \Omega \times \Theta \to \mathbb{R}$: a random field (quasi-log likelihood)
- $a_n \in \mathbf{GL}(\mathbb{R}^{\mathbf{p}})$, $a_n \to 0$ $(n \to \infty)$
- $\theta = \theta^* + a_n u$
- $\mathbb{U}_n = \{u \in \mathbb{R}^{\mathbf{p}};\ \theta^* + a_n u \in \Theta\} \quad (\subset T_{\theta^*}\Theta)$
- Quasi-likelihood ratio process

$$\mathbb{Z}_n(u) = \exp\left\{\mathbb{H}_n(\theta^* + a_n u) - \mathbb{H}_n(\theta^*)\right\} \quad (u \in \mathbb{U}_n)$$

Locally asymptotically quadratic random filed

- $\mathbb{Z}_n$ is Locally Asymptotically Quadratic (LAQ) if

$$\mathbb{Z}_n(u) = \exp\left(\Delta_n[u] - \frac{1}{2}\Gamma[u^{\otimes 2}] + r_n(u)\right)$$

  - Δ_n: a random vector (linear form)
  - Γ: a random bilinear form
  - $r_n(u) \to^p 0$ as $n \to \infty$

- Notation.
  $v[u] = \sum_i v_i u^i$,
  $\Gamma[u^{\otimes 2}] = \sum_{i,j} \Gamma_{i,j} u^i u^j$
  for $v = (v_i)$, $\Gamma = (\Gamma_{i,j})$ and $u = (u^i)$.

Ibragimov-Khasminskii theory for QLA

- $\mathbb{Z}_n$ is Locally Asymptotically Quadratic (LAQ) :

$$\mathbb{Z}_n(u) = \exp\left(\Delta_n[u] - \frac{1}{2}\Gamma[u^{\otimes 2}] + r_n(u)\right)$$

- Polynomial type large deviation inequality

$$P\left[\sup_{u\in\mathbb{V}_n(r)} \mathbb{Z}_n(u) \geq e^{-r^\alpha}\right] \leq \frac{C_L}{r^L} \quad (r > 0,\ n \in \mathbb{N})$$

where $\mathbb{V}_n(r) = \{u \in \mathbb{U}_n;\ |u| \geq r\}$, $\mathbb{U}_n = \{u \in \mathbb{R}^{\mathbf{p}};\ \theta^* + a_n u \in \Theta\}$

- If $(\Delta_n, \Gamma) \to^d (\Delta, \Gamma)$, then the convergence of the likelihood ratio process

$$\mathbb{Z}_n \to^d \mathbb{Z} \tag{38}$$

on some function space $\widehat{C}(\mathbb{R}^{\mathbf{p}})$ of continuous functions (for properly extended $\mathbb{Z}_n$).

- $\mathbb{Z}(u) = \exp\left(\Delta[u] - \frac{1}{2}\Gamma[u^{\otimes 2}]\right)$

Ibragimov-Khasminskii theory for QLA

- $\hat{\theta}_n^M$: the quasi-maximum likelihood estimator (QMLE)

$$\mathbb{H}_n\big(\hat{\theta}_n^M\big) = \max_{\theta\in\overline{\Theta}} \mathbb{H}_n(\theta)$$

- $\hat{\theta}_n^B$: the quasi-Bayesian estimator (QBE) for a continuous prior density $\pi : \Theta \to \mathbb{R}_+$ defined by

$$\hat{\theta}_n^B = \left(\int_\Theta \exp(\mathbb{H}_n(\theta))\pi(\theta)d\theta\right)^{-1} \int_\Theta \theta \exp(\mathbb{H}_n(\theta))\pi(\theta)d\theta.$$

Ibragimov-Khasminskii theory for QLA: summary

- $\mathbb{H}_n : \Omega \times \Theta \to \mathbb{R}$: a LAQ random field (quasi-log likelihood)
- $\mathbb{Y}_n(\theta) = \frac{1}{b_n}\big(\mathbb{H}_n(\theta) - \mathbb{H}_n(\theta^*)\big) \to^{L^p} \mathbb{Y}(\theta)$
- $\exists$ a positive r.v. χ_0 and a positive constant ρ s.t.

$$\mathbb{Y}(\theta) = \mathbb{Y}(\theta) - \mathbb{Y}(\theta^*) \le -\chi_0|\theta - \theta^*|^\rho$$

and $P\big[\chi_0 \le r^{-1}\big] \le \dfrac{C_L}{r^L} \quad (r > 0)$

Then

- PLD $P\left[\sup_{u\in\mathbb{V}_n} \mathbb{Z}_n(u) \ge e^{-r^\alpha}\right] \le \dfrac{C_L}{r^L} \quad (r > 0, n \in \mathbb{N})$

  for $\mathbb{Z}_n(u) = \exp\left\{\mathbb{H}_n(\theta^* + a_n u) - \mathbb{H}_n(\theta^*)\right\}$
- $\sup_n \left\|a_n^{-1}(\hat{\theta}_n^{\mathsf{A}} - \theta^*)\right\|_p < \infty \ (\mathsf{A} = M, B)$
- $a_n^{-1}(\hat{\theta}_n^{\mathsf{A}} - \theta^*) \to^d \Gamma^{-1}\Delta$ from $\mathbb{Z}_n \to^d \mathbb{Z} = e^{\Delta[u] - \frac{1}{2}\Gamma[u^{\otimes 2}]}$

A sufficient condition for the polynomial type large deviation (PLD) inequality

- $b_n = \{\lambda_{\min}(a_n' a_n)\}^{-1}$ (e.g. $b_n = n$ for $a_n = n^{-1/2} I_p$)
- For $\mathbb{Y}_n(\theta) = \frac{1}{b_n}\big(\mathbb{H}_n(\theta) - \mathbb{H}_n(\theta^*)\big)$,

  $\mathbb{Y}_n(\theta) \to^{L^p} \mathbb{Y}(\theta)$ (and slightly more)
- Assumption.
  $\exists$ a positive r.v. χ_0 and a positive constant ρ s.t.

$$\mathbb{Y}(\theta) = \mathbb{Y}(\theta) - \mathbb{Y}(\theta^*) \le -\chi_0 |\theta - \theta^*|^\rho,$$

  and for $L > 0$,

$$P\big[\chi_0 \le r^{-1}\big] \le \frac{C_L}{r^L} \quad (r > 0)$$

- Then the PLD inequality holds. (Y 2011)

$$P\left[\sup_{u \in \mathbb{V}_n(r)} \mathbb{Z}_n(u) \ge e^{-r^\alpha}\right] \le \frac{C_L}{r^L} \quad (r > 0,\ n \in \mathbb{N})$$

Example: stochastic regression model

- An **m**-dimensional Itô process satisfying the stochastic differential equation

$$dY_t = b_t dt + \sigma(X_t, \theta) dw_t, \quad t \in [0, T],$$

- Data: $\mathbf{Z}_n = (X_{t_j}, Y_{t_j})_{0 \le j \le n}$ with $t_j = jh$ for $h = h_n = T/n$.
- T fixed

Stochastic regression model

- An $\mathbf{m}$-dimensional Itô process satisfying the stochastic differential equation

$$dY_t = b_t dt + \sigma(X_t, \theta) dw_t, \quad t \in [0, T] \tag{39}$$

- w: an r-dimensional standard Wiener process
- b and X: progressively measurable processes with values in $\mathbb{R}^{\mathbf{m}}$ and $\mathbb{R}^{\mathbf{d}}$, respectively. b is unobservable, completely unknown.
- σ: an $\mathbb{R}^{\mathbf{m}} \otimes \mathbb{R}^r$-valued function defined on $\mathbb{R}^{\mathbf{d}} \times \overline{\Theta}$,
- Θ: a bounded domain in $\mathbb{R}^{\mathbf{p}}$
- θ^* denotes the true value of θ.
- Data: $\mathbf{Z}_n = (X_{t_j}, Y_{t_j})_{0 \le j \le n}$ with $t_j = jh$ for $h = h_n = T/n$. T is fixed.
- For example, when $b_t = b(Y_t, t)$ and $X_t = (Y_t, t)$, Y is the time-inhomogeneous diffusion process.

Quasi-likelihood

- **Quasi-log likelihood function:**

$$\mathbb{H}_n(\boldsymbol{\theta}) = -\frac{n\mathbf{m}}{2}\log(2\pi h) - \frac{1}{2}\sum_{j=1}^{n}\Big\{\log\det \boldsymbol{S}(\boldsymbol{X}_{t_{j-1}},\boldsymbol{\theta}) + h^{-1}\boldsymbol{S}^{-1}(\boldsymbol{X}_{t_{j-1}},\boldsymbol{\theta})[(\boldsymbol{\Delta}_j\boldsymbol{Y})^{\otimes\mathbf{2}}]\Big\},$$

$$\boldsymbol{S} = \boldsymbol{\sigma}^{\circledast\mathbf{2}} = \boldsymbol{\sigma}\boldsymbol{\sigma}^{\star},$$
$$\boldsymbol{\Delta}_j\boldsymbol{Y} = \boldsymbol{Y}_{t_j} - \boldsymbol{Y}_{t_{j-1}}.$$

- **a local Gaussian approximation of the transition probability**

$$d\boldsymbol{Y}_t = \boldsymbol{b}_t dt + \boldsymbol{\sigma}(\boldsymbol{X}_t,\boldsymbol{\theta})d\boldsymbol{w}_t, \quad t \in [0,\boldsymbol{T}],$$

QLA estimators

- $\hat{\theta}_n^M$: the maximum likelihood type estimator defined as

$$\mathbb{H}_n(\hat{\theta}_n^M) = \sup_{\theta \in \overline{\Theta}} \mathbb{H}_n(\theta). \tag{40}$$

- $\hat{\theta}_n^B$: the Bayes type estimator for a prior density $\pi : \Theta \to \mathbb{R}_+$ defined as

$$\hat{\theta}_n^B = \left(\int_\Theta \exp(\mathbb{H}_n(\theta)) \pi(\theta) d\theta \right)^{-1} \int_\Theta \theta \exp(\mathbb{H}_n(\theta)) \pi(\theta) d\theta. \tag{41}$$

We assume that π is continuous and $0 < \inf_{\theta \in \Theta} \pi(\theta) \leq \sup_{\theta \in \Theta} \pi(\theta) < \infty$.

