Fast and asymptotically efficient inference for large and high-frequency data

Alexandre Brouste, Le Mans Université

ICIAM 2023, Tokyo, August 25^{th}

Let Θ be an open set of $\mathbb{R}^d$. For $n \geq 1$, let

$$\mathfrak{E}^n = (\Omega^n, \mathcal{F}^n, \{P^n_\vartheta, \vartheta \in \Theta\})$$

be a sequence of statistical experiment (generated by the sample $X^{(n)} = (X_1, X_2, \ldots, X_n)$).

For $\vartheta \in \Theta$, we represent

$$\widetilde{\vartheta} = \vartheta + \varphi_n(\vartheta)u, \quad u \in \mathcal{A}^n(\vartheta)$$

where

$$\mathcal{A}^n(\vartheta) = \left\{ u \in \mathbb{R}^d ; \vartheta + \varphi_n(\vartheta)u \in \Theta \right\}. \quad (1)$$

Here $(\varphi_n(\vartheta), n \geq 1)$ is a sequence of $d \times d$ nondegenerate matrix with decreasing norm, $\|\varphi_n(\vartheta)\| \to 0$ as $n \to 0$.

For $n \geq 1$, let

$$\mathcal{E}^n(\vartheta) = \left(\Omega^n, \mathcal{F}^n, \left\{P^n_{u,\vartheta}, u \in \mathcal{A}^n(\vartheta)\right\}\right)$$

be the corresponding localized statistical experiment where $P^n_{u,\vartheta} = P^n_{\vartheta+\varphi_n(\vartheta)u}$.

If we can give a reasonable sense of the limit of the sequence of localized statistical experiment $(\mathcal{E}^n(\vartheta_0), n \geq 1)$ to a "simple" canonical experiment (for which the optimal decision can be defined), then

– Is this limiting optimal decision is an optimal decision in the localized statistical experiments ?

– Can we build a "global" optimal decision in the initial corresponding statistical experiment ?

The sequence of statistical experiment satisfies the LAMN property if

$$\log \frac{dP_{n,\vartheta+\varphi_n(\vartheta)u}}{dP_{n,\vartheta}} = u^T S_n(\vartheta) - \frac{1}{2} u^T J_n(\vartheta) u + o_{P_{n,\vartheta}}(1), \quad n \to \infty \tag{2}$$

and

$$\mathcal{L}\left((S_n(\vartheta), J_n(\vartheta)) \,|\, P_{n,\vartheta}\right) \Longrightarrow \mathcal{L}\left((S(\vartheta), J(\vartheta))\right) \quad \text{weakly,} \tag{3}$$

where $S(\vartheta)$ is conditionally multivariate Gaussian.

The sequence $Z_n(\vartheta) = J_n(\vartheta)^{-1} S_n(\vartheta)$ is called the central sequence. We denote $Z(\vartheta) = J(\vartheta)^{-1} S(\vartheta)$.

For an arbitrary choice of a constant $0 < c < \infty$,

$$\liminf_{n\to\infty} \sup_{\|\varphi_n(\vartheta)^{-1}(\widetilde{\vartheta}-\vartheta)\|\leq c} \mathrm{E}_{\widetilde{\vartheta}}\left(\ell\left(\varphi_n(\vartheta)^{-1}(T_n-\widetilde{\vartheta})\right)\right) \geq \mathrm{E}_0(\ell(Z))$$

for a large class of loss function ℓ.

Moreover, a sequence $(T_n)_n$ satisfying the coupling property

$$\varphi_n(\vartheta)^{-1}(T_n-\vartheta) = Z_n(\vartheta) + o_{P_{n,\vartheta}}(1), \quad n\to\infty, \tag{4}$$

attain the local asymptotic minimax bound at ϑ.

The project EFFI (`www.effi-stats.fr`) aims at popularize the generic Le Cam one-step method to give asymptotically efficient and fast estimation procedures.

The one-step modification $(\overline{\vartheta}_n)_n$

$$\overline{\vartheta}_n = \vartheta_n^* + \varphi_n(\vartheta_n^*) J_n(\vartheta_n^*)^{-1} S_n(\vartheta_n^*)$$

satisfies the coupling property (4) and achieves asymptotical efficiency in a LAMN experiment.