The Key index for separation of distributions (1)

- $\mathbb{Y}_n(\theta) = \frac{1}{n}\big(\mathbb{H}_n(\theta) - \mathbb{H}_n(\theta^*)\big)$: **quasi-log likelihood ratio**
- $\mathbb{Y}$: **the limit random field of** $\mathbb{Y}_n$**, more precisely,**

$$\sup_{n\in\mathbb{N}} \left\| \sup_{\theta\in\Theta} \sqrt{n}|\mathbb{Y}_n(\theta) - \mathbb{Y}(\theta)| \right\|_q < \infty$$

for every $q > 1$**. Next slide for** $\mathbb{Y}(\theta)$**.**

Quasi-log likelihood function:

$$\mathbb{H}_n(\theta) = -\frac{nm}{2}\log(2\pi h) - \frac{1}{2}\sum_{j=1}^{n}\Big\{ \log\det S(X_{t_{j-1}},\theta) + h^{-1}S^{-1}(X_{t_{j-1}},\theta)[(\Delta_j Y)^{\otimes 2}] \Big\},$$

$$S = \sigma^{\circledast 2} = \sigma\sigma^\star, \quad \Delta_j Y = Y_{t_j} - Y_{t_{j-1}}.$$

The Key index for separation of distributions (2)

Key index:

$$\chi_0 = \inf_{\theta \neq \theta^*} \frac{-\mathbb{Y}(\theta)}{|\theta - \theta^*|^2}.$$

For the stochastic regression model,

$$\mathbb{Y}(\theta) = -\frac{1}{2T}\int_0^T \left\{ \log\left(\frac{\det S(X_t,\theta)}{\det S(X_t,\theta^*)}\right) + \mathrm{tr}\left(S^{-1}(X_t,\theta)S(X_t,\theta^*) - I_{\mathsf{m}} \right) \right\} dt.$$

[H2] For every $L > 0$, there exists $C_L > 0$ such that

$$P\left[\chi_0 \leq r^{-1}\right] \leq \frac{C_L}{r^L}$$

for all $r > 0$.

QLA for volatility under $[H2]$:

Theorem 6. (Uchida and Y SPA2013)
Under $[H2]$ (non-degeneracy of χ_0),

(a) $\sqrt{n}(\hat{\theta}_n^{\mathbf{A}} - \theta^*) \to^{d_s} \Gamma^{-1/2}\zeta$ for $\mathbf{A} \in \{M, B\}$

(b) For all continuous functions $h \in C_p(\mathbb{R}^{\mathbf{p}})$,

$$E\left[h(\sqrt{n}(\hat{\theta}_n^{\mathbf{A}} - \theta^*))\right] \to \mathbb{E}\left[h(\Gamma^{-1/2}\zeta)\right] \quad (n \to \infty)$$

for $\mathbf{A} \in \{M, B\}$. Here Γ is a random matrix and ζ is a standard Gaussian random vector $\perp\!\!\!\perp \Gamma$.

$$dY_t = b_t dt + \sigma(X_t, \theta) dw_t, \quad t \in [0, T].$$

$$\Gamma[u^{\otimes 2}] = \frac{1}{2T}\int_0^T \mathrm{Tr}\Big((\partial_\theta S[u])(S)^{-1}(\partial_\theta S[u])(S)^{-1}(X_t, \theta^*)\Big)dt$$

How to obtain the nondegeneracy of the Key index

To verify

[H2] $P\left[\chi_0 \le r^{-1}\right] \le \dfrac{C_L}{r^L}$ $(r > 0)$,

we have

- **Analytic criterion**
- **Geometric criterion**

See Uchida and Yoshida (SPA 2013).

NB. Key index:

$$\chi_0 = \inf_{\theta \neq \theta^*} \frac{-\mathbb{Y}(\theta)}{|\theta - \theta^*|^2}.$$

$$\mathbb{Y}(\theta) = -\frac{1}{2T}\int_0^T \left\{ \log \frac{\det S(X_t, \theta)}{\det S(X_t, \theta^*)} + \mathrm{tr}\left(S^{-1}(X_t, \theta) S(X_t, \theta^*) - I_{\mathsf{m}} \right) \right\} dt.$$

QLA of volatility: Summary

- QLA for the stochastic regression model

$$dY_t = b_t dt + \sigma(X_t, \theta) dw_t, \quad t \in [0, T].$$

In particular,

- QMLE and QBE are asymptotically mixed normal.
- Tail probability estimates for QLA estimators:

$$\sup_n P\left[\sqrt{n}|\hat{\theta}_n^{\mathbf{A}} - \theta| > r\right] \lesssim \frac{1}{r^L} \quad (r > 0)$$

  for $\mathbf{A} \in \{M, B\}$.

- Convergence of moments of the estimators.

Example: asymptotic expansion of the quasi-Likelihood estimators for volatility

- An **m**-dimensional Itô process satisfying the stochastic differential equation

$$dY_t = b_t dt + \sigma(X_t, \theta) dw_t, \quad t \in [0, T]$$

- Data: $\mathbf{Z}_n = (X_{t_j}, Y_{t_j})_{0 \le j \le n}$ with $t_j = jh$ for $h = h_n = T/n$, T fixed
- Asymptotic expansion of $\hat{\theta}_n^{\mathbf{A}}$ ($\mathbf{A} = M, B$)

$$\sup_{f \in \mathcal{E}(M,\gamma)} \left| E\big[f(\hat{u}_n^{\mathbf{A}})\big] - \int f(z) p_n(z) dz \right| = o(n^{-1/2})$$

  - asymptotic expansion for martingales with a mixed normal limit (Y 2013; Nualart and Y 2019)
  - The boundedness of $\hat{u}_n^{\mathbf{A}}$ in $L^{\infty-}$ is used in the proof.

Quasi-Likelihood Analysis for volatility of a process with jumps

- **Joint work with Haruhiko Inatsugu (AISM2021; an updated version arXiv:1806.10706v3)**

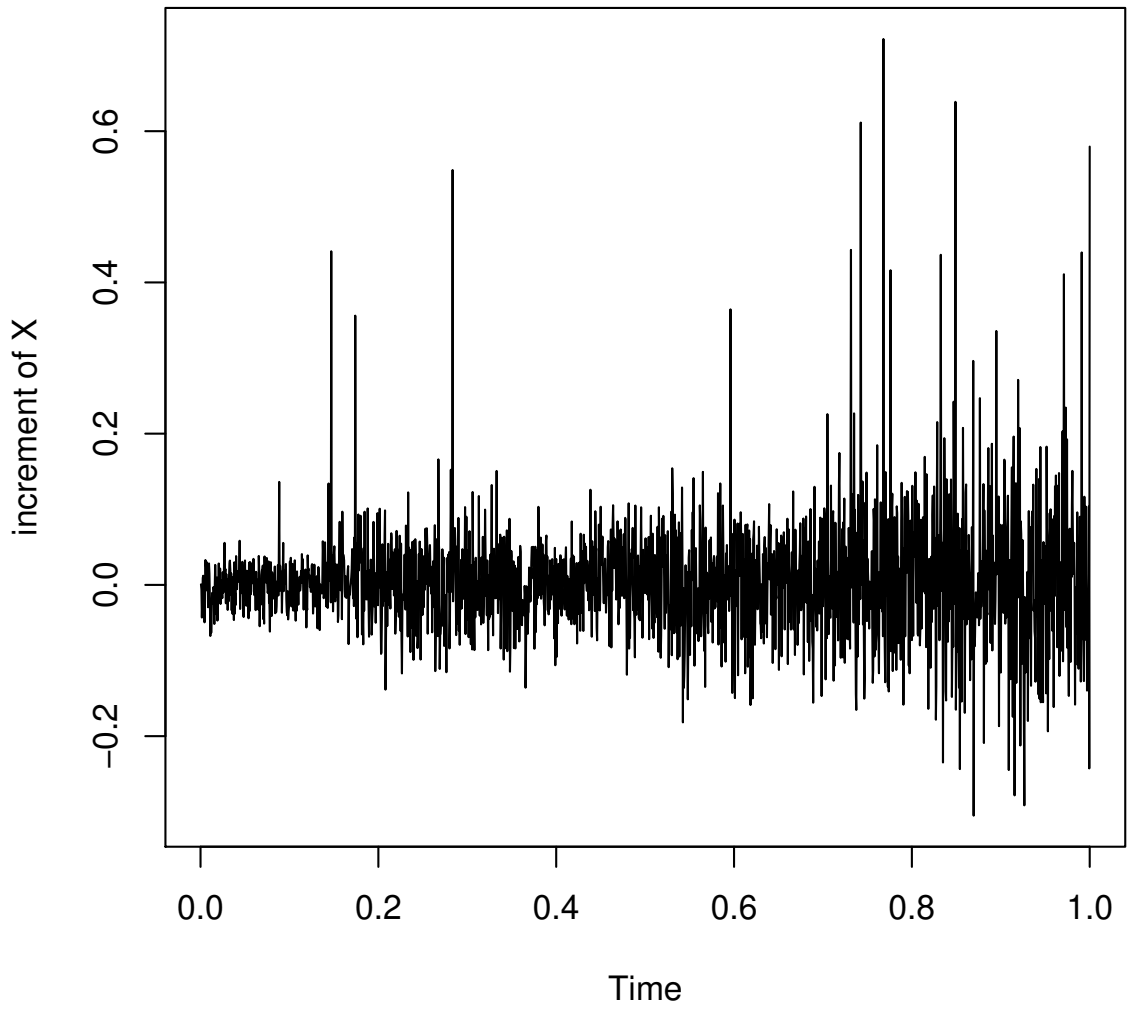
Estimation of the volatility parameter

- an **m-dimensional semimartingale** $Y = (Y_t)_{t\in[0,T]}$

$$Y_t = Y_0 + \int_0^t b_s ds + \int_0^t \sigma(X_s, \theta) dw_s + J_t, \ t \in [0, T]$$

- data $(X_{t_j}, Y_{t_j})_{j=0,1,\ldots,n}$, $t_j = t_j^n = jT/n$.
- T fixed
- That is, the observations are high frequency data.
- The processes b and J are unobservables.
- We want to estimate the true value $\theta^* \in \Theta$ of θ

$$\Delta_j Y = Y_{t_j} - Y_{t_{j-1}}$$



Semimartingale

- an $\mathbf{m}$-dimensional semimartingale $Y = (Y_t)_{t\in[0,T]}$

$$Y_t = Y_0 + \int_0^t b_s ds + \int_0^t \sigma(X_s, \theta) dw_s + J_t,\ t \in [0, T]$$

  - $\sigma : \mathbb{R}^{\mathbf{d}} \times \bar{\Theta} \to \mathbb{R}^{\mathbf{m}} \otimes \mathbb{R}^{\mathbf{r}}$: a continuous function.
  - $J = (J_t)_{t\in[0,T]}$: the jump part of Y, i.e., $J_t = \sum_{s\in[0,t]} \Delta Y_s$, where $\Delta Y_s = Y_s - Y_{s-}$ and $\Delta Y_0 = 0$.

- Assume $\sum_{t\in[0,T]} 1_{\{\Delta J_t \neq 0\}} < \infty$ a.s.

Semimartingale

- an **m**-dimensional semimartingale $Y = (Y_t)_{t\in[0,T]}$

$$Y_t = Y_0 + \int_0^t b_s ds + \int_0^t \sigma(X_s, \theta) dw_s + J_t,\ t \in [0, T]$$

  - a stochastic basis $(\Omega, \mathcal{F}, \mathrm{F}, P)$ with a filtration $\mathrm{F} = (\mathcal{F}_t)_{t\in[0,T]}$.
  - $b = (b_t)_{t\in[0,T]}$: an **m**-dim. progressively measurable process
  - $X = (X_t)_{t\in[0,T]}$: an **d**-dim. càdlàg adapted process
  - $w = (w_t)_{t\in[0,T]}$: an **r**-dim. standard F-Wiener process
  - θ: a parameter in the closure of an open set Θ in $\mathbb{R}^{\mathbf{p}}$

Semimartingale

- This model is a stochastic regression model, but for example, it can express a diffusion type process with jumps ΔJ^X contaminated by exogenous jump noise J^Y:

$$\begin{cases} Y_t = X_t + J_t^Y, \\ X_t = X_0 + \int_0^t b_s ds + \int_0^t \sigma(X_s, \theta) dw_s + J_t^X, \end{cases}$$

with $J = J^X + J^Y$, and as a special case, a jump-diffusion process.

Detection of jumps is indispensable

- Convergence

$$\sum_{j=1}^{n}(\Delta_j Y)^2 \to^p \int_0^T \sigma(X_s,\theta)^2 ds + \sum_{s\le T}(\Delta J_s)^2$$

(one-dimensional case)

- Without removing ΔJ_s, cannot estimate θ.
- Jump filter decides whether $\Delta_j Y$ has jumps

$$\text{Data } \Delta_j Y = Y_{t_j} - Y_{t_{j-1}} \begin{cases} \nearrow & \text{“}\Delta_j Y \text{ has no jump”} \\ \searrow & \text{“}\Delta_j Y \text{ has jumps”} \end{cases}$$

$$Y_t = Y_0 + \int_0^t b_s ds + \int_0^t \sigma(X_s,\theta) dw_s + J_t,\ t \in [0,T]$$

A classical approach: local filter

- A traditional approach for detection of jumps is by a deterministic threshold

$$|\Delta_j Y| > Ch^{\rho} \quad \Rightarrow \quad \text{“}\exists \text{ jumps ”}$$

  with $\rho \in (0, 1/2)$, $h = T/n$ and $\Delta_j Y = Y_{t_j} - Y_{t_{j-1}}$.
- Gives theoretically efficient estimators. Shimizu and Yoshida (SISP2016), Ogihara and Yoshida (SISP2011)
- However, in practice, the performance of the filter strongly depends on the tuning parameters.
- “global filter” will be proposed

A classical approach: local filter

- A traditional approach for detection of jumps is by a deterministic threshold $\left|\Delta_j Y\right| > Ch^{\rho}$ with $\rho \in (0, 1/2)$, $h = T/n$ and $\Delta_j Y = Y_{t_j} - Y_{t_{j-1}}$.
- See Shimizu and Yoshida (SISP2016), Ogihara and Yoshida (SISP2011) for theoretically efficient estimators
- However, in practice, it is difficult to choose tuning parameters C and ρ that give stably precise estimate of the unknown parameter.
- This is a "local" filter because each test of jump is carried out based on one increment $\Delta_j Y$.
- In today's talk, we will propose global filters that dramatically improve estimation of volatility.

Global filter with moving threshold

- $\delta_1 \in (0, 1/2)$.
- $s_n := n - \lfloor Bn^{\delta_1} \rfloor$ with constants $B > 0$.
- $\alpha_n := 1 - s_n/n$. Then $\alpha_n \sim Bn^{\delta_1 - 1}$
- Global jump filter

$$\mathcal{K}_n = \left\{ j \in \{1, ..., n\}; V_j < V_{(s_n)} \right\}$$

where

$$V_j = |(\mathfrak{S}_{n,j-1})^{-1/2} \Delta_j Y|$$

with some positive definite random matrix $\mathfrak{S}_{n,j-1}$, and $V_{(j)}$ is the j-th order statistic of $V_1, ..., V_n$.

$\Delta_j Y = Y_{t_j} - Y_{t_{j-1}}$, $t_j = jT/n$

Global filter with moving threshold: an aggressive quasi-likelihood

Define $\mathbb{H}_n(\theta)$ by

$$\mathbb{H}_n(\theta) = -\frac{1}{2}\sum_{j\in\mathcal{K}_n}\Big\{(q_n)^{-1}h^{-1}S(X_{t_{j-1}},\theta)^{-1}\big[(\Delta_j Y)^{\otimes 2}\big]K_{n,j} + (p_n)^{-1}\log\det S(X_{t_{j-1}},\theta)\Big\}. \tag{42}$$

- The factors $\mathfrak{S}_{n,j-1}$, q_n and p_n can freely be chosen if $\mathfrak{S}_{n,j-1}$ and its inverse are uniformly bounded in $L^{\infty-}$ and if q_n and p_n are sufficiently close to 1. $\bar{S}_{n,j-1}$, $q(\alpha_n)$ and $p(\alpha_n)$ are natural choices for $\mathfrak{S}_{n,j-1}$, q_n and p_n, respectively.
- Asymptotic theoretically, the factors $(q_n)^{-1}$ and $(p_n)^{-1}$ can be replaced by 1, and we may take $\mathfrak{S}_{n,j-1} = I_m$.

$$\mathcal{K}_n = \big\{j \in \{1,...,n\}; V_j < V_{(s_n)}\big\},$$

Global filter with moving threshold: QMLE and QBE

- $\hat{\theta}_n^{M,\alpha_n}$: a QMLE of θ with respect to $\mathbb{H}_n$
- The quasi-Bayesian estimator (QBE) $\hat{\theta}_n^{B,\alpha_n}$ of θ:

$$\hat{\theta}_n^{B,\alpha_n} = \left[\int_\Theta \exp\left(\mathbb{H}_n(\theta)\right)\varpi(\theta)d\theta\right]^{-1} \times \int_\Theta \theta \exp\left(\mathbb{H}_n(\theta)\right)\varpi(\theta)d\theta,$$

where ϖ is a continuous function on Θ satisfying $0 < \inf_{\theta\in\Theta}\varpi(\theta) \leq \sup_{\theta\in\Theta}\varpi(\theta) < \infty$.

$$\mathbb{H}_n(\theta) = -\frac{1}{2}\sum_{j\in\mathcal{K}_n}\Big\{(q_n)^{-1}h^{-1}S(X_{t_{j-1}},\theta)^{-1}\left[\left(\Delta_j Y\right)^{\otimes 2}\right]K_{n,j} + (p_n)^{-1}\log\det S(X_{t_{j-1}},\theta)\Big\}.$$

Global filter with moving threshold: Polynomial type large deviation inequality

- $\mathbb{U}_n = \{u \in \mathbb{R}^{\mathbf{p}};\ \theta^* + n^{-1/2}u \in \Theta\}$.
- **Define the quasi-likelihood ratio random field $\mathbb{Z}_n$ by**

$$\mathbb{Z}_n(u) = \exp\left\{\mathbb{H}_n(\theta^* + n^{-1/2}u) - \mathbb{H}_n(\theta^*)\right\} \quad (u \in \mathbb{U}_n)$$

NB

$$\mathbb{H}_n(\theta) = -\frac{1}{2}\sum_{j\in\mathcal{K}_n}\left\{(q_n)^{-1}h^{-1}S(X_{t_{j-1}},\theta)^{-1}\left[\left(\Delta_j Y\right)^{\otimes 2}\right]K_{n,j} + (p_n)^{-1}\log\det S(X_{t_{j-1}},\theta)\right\}.$$

Global filter with moving threshold: Polynomial type large deviation inequality

- Let

$$\mathbb{Y}(\theta) = -\frac{1}{2T}\int_0^T \left\{ \mathrm{Tr}\left(S(X_t,\theta)^{-1} S(X_t,\theta^*) - I_{\mathsf{m}} \right) + \log \frac{\det S(X_t,\theta)}{\det S(X_t,\theta^*)} \right\} dt.$$

- The key index χ_0 is defined by

$$\chi_0 = \inf_{\theta \neq \theta^*} \frac{-\mathbb{Y}(\theta)}{|\theta - \theta^*|^2}.$$

- Non-degeneracy of χ_0 plays an essential role in the QLA.