In order to make it work (see (Höpfner, 2014)), one generally requires 4 conditions:

1 – Rate-efficient consistency: Let $(\vartheta_n^*)_n$ be a initial sequence of guess estimators for which

$$\left(\varphi_n(\vartheta)^{-1}(\vartheta_n^* - \vartheta)\right)_n$$

is tight in $\mathbb{R}^d$ for every $\vartheta \in \Theta$.

2 – Uniform continuity of J_n:

$$\sup_{|\varphi_n(\vartheta)^{-1}(\xi-\vartheta)|\leq c} |J_n(\xi) - J_n(\vartheta)| = o_{P_{n,\vartheta}(1)}. \quad (5)$$

3 – "Regularity" of the score

$$\sup_{|\varphi_n(\vartheta)^{-1}(\xi-\vartheta)|\le c} \left| S_n(\xi) - \left\{ S_n(\vartheta) - J_n(\vartheta)\left(\varphi_n(\vartheta)^{-1}(\xi-\vartheta)\right)\right\}\right| = o_{P_{n,\vartheta}(1)}. \tag{6}$$

4 – Local scale

$$\sup_{|\varphi_n(\vartheta)^{-1}(\xi-\vartheta)|\le c} \left|\varphi_n^{-1}(\vartheta)\varphi_n(\xi) - I_p\right| \longrightarrow 0. \tag{7}$$

We sometimes use for different applications the following improved one-step procedure. Let

$$\overline{\vartheta}_n = \vartheta_n^* + \varphi_n(\vartheta_n^*) J(\vartheta_n^*)^{-1} S_n(\vartheta_n^*).$$

We impose a Lipshitz continuity on the observed information matrix and slower consistency of the initial guess estimator, namely $\left(\eta_n^{-1}(\vartheta_n^* - \vartheta)\right)_n$ is tight in $\mathbb{R}^d$ for every $\vartheta \in \Theta$ with

$$\|\eta_n^T \varphi_n(\vartheta)^{-1} \eta_n\| \to 0.$$

These procedures have been applied (non-exhaustive list) to:

- i.i.d. setting with the OneStep R package (B., Dutang, Noutsa Mienedou, 21), OS-SGD (Bensoussan, B., Esstafa, preprint), stable vectors (B. and Masuda, 18);
- Generalized Linear Models: general (Lumley, 19), categorical explanatory variables (B. et al., preprint);
- Markov processes (Kutoyants, Motrunich, 16), Diffusion processes (Kamatani, Uchida, 15) ;
- Counting processes: Inhomogeneous Poisson Process (Dabye, Gounoung, Kutoyants, 18), Hawkes processes (B., Farinetto, 23);
- Long memory processes : fBm (B., Soltane, Votsi, 20), FARIMA (Ben Hariz, B., Esstafa, Soltane, 23), ;
- Noisy observations: Kalman filter (Kutoyants, 23);
- and surely others : α-CIR with jumps (Bayraktar, Clément, preprint), ...

Time series

Yury Kutoyants, On Adaptive Kalman Filtration
Youssef Esstafa, Weak FARIMA models
Samir Ben Hariz, Fast inference for stationary time series

Stochastic differential equations

Laurent Denis, LAMN property for SDE driven by a stable Lévy process
Hiroki Masuda, Asymptotics for Student-Lévy regression
Elise Bayraktar, High-frequency estimation of pure jump alpha-CIR process
Ahmed Kebaier, Local asymptotic properties for the growth rate of a jump-type CIR process (on Zoom)

Fractional processes / Rough volatility models

Grégoire Szymanski, Statistical inference for rough volatility: minimax theory
Mikko Pakkanen, A GMM approach to estimate the roughness of stochastic volatility (on Zoom)
Tetsuya Takabatake, Asymptotically Efficient Estimation for Fractional Brownian Motion with Additive Noise
Mathieu Rosenbaum, Understanding how market impact shapes rough fractional volatility (on Zoom)