[F3] For every positive number L, there exists a constant C_L such that

$$P\left[\chi_0 < r^{-1}\right] \leq C_L\, r^{-L} \qquad (r > 0).$$

Global filter with moving threshold: Polynomial type large deviation inequality

- A polynomial type large deviation inequality is given by the following theorem.

  Theorem 7. Suppose that $[F1]_4$, $[F2']$ and $[F3]$ are fulfilled. Let $c_0 \in (1,2)$. Then, for every positive number L, there exists a constant $C(c_0, L)$ such that

$$P\left[\sup_{u \in \mathbb{V}_n(r)} \mathbb{Z}_n(u) \geq e^{-r^{c_0}}\right] \leq \frac{C(c_0, L)}{r^L}$$

  for all $r > 0$ and $n \in \mathbb{N}$.

- $\mathbb{V}_n(r) = \{u \in \mathbb{U}_n;\ |u| \geq r\}$.

Global filter with moving threshold: rate of convergence of the estiamtors

The polynomial type large deviation inequality for $\mathbb{Z}_n$ in Theorem 7 ensures $L^{\infty-}$-boundedness of the QLA estimators.

Proposition 8. Suppose that $[F1]_4$, $[F2']$ and $[F3]$ are satisfied. Then

$$\sup_{n\in\mathbb{N}} \left\| \sqrt{n}\left(\hat{\theta}_n^{\mathbf{A},\alpha_n} - \theta^*\right) \right\|_p < \infty \qquad (\mathbf{A} = M, B)$$

for every $p > 1$.

Global filter with moving threshold: condition

[F2′] (i) The positive-definite measurable random matrices $\mathfrak{S}_{n,j-1}$ ($k \in \{1,...,\mathsf{k}\}$, $n \in \mathbb{N}$, $j \in \{1,...,n\}$) satisfy

$$\sup_{\substack{k\in\{1,...,\mathsf{k}\}\\ n\in\mathbb{N},\, j\in\{1,...,n\}}} \left(\|\mathfrak{S}_{n,j-1}\|_p + \|(\mathfrak{S}_{n,j-1})^{-1}\|_p\right) < \infty$$

for every $p > 1$.

(ii) Positive numbers q_n and p_n satisfy $|q_n - 1| = o(n^{-1/2})$ and $|1 - p_n| = o(n^{-1/2})$.

Global filter with moving threshold: limit theorem and convergence of moments

---

$[\mathrm{F1}']_\kappa$ Conditions (ii), (iii) and (iv) of $[F1]_\kappa$ are satisfied in addition to

(i) the process X has a representation

$$X_t = X_0 + \int_0^t \tilde{b}_s ds + \int_0^t \tilde{a}_s d\tilde{w}_s + J_t^X \qquad (t \in [0,T])$$

where $J^X = (J_t^X)_{t\in[0,T]}$ is a càdlàg adapted pure jump process, $\tilde{w} = (\tilde{w}_t)_{t\in[0,T]}$ is an $\mathbf{r}_1$-dimensional F-Wiener process, $\tilde{b} = (\tilde{b}_t)_{t\in[0,T]}$ is a $\mathbf{d}$-dimensional càdlàg adapted process and $\tilde{a} = (\tilde{a}_t)_{t\in[0,T]}$ is a progressively measurable processes taking values in $\mathbb{R}^{\mathbf{d}} \otimes \mathbb{R}^{\mathbf{r}_1}$. Moreover,

$$\|X_0\|_p + \sup_{t\in[0,T]} \left(\|\tilde{b}_t\|_p + \|\tilde{a}_t\|_p + \|J_t^X\|_p\right) < \infty$$

for very $p > 1$.

- The Wiener process $\tilde{w}$ is possibly correlated with w.
- The $\mathbf{p} \times \mathbf{p}$ symmetric matrix Γ is defined by the following formula:

$$\Gamma[u^{\otimes 2}] = \frac{1}{2T} \int_0^T \mathrm{Tr}\Big((\partial_\theta S[u])(S)^{-1} (\partial_\theta S[u])(S)^{-1} (X_t, \theta^*) \Big) dt,$$

where $u \in \mathbb{R}^{\mathbf{p}}$.

Global filter with moving threshold: limit theorem and convergence of moments

Recall that

- $\hat{\theta}_n^{M,\alpha_n}$: a QMLE of θ with respect to $\mathbb{H}_n$ given by the aggressive quasi-likelihood $\mathbb{H}_n(\theta)$ of (42).
- $\hat{\theta}_n^{B,\alpha_n}$ denotes the quasi-Bayesian estimator (QBE) of θ with respect to $\mathbb{H}_n$ defined by (42).
- $\hat{u}_n^{\mathsf{A},\alpha_n} := \sqrt{n}\big(\hat{\theta}_n^{\mathsf{A},\alpha_n} - \theta^*\big)$ for $\mathsf{A} \in \{M, B\}$.

Theorem 9. Suppose that $[F1']_4$, $[F2']$ and $[F3]$ are fulfilled. Then

$$E\big[f\big(\hat{u}_n^{\mathsf{A},\alpha_n}\big)\Phi\big] \to \mathbb{E}\big[f\big(\Gamma^{-1/2}\zeta\big)\Phi\big]$$

as $n \to \infty$ for $\mathsf{A} \in \{M, B\}$, any continuous function f of at most polynomial growth, and any $\mathcal{F}$-measurable random variable $\Phi \in \cup_{p>1} L^p$.

Global jump filter: summary

- model

$$Y_t = Y_0 + \int_0^t b_s ds + \int_0^t \sigma(X_s, \theta) dw_s + J_t, \ t \in [0, T]$$

- quasi-log likelihood function with a global filter

$$\mathbb{H}_n(\theta) = -\frac{1}{2} \sum_{j \in \mathcal{K}_n} \Big\{ (q_n)^{-1} h^{-1} S(X_{t_{j-1}}, \theta)^{-1} \big[(\Delta_j Y)^{\otimes 2} \big] K_{n,j} + (p_n)^{-1} \log \det S(X_{t_{j-1}}, \theta) \Big\}.$$

- global jump filter $\mathcal{K}_n$ based on the order statistics
- QLA applied
- $\sup_{n \in \mathbb{N}} \big\| n^{1/2} \big(\hat{\theta}_n^{\mathbf{A}, \alpha} - \theta^* \big) \big\|_p < \infty \quad (\mathbf{A} = M, B)$
- $n^{1/2} \big(\hat{\theta}_n^{\mathbf{A}, \alpha_n} - \theta^* \big) \to^{d_s} \Gamma^{-1/2} \zeta$
- Inatsugu and Y (AISM2021; arXiv:1806.10706v3 updated)

Quasi-Likelihood Analysis applied to various problems

Inference for diffusion processes

- **Kutoyants (Herdermann1984, Springer2004)**
- **Prakasa Rao (MathOperationsforsch1983, Contemp-Math1988)**
- **Dohnal (JApplProbab1987)**
- **Yoshida (JMA1922, AISM2011)**
- **Genon-Catalot and Jacod (AIHP1993)**
- **Bibby and Soerensen (Bernoulli1995)**
- **Kessler (ScandJStatist1997)**
- **Genon-Catalot, Jeantheau and Laredo (Bernoullli1999)**

QLA for ergodic diffusion processess

- QLA for ergodic diffusion: Y (AISM2011)
  - QMLE, adaptive Bayes estimator
- QLA and adaptive procedures: Uchida and Y (SPA2012)
  - reduced sampling rate from $nh^2 \to 0$ to $nh^p \to 0$
  - PLD, moment convergence
- Adaptive Bayes type estimators of ergodic diffusion processes: Uchida and Y (SISP2014)
- Information criterion CIC: Uchida (AISM2010)
  - PLD, Malliavin calculus
- Misspecification affects convergence rate of the volatility parameter: Uchida and Y (ESAIM2011)
  - The convergence rate of the volatility estimator becomes $\sqrt{T}$, not $\sqrt{n}$.
- Hybrid multi-step estimators for stochastic differential equations: Kamatani and Uchida (2015)

QLA for ergodic jump-diffusion processes

- **Ergodic jump-diffusion processes:**
  Shimizu and Y (SISP2006)

$$dX_t = a(X_t, \theta)dt + b(X_t, \sigma)dw_t + \int c(X_{t-}, z, \theta)\tilde{\mu}(dt, dz),$$
$$\mu \sim dt f(z, \theta)dz$$

- **QLA for ergodic jump-diffusion:**
  Ogihara and Y (SISP2011)
- **Gaussian quasi-likelihood random fields for ergodic Lévy driven SDE:**
  Masuda (2013AS)

More with QL methods

- **QLA of volatility in nonsynchronous sampling scheme: Ogihara and Y (SPA2014)**
- **Model selection for volatility prediction: Uchida and Y (2015)**
- **Change point problem for volatility: Iacus and Y (SPA2012)**
- **Distcriminant analysis: Uchida (SAPS2013)**
- **Moment convergence in regularized estimations: Masuda and Yusuke Shimizu (arXiv2014)**
- **Schwarz type model comparison for LAQ models: Eguchi and Masuda (Bernoulli2018)**
- **AIC for the non-concave penalized likelihood method: Yuta Umezu, Yusuke Shimizu, Masuda, Ninomiya (AISM2019)**

More on QL methods

- **Point process Clinet and Y (SPA2017)**
- **Regularization and sparse estimation Kinoshita and Y (arXiv2020)**
- **Least square approximation method for stochastic processes Suzuki and Y (JJSDS2020)**
- **Limit Order Book Muni Toke and Y (Quantitative Finance2017, Quantitative Finance2019, JJSDS2021?)**
- **Global jump filters Inatsugu and Y (AISM2021)**
- **Diffusion plus noise Nakakita, Kaino and Uchida (AISM2021)**

Global jump filters with a fixed α

Global jump filter: fixed α

- For $\alpha \in [0,1)$, Global jump filter is constructed:
- $\mathcal{J}_n(\alpha) = \left\{ j \in \{1,...,n\};\ V_j < V_{(s_n(\alpha))} \right\}$
  - $V_j = \left| (\bar{S}_{n,j-1})^{-1/2} \Delta_j Y \right|$
  - $s_n(\alpha) = \lfloor n(1-\alpha) \rfloor$
  - $V_{(r)}$: r-th order statistic of $\{V_j\}_{j=1,...,n}$
- $\bar{S}_{n,j-1}$ is an initial estimator of $S(X_{t_{j-1}}, \theta^*)$ up to a (possibly unknown) scaling constant, that is,

$$\exists c > 0, \quad \forall j, \quad S(X_{t_{j-1}}, \theta^*) \sim c\bar{S}_{n,j-1}$$

$S = \sigma\sigma^\star$

$Y_t = Y_0 + \int_0^t b_s ds + \int_0^t \sigma(X_s, \theta) dw_s + J_t,\ t \in [0,T]$

Spot volatility estimator for the global jump filter

- prepare statistics $\bar{S}_{n,j-1}$ $(k = 1, ..., \mathsf{k};\ j = 1, ..., n;$ $n \in \mathbb{N})$ such that
- each $\bar{S}_{n,j-1}$ is an initial estimator of $S(X_{t_{j-1}}, \theta^*)$ up to a (possibly unknown) scaling constant, that is,

$$\exists c > 0, \quad \forall j, \quad S(X_{t_{j-1}}, \theta^*) \sim c\bar{S}_{n,j-1}$$

  as precisely stated later.

- E.g., if $S(X_{t_{j-1}}, \theta^*)$ is a constant scaler matrix, this property is obvious.
- If $\bar{S}_{n,j-1}$ is a suitable estimator of $S(X_{t_{j-1}}, \theta^*)$, then we can take $c = 1$.
- This condition imposes no restriction because one can construct a suitable "local" estimator of $S(X_{t_{j-1}}, \theta^*)$.
- We do not assume that $\bar{S}_{n,j-1}$ is $\mathcal{F}_{t_{j-1}}$-measurable.

An example of the initial estimator

- **min type estimator.** Let $(\bar{i}_n)$ be a diverging sequence of positive integers, e.g., $\bar{i}_n \sim h^{-1/2}$. Let

$$\hat{S}_{n,j-1} = \frac{\sum_{i=-\bar{i}_n}^{\bar{i}_n} \left(\Delta_{j-i}Y\right)^{\otimes 2} 1_{\left\{|\Delta_{j-i-K+1}Y| \wedge \cdots \wedge |\Delta_{j-i-1}Y| \geq |\Delta_{j-i}Y|\right\}}}{h \max\left\{1, \sum_{i=-\bar{i}_n}^{\bar{i}_n} 1_{\left\{|\Delta_{j-i-K+1}Y| \wedge \cdots \wedge |\Delta_{j-i-1}Y| \geq |\Delta_{j-i}Y|\right\}}\right\}}.$$

  Here $\Delta_j Y$ reads 0 when $j \leq 0$ or $j > n$. An example of $\bar{S}_{n,j-1}$ is

$$\bar{S}_{n,j-1} = \hat{S}_{n,j-1} 1_{\{\lambda_{min}(\hat{S}_{n,j-1}) > 2^{-1}\epsilon_0\}} + 2^{-1}\epsilon_0 I_{\mathsf{m}} 1_{\{\lambda_{min}(\hat{S}_{n,j-1}) \leq 2^{-1}\epsilon_0\}},$$

  suppose that $\inf_{x,\theta} \lambda_{min}(S(x,\theta)) \geq \epsilon_0$ for some positive constant ϵ_0, where λ_{min} is the minimum eigenvalue of the matrix.

- A localized version of the global jump filter (LGRV) serves as $\bar{S}_{n,j-1}$, and better than the "min" type volatility estimators.

Global jump filter

- The α-quasi-log likelihood function $\mathbb{H}_n(\theta;\alpha)$ is defined by

$$\mathbb{H}_n(\theta;\alpha) = -\frac{1}{2}\sum_{j\in\mathcal{J}_n(\alpha)}\Big\{q(\alpha)^{-1}h^{-1}S(X_{t_{j-1}},\theta)^{-1}\big[(\Delta_j Y)^{\otimes 2}\big]K_{n,j} + p(\alpha)^{-1}\log\det S(X_{t_{j-1}},\theta)\Big\}$$

- $\Delta_j Y = Y_{t_j} - Y_{t_{j-1}}$
- $q(\alpha)$, $p(\alpha)$: balancing constants
- The cap $K_{n,j} = 1_{\left\{|\Delta_j Y| < C_* n^{-\frac{1}{4}}\right\}}$ is not essential and can be removed if a suitable integrability condition is assumed for the jump sizes of J.
- $\mathcal{J}_n(\alpha) = \left\{j \in \{1,...,n\};\ V_j < V_{(s_n(\alpha))}\right\}$

Global jump filter

- Let

$$q(\alpha) = \frac{\mathrm{Tr}\left(\int_{\{|z| \le c(\alpha)^{1/2}\}} z^{\otimes 2} \phi(z; 0, I_{\mathsf{m}}) dz \right)}{\mathrm{Tr}\left(\int_{\mathbb{R}^{\mathsf{m}}} z^{\otimes 2} \phi(z; 0, I_{\mathsf{m}}) dz \right)},$$

$\phi(z; \mu, C)$: the density function of the multivariate $N(\mu, C)$. Equivalently,

$$\begin{aligned} q(\alpha) &= \mathsf{m}^{-1} \mathrm{Tr}\left(\int_{\{|z| \le c(\alpha)^{1/2}\}} z^{\otimes 2} \phi(z; 0, I_{\mathsf{m}}) dz \right) \\ &= \mathsf{m}^{-1} E[V 1_{\{V \le c(\alpha)\}}], \end{aligned}$$

for a random variable $V \sim \chi^2(\mathsf{m})$,

- $c(\alpha)$ is determined by

$$P[V \le c(\alpha)] = 1 - \alpha =: p(\alpha).$$

Global jump filter

- For a tensor $T = (T_{i_1,...,i_k})_{i_1,...,i_k}$, we write

$$T[x_1, ..., x_k] = T[x_1 \otimes \cdots \otimes x_k] = \sum_{i_1,...,i_k} T_{i_1,...,i_k} x_1^{i_1} \cdots x_k^{i_k}$$

for $x_1 = (x_1^{i_1})_{i_1}$, ..., $x_k = (x_k^{i_k})_{i_k}$. We denote $u^{\otimes r} = u \otimes \cdots \otimes u$ (r times). Brackets [ ] stand for a multilinear mapping.

- If $\alpha = 0$, then $\mathcal{J}_n(\alpha) = \{1, ..., n\}$, $c(\alpha) = +\infty$, $p(\alpha) = 1$ and $q(\alpha) = 1$, so the k-th component of $\mathbb{H}_n(\theta; \alpha)$ essentially becomes the ordinary quasi-log likelihood function by local Gaussian approximation.

α-QMLE

- The α-quasi-maximum likelihood estimator of θ (α-QMLE) is any measurable mapping $\hat{\theta}_n^{M,\alpha}$ characterized by

$$\mathbb{H}_n(\hat{\theta}_n^{M,\alpha};\alpha) = \max_{\theta\in\bar{\Theta}} \mathbb{H}_n(\theta;\alpha).$$

- We can construct a quasi-Bayesian estimator (QBE) closely related to $\mathbb{H}_n(\theta;\alpha)$.

$$\mathbb{H}_n(\theta;\alpha) = -\frac{1}{2}\sum_{j\in\mathcal{J}_n(\alpha)} \Big\{ q(\alpha)^{-1}h^{-1}S(X_{t_{j-1}},\theta)^{-1}\big[(\Delta_j Y)^{\otimes 2}\big]K_{n,j} + p(\alpha)^{-1}\log\det S(X_{t_{j-1}},\theta)\Big\}$$

Technical aspects in mathematical validation of the estimators

- The global filter depends on all the data.
- The martingale structure has been destroyed in $\mathbb{H}_n(\theta, \alpha)$.
- Recovering it asymptotically is the key to the validation.
- Our argument deeply relies on the theory of quasi-likelihood analysis (QLA, Y AISM2011).

Conditions

- $\kappa_n = O(n^{1/2})$, $I_{n,j} = \{i \in \{1, ..., n\};\ |i - j| \le \kappa_n\}$
- $\mathtt{I}_{n,j} = \cup_{k \in I_{n,j}} [t_{k-1}, t_k]$. The index set $\mathsf{L}_n^{(k)}$ defined by

$$\mathsf{L}_n = \{j \in \{1, ..., n\};\ N(\mathtt{I}_{n,j}) + N^X(\mathtt{I}_{n,j}) \neq 0\}.$$

[F2] (i) $\bar{S}_{n,j-1}$ are symmetric, invertible and $\sup_{n \in \mathbb{N}} \max_{j=1,...,n} \left\| (\bar{S}_{n,j-1})^{-1} \right\|_p < \infty$ for every $p > 1$.

(ii) There exist positive constants γ_0 and c such that

$$\sup_{n \in \mathbb{N}} \max_{j=1,...,n} n^{\gamma_0} \left\| \left(S(X_{t_{j-1}}, \theta^*) - c\bar{S}_{n,j-1} \right) 1_{j \in (\mathsf{L}_n)^c} \right\|_p < \infty$$

for every $p > 1$.

- The values of c can be unknown since the function $\mathbb{H}_n$ does not involve c.
- When $S(X_{t_{j-1}}, \theta^*)$ is a constant scaler matrix, Condition [$F2$] is satisfied simply by $\bar{S}_{n,j-1} = I_m$.

Polynomial type large deviation inequality for $\mathbb{H}_n^{\beta}(\theta;\alpha)$

- γ_0: the rate of convergence of the initial spot volatility estimator

$$\sup_{n\in\mathbb{N}} \max_{j=1,\dots,n} n^{\gamma_0} \left\| \left(S(X_{t_{j-1}}, \theta^*) - c\bar{S}_{n,j-1} \right) 1_{j\in(\mathsf{L}_n)^c} \right\|_p < \infty$$

Usually $\gamma_0 \in (0, 1/2]$.

- For $\beta \in (0, \gamma_0)$, define an annealed random field by

$$\mathbb{H}_n^{\beta}(\theta;\alpha) = n^{-1+2\beta}\mathbb{H}_n(\theta;\alpha)$$

$$\begin{aligned}
&\mathbb{H}_n(\theta;\alpha) \\
&= -\frac{1}{2} \sum_{j\in\mathcal{J}_n(\alpha)} \Big\{ q(\alpha)^{-1} h^{-1} S(X_{t_{j-1}}, \theta)^{-1} \left[(\Delta_j Y)^{\otimes 2} \right] K_{n,j} \\
&\quad + p(\alpha)^{-1} \log\det S(X_{t_{j-1}}, \theta) \Big\}
\end{aligned}$$

Polynomial type large deviation inequality for $\mathbb{H}_n^{\beta}(\theta;\alpha)$

- Let

$$\begin{aligned}\mathbb{Y}_n(\theta;\alpha) &= n^{-2\beta}\{\mathbb{H}_n^{\beta}(\theta;\alpha) - \mathbb{H}_n^{\beta}(\theta^*;\alpha)\} \\ &= n^{-1}\{\mathbb{H}_n(\theta;\alpha) - \mathbb{H}_n(\theta^*;\alpha)\}.\end{aligned}$$

- Let

$$\begin{aligned}\mathbb{Y}(\theta) = -\frac{1}{2T}\int_0^T \Bigg\{ &\mathrm{Tr}\Big(S(X_t,\theta)^{-1} S(X_t,\theta^*) - I_{\mathsf{m}} \Big) \\ &+ \log \frac{\det S(X_t,\theta)}{\det S(X_t,\theta^*)} \Bigg\} dt.\end{aligned}$$

Polynomial type large deviation inequality

- The key index χ_0 is defined by

$$\chi_0 = \inf_{\theta \neq \theta^*} \frac{-\mathbb{Y}(\theta)}{|\theta - \theta^*|^2}.$$

- Non-degeneracy of χ_0 plays an essential role in the QLA.

[F3] For every positive number L, there exists a constant C_L such that

$$P\left[\chi_0 < r^{-1}\right] \leq C_L\, r^{-L} \qquad (r > 0).$$

Polynomial type large deviation inequality

- An analytic criterion and a geometric criterion are known to insure Condition $[F3]$ when X is a non-degenerate diffusion process. See Uchida and Yoshida (SPA2013) for details.
- Since the proof of this fact depends on short time asymptotic properties, we can modify it by taking the same approach before the first jump even when X has finitely active jumps.
- On the other hand, those criteria can apply to the jump diffusion X without remaking them if we work under localization.

Polynomial type large deviation inequality

- $\mathbb{U}_n^{\beta} = \{u \in \mathbb{R}^{\mathbf{p}};\ \theta^* + n^{-\beta}u \in \Theta\}$
- $\mathbb{V}_n^{\beta}(r) = \{u \in \mathbb{U}_n^{\beta};\ |u| \geq r\}$
- The quasi-likelihood ratio random field $\mathbb{Z}_n^{\beta}(\cdot;\alpha)$ of order β is defined by

$$\mathbb{Z}_n^{\beta}(u;\alpha) = \exp\left\{\mathbb{H}_n^{\beta}(\theta^* + n^{-\beta}u;\alpha) - \mathbb{H}_n^{\beta}(\theta^*;\alpha)\right\}\ (u \in \mathbb{U}_n^{\beta})$$

Theorem 10. Suppose that $[F1]_4$, $[F2]$ and $[F3]$ are fulfilled. Let $c_0 \in (1,2)$. Then, for every positive number L, there exists a constant $C(\alpha,\beta,c_0,L)$ such that

$$P\left[\sup_{u\in\mathbb{V}_n^{\beta}(r)} \mathbb{Z}_n^{\beta}(u;\alpha) \geq e^{-r^{c_0}}\right] \leq \frac{C(\alpha,\beta,c_0,L)}{r^L}$$

for all $r > 0$ and $n \in \mathbb{N}$.

Rate of convergence of the α-QMLE

- By definition, an α-QMLE $\hat{\theta}_n^{M,\alpha}$ of θ with respect to $\mathbb{H}_n(\cdot;\alpha)$ is a QMLE with respect to $\mathbb{H}_n^{\beta}(\cdot;\alpha)$.
- As usual in the QLA theory, the following rate of convergence is a consequence of Theorem 10 (the PLD inequality for $\mathbb{Z}_n^{\beta}(u;\alpha)$).

Proposition 11. Suppose that $[F1]_4$, $[F2]$ and $[F3]$ are satisfied. Then

$$\sup_{n\in\mathbb{N}} \left\| n^{\beta}\left(\hat{\theta}_n^{M,\alpha} - \theta^*\right)\right\|_p < \infty$$

for every $p > 1$ and every $\beta < \gamma_0 \leq \frac{1}{2}$.

$$\mathbb{H}_n(\theta;\alpha) = -\frac{1}{2}\sum_{j\in\mathcal{J}_n(\alpha)} \Big\{ q(\alpha)^{-1} h^{-1} S(X_{t_{j-1}},\theta)^{-1}\left[(\Delta_j Y)^{\otimes 2}\right] K_{n,j} + p(\alpha)^{-1}\log\det S(X_{t_{j-1}},\theta)\Big\}$$

(α, β)-quasi-Bayesian estimator

- **The (α, β)-quasi-Bayesian estimator ((α, β)-QBE) $\hat{\theta}_n^{B,\alpha,\beta}$ of θ is defined by**

$$\hat{\theta}_n^{B,\alpha,\beta} = \left[\int_\Theta \exp\left(\mathbb{H}_n^\beta(\theta;\alpha)\right) \varpi(\theta) d\theta \right]^{-1} \times \int_\Theta \theta \exp\left(\mathbb{H}_n^\beta(\theta;\alpha)\right) \varpi(\theta) d\theta,$$

where ϖ is a continuous function on Θ satisfying $0 < \inf_{\theta\in\Theta} \varpi(\theta) \leq \sup_{\theta\in\Theta} \varpi(\theta) < \infty$.

$$\mathbb{H}_n^\beta(\theta;\alpha) = n^{-1+2\beta} \mathbb{H}_n(\theta;\alpha)$$

$$\mathbb{H}_n(\theta;\alpha) = -\frac{1}{2} \sum_{j\in\mathcal{J}_n(\alpha)} \Big\{ q(\alpha)^{-1} h^{-1} S(X_{t_{j-1}},\theta)^{-1} \left[(\Delta_j Y)^{\otimes 2} \right] K_{n,j} + p(\alpha)^{-1} \log\det S(X_{t_{j-1}},\theta) \Big\}$$

Rate of convergence of (α, β)-QBE

- **Once again Theorem 10 ensures $L^{\infty-}$-boundedness of the error of the (α, β)-QBE:**

Proposition 12. Suppose that $[F1]_4$, $[F2]$ and $[F3]$ are satisfied. Let $\beta \in (0, \gamma_0)$. Then

$$\sup_{n \in \mathbb{N}} \left\| n^{\beta} \left(\hat{\theta}_n^{B,\alpha,\beta} - \theta^* \right) \right\|_p < \infty$$

for every $p > 1$.

Global jump filters

- **Fixed α**
  - **jumps are removed surely (e.g. $\alpha = 0.2, 0.5, 0.8$)**
  - **conservative and does not attain the optimal rate $n^{-1/2}$ of convergence.**
- **Shrinking $\alpha = \alpha_n \to 0$**
  - **efficient estimation by aggressively taking bigger increments**

QLA for volatility parameter: summary

- model

$$Y_t = Y_0 + \int_0^t b_s ds + \int_0^t \sigma(X_s, \theta) dw_s + J_t, \ t \in [0, T]$$

- α-quasi-log likelihood function (fixed α case)

$$\mathbb{H}_n(\theta; \alpha) = -\frac{1}{2} \sum_{j \in \mathcal{J}_n(\alpha)} \Big\{ q(\alpha)^{-1} h^{-1} S(X_{t_{j-1}}, \theta)^{-1} \big[(\Delta_j Y)^{\otimes 2} \big] K_{n,j} + p(\alpha)^{-1} \log \det S(X_{t_{j-1}}, \theta) \Big\}$$

- global jump filter $\mathcal{J}_n(\alpha)$ based on the order statistics
- QLA applied
- fixed $\alpha \ \Rightarrow \ \sup_{n \in \mathbb{N}} \big\| n^{\beta} \big(\hat{\theta}_n^{M,\alpha} - \theta^* \big) \big\|_p \ < \ \infty \quad (\beta < \gamma_0 \leq 1/2)$
- shrinking $\alpha = \alpha_n \to 0 \ \Rightarrow \ n^{1/2} \big(\hat{\theta}_n^{M,\alpha_n} - \theta^* \big) \to^{d_s} \Gamma^{-1/2} \zeta$
- similar results for the quasi-Bayesian estimators
- Inatsugu and Y (AISM2021; arXiv:1806.10706v3 updated)

Numerical studies

- **We use the one-dimensional Ornstein-Uhlenbeck process with jumps**

$$dX_t = -\eta X_t dt + \sigma dW_t + dJ_t. \tag{43}$$

- J **is a one-dimensional compound Poisson process defined by**

$$J_t = \sum_{i=1}^{N_t} \xi_i, \qquad \xi_i \sim \mathcal{N}(0, \varepsilon),$$

where $\varepsilon > 0$ **and** $N = \{N_t\}$ **is a Poisson process with intensity** $\lambda > 0$**.**

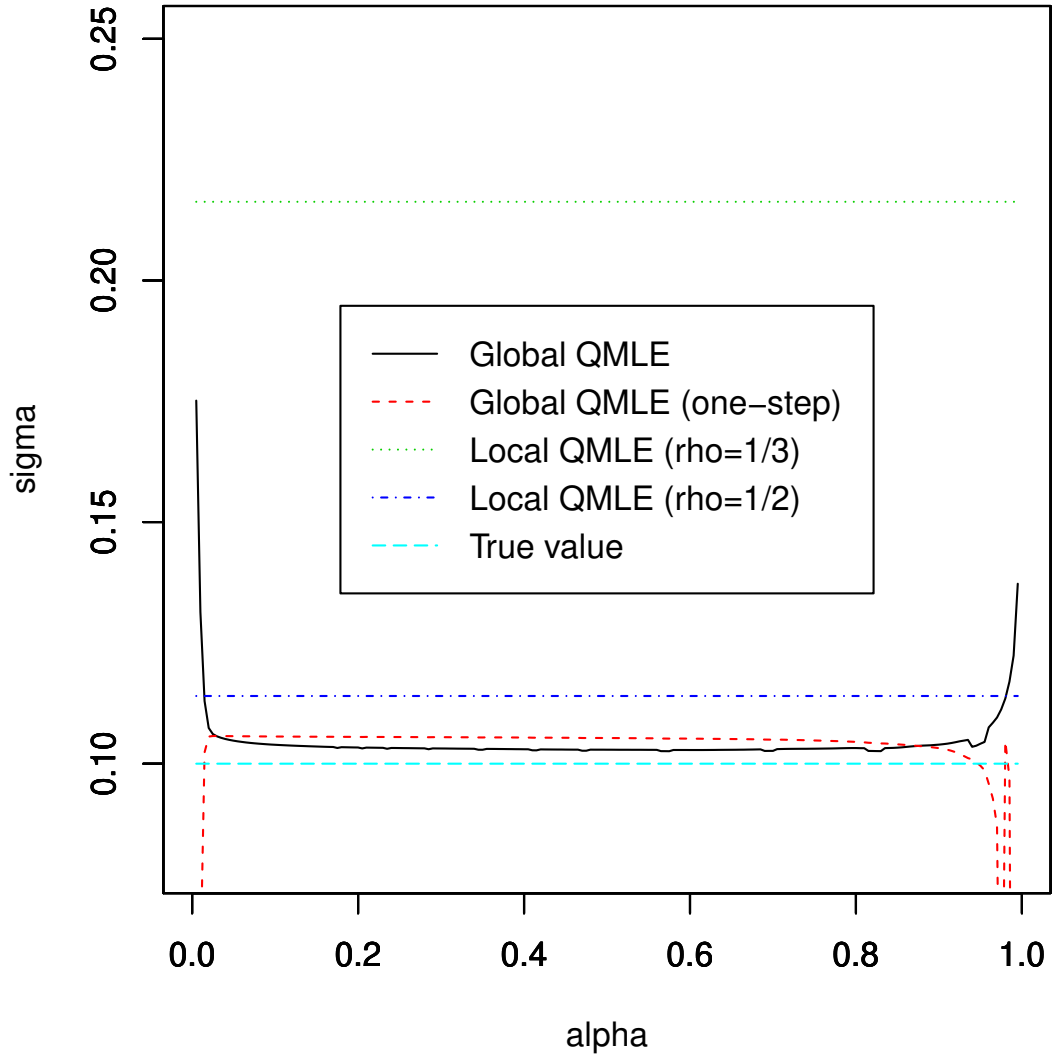
- **The parameter** η**,** ε**, and** λ **are nuisance parameters,**
- σ **is estimated from the discretely observed data** $\{X_{t_j^n};\ j = 1, 2, ..., n\}$**.**

Numerical studies

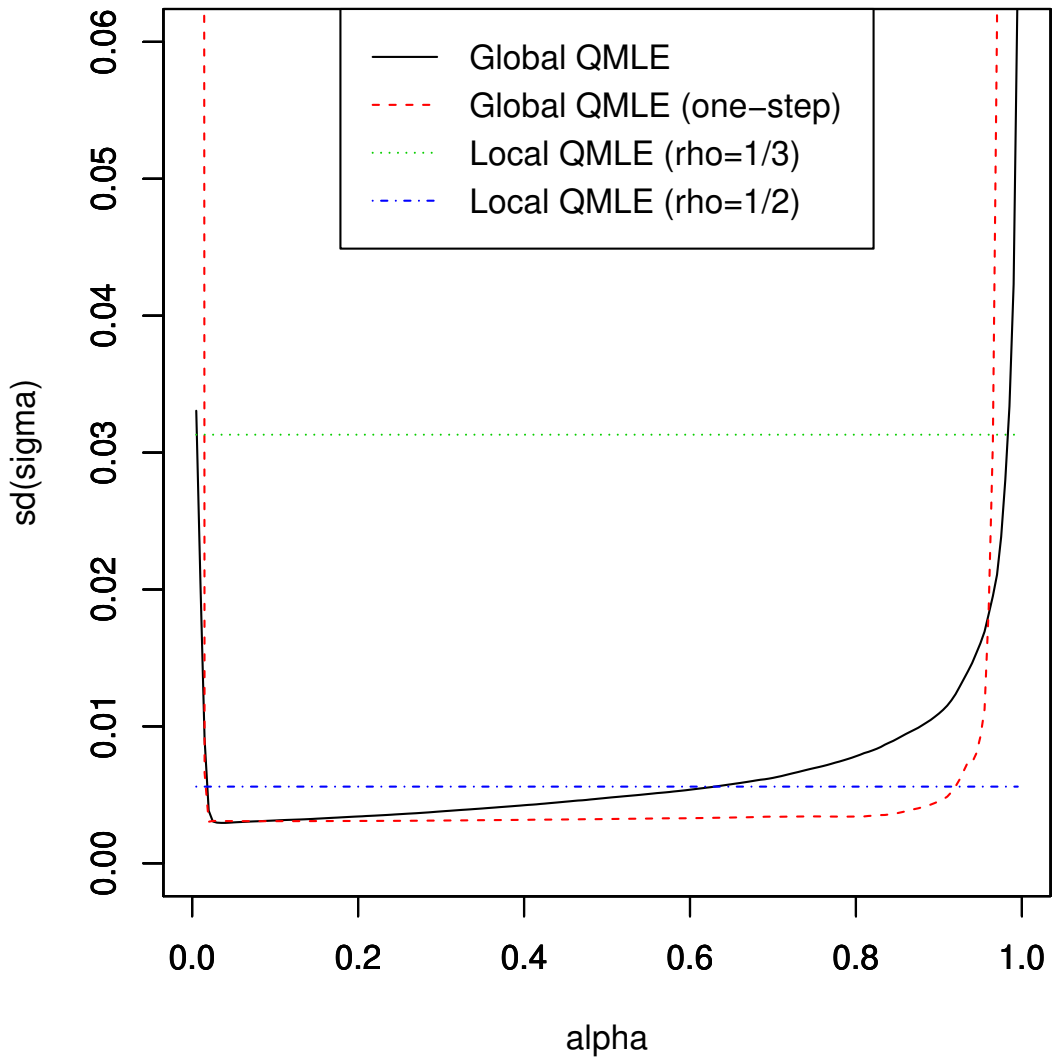
Compared four QMLEs

- **Global QMLE**: the QMLE by the global filter for given values of α
- **Global QMLE (one-step)**: the one-step estimator using a Global QMLE as the initial estimator
- **Local QMLE ($\rho = 1/3$)**: the threshold method with the bound $h^{1/3}$ $(|\Delta_j X| \gtrsim h^{1/3})$
- **Local QMLE ($\rho = 1/2$)**: the threshold method with the bound $h^{1/2}$ $(|\Delta_j X| \gtrsim h^{1/2})$ however theoretically banned
- simulation by YUIMA

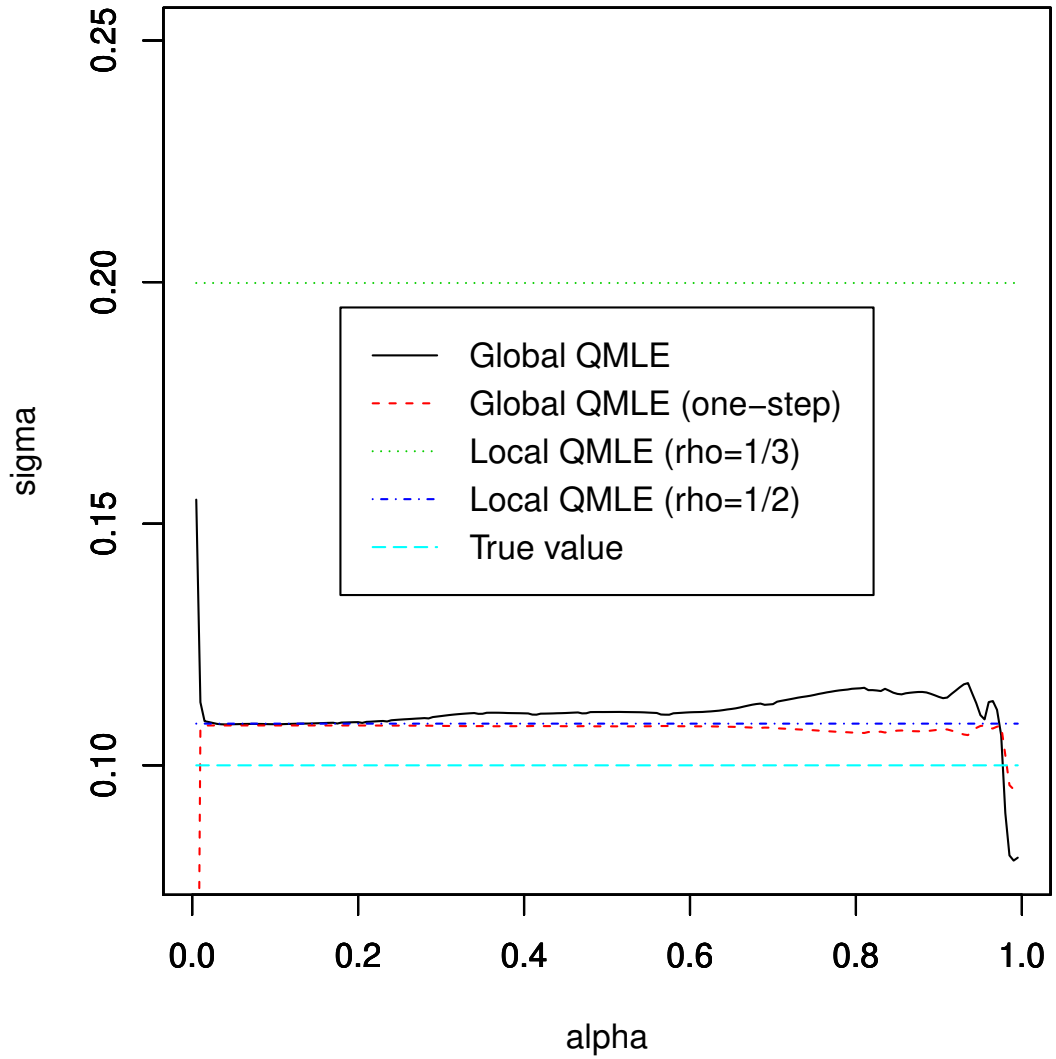
Numerical studies: mean (100 outcomes)



Numerical studies: standard deviation (100 outcomes)



Numerical studies: estimated value, given a path



Efficient one-step estimators

Efficient one-step estimators

Recall $\mathbb{H}_n(\theta)$ in (42):

$$\mathbb{H}_n(\theta) = -\frac{1}{2}\sum_{k=1}^{\mathsf{k}}\sum_{j\in\mathcal{K}_n}\Big\{(q_n)^{-1}h^{-1}S(X_{t_{j-1}},\theta)^{-1}\big[(\Delta_j Y)^{\otimes 2}\big]K_{n,j} + (p_n)^{-1}\log\det S(X_{t_{j-1}},\theta)\Big\}.$$

- $s_n = n - B\lfloor n^{\delta_1}\rfloor$ with constants $B > 0$ and $\delta_1 \in (0, 1/2)$.
- $\alpha_n = 1 - s_n/n$ and $\alpha_n = (\alpha_n^{(1)}, ..., \alpha_n^{(\mathsf{k})})$.
- Then $\alpha_n \sim Bn^{\delta_1 - 1}$.
- $\mathcal{K}_n = \big\{j \in \{1, ..., n\}; V_j < V_{(s_n)}\big\}$, where

$$V_j = |(\mathfrak{S}_{n,j-1})^{-1/2}\Delta_j Y|$$

with some positive definite random matrix $\mathfrak{S}_{n,j-1}$, and $V_{(j)}$ is the j-th order statistic of $V_1, ..., V_n$.

Efficient one-step estimators

- It was found that the α-QMLE $\hat{\theta}_n^{M,\alpha}$ and the (α,β)-QBE $\hat{\theta}_n^{B,\alpha,\beta}$ based on a fixed α-threshold are consistent.
- However they have pros and cons.
  - They are expected to remove jumps completely
  - but they are conservative and the rate of convergence is not optimal.
- We try to recover efficiency by combining the fixed α-method with the aggressive random field $\mathbb{H}_n$ of (42).

Efficient one-step estimators

- **Suppose that $\kappa \in \mathbb{N}$ satisfies $\kappa > 1 + (2\gamma_0)^{-1}$.**
- **Assume $[F1']_\kappa$, $[F2]$, $[F2']$ and $[F3]$.**
- **$\hat{\theta}_n^{M,\alpha}$ attains $n^{-\beta}$-consistency for any $\beta \in \left(2^{-1}(\kappa-1)^{-1}, \gamma_0\right)$, i.e.**

$$\beta(\kappa - 1) > 1/2.$$

Efficient one-step estimators

By some differential calculus, a sequence of $\mathbb{R}^{\mathbf{p}}$-valued random functions

$$\boldsymbol{A}_i(\boldsymbol{\theta}_0) \qquad (i = 1, ..., \boldsymbol{\kappa} - 2)$$

are defined on $\{\boldsymbol{\theta}_0 \in \boldsymbol{\Theta};\ \partial_{\boldsymbol{\theta}}^2 \mathbb{H}_n(\boldsymbol{\theta}_0)$ is invertible$\}$. For example, when $\boldsymbol{\kappa} = 4$,

$$\boldsymbol{A}_1(\boldsymbol{\theta}_0) = -\big(\partial_{\boldsymbol{\theta}}^2 \mathbb{H}_n(\boldsymbol{\theta}_0)\big)^{-1}[\partial_{\boldsymbol{\theta}} \mathbb{H}_n(\boldsymbol{\theta}_0)],$$

$$\boldsymbol{A}_2(\boldsymbol{\theta}_0) = -\frac{1}{2}\big(\partial_{\boldsymbol{\theta}}^2 \mathbb{H}_n(\boldsymbol{\theta}_0)\big)^{-1}\big[\partial_{\boldsymbol{\theta}}^3 \mathbb{H}_n(\boldsymbol{\theta}_0)[\boldsymbol{A}_1(\boldsymbol{\theta}_0)^{\otimes \mathbf{2}}]\big].$$

Efficient one-step estimators: with conservative initial QMLE

Let

$$\mathfrak{M}_n = \Big\{ \hat{\boldsymbol{\theta}}_n^{M,\boldsymbol{\alpha}} \in \Theta,\ \det \partial^2_{\boldsymbol{\theta}} \mathbb{H}_n(\hat{\boldsymbol{\theta}}_n^{M,\boldsymbol{\alpha}}) \neq 0,\ \hat{\boldsymbol{\theta}}_n^{M,\boldsymbol{\alpha}} + \sum_{i=1}^{\kappa-2} \boldsymbol{A}_i(\hat{\boldsymbol{\theta}}_n^{M,\boldsymbol{\alpha}}) \in \Theta \Big\}.$$

Define $\check{\boldsymbol{\theta}}_n^{M,\boldsymbol{\alpha}}$ by

$$\check{\boldsymbol{\theta}}_n^{M,\boldsymbol{\alpha}} = \begin{cases} \hat{\boldsymbol{\theta}}_n^{M,\boldsymbol{\alpha}} + \sum_{i=1}^{\kappa-2} \boldsymbol{A}_i(\hat{\boldsymbol{\theta}}_n^{M,\boldsymbol{\alpha}}) & \text{on } \mathfrak{M}_n \\ \boldsymbol{\theta}_* & \text{on } \mathfrak{M}_n^{\boldsymbol{c}} \end{cases}$$

where $\boldsymbol{\theta}_*$ is an arbitrary value in Θ.

Efficient one-step estimators: with conservative initial QMLE

We obtain

$$\left\|\hat{\boldsymbol{\theta}}_{\boldsymbol{n}}^{\boldsymbol{M},\boldsymbol{\alpha_n}} - \check{\boldsymbol{\theta}}_{\boldsymbol{n}}^{\boldsymbol{M},\boldsymbol{\alpha}}\right\|_{\boldsymbol{p}} = \boldsymbol{O}(\boldsymbol{n}^{-\boldsymbol{\beta}(\boldsymbol{\kappa}-1)})$$

(recall $\boldsymbol{\beta}(\boldsymbol{\kappa}-1) > 1/2$) and in particular this implies

$$\left\|\check{\boldsymbol{\theta}}_{\boldsymbol{n}}^{\boldsymbol{M},\boldsymbol{\alpha}} - \boldsymbol{\theta}^*\right\|_{\boldsymbol{p}} = \boldsymbol{O}(\boldsymbol{n}^{-1/2})$$

for every $\boldsymbol{p} > 1$.

Efficient one-step estimators: with conservative initial QBE

Let

$$\mathfrak{B}_n = \Big\{ \hat{\theta}_n^{B,\alpha,\beta} \in \Theta,\ \det \partial_\theta^2 \mathbb{H}_n(\hat{\theta}_n^{B,\alpha,\beta}) \neq 0,$$

$$\hat{\theta}_n^{B,\alpha,\beta} + \sum_{i=1}^{\kappa-2} A_i(\hat{\theta}_n^{B,\alpha,\beta}) \in \Theta \Big\}.$$

Define $\check{\theta}_n^{B,\alpha,\beta}$ by

$$\check{\theta}_n^{B,\alpha,\beta} = \begin{cases} \hat{\theta}_n^{B,\alpha,\beta} + \sum_{i=1}^{\kappa-2} A_i(\hat{\theta}_n^{B,\alpha,\beta}) & \text{on } \mathfrak{B}_n \\ \theta_* & \text{on } \mathfrak{B}_n^c. \end{cases}$$

Then we obtain $\left\| \hat{\theta}_n^{M,\alpha_n} - \check{\theta}_n^{B,\alpha,\beta} \right\|_p = O(n^{-\beta(\kappa-1)})$ and

$$\left\| \check{\theta}_n^{B,\alpha,\beta} - \theta^* \right\|_p = O(n^{-1/2})$$

for every $p > 1$.

Efficient one-step estimators

Write $\check{u}_n^{\mathsf{A}} = \sqrt{n}\big(\check{\theta}_n^{\mathsf{A}} - \theta^*\big)$ for A ="M, α" and "B, α, β". Thus, we have obtained the following result from Theorem 9 for $\hat{\theta}_n^{M,\alpha_n}$.

Theorem 13. Suppose that $\beta < \gamma_0$ and that an integer κ satisfies $\kappa > 1 + (2\gamma_0)^{-1}$. Suppose that $[F1']_\kappa$, $[F2]$, $[F2']$ and $[F3]$ are fulfilled. Then

$$E\big[f\big(\check{u}_n^{\mathsf{A}}\big)\Phi\big] \to \mathbb{E}\big[f\big(\Gamma^{-1/2}\zeta\big)\Phi\big]$$

as $n \to \infty$ for A ="M, α" and "B, α, β", any continuous function f of at most polynomial growth, and any $\mathcal{F}$-measurable random variable $\Phi \in \cup_{p>1} L^p$.