Parameter estimation of discretely observed interacting particle systems

Vytautė Pilipauskaitė  
(Aalborg University, Denmark)

joint work with  
Chiara Amorino, Akram Heidari, Mark Podolskij  
(University of Luxembourg)

2022–11–30
Plan of this talk

1. Model
2. Problem
3. Assumptions
4. Examples
5. Main results & proofs
6. Open problems
1. Model

Applications in

- mathematical biology (large societies of animals, neural networks)
- social sciences (multiagents’ opinion, consensus)
- finance (systemic risk)
- mean-field games (a large number of players)
An interacting particle system:

\[ (X^{1,N}_t, \ldots, X^{N,N}_t)_{t \in [0,T]}, \]

which satisfies

\[ \begin{cases} 
    dX^{i,N}_t = b(X^{i,N}_t, \mu^N_t, \theta_1)dt + a(X^{i,N}_t, \mu^N_t, \theta_2)dW^i_t, & i = 1, \ldots, N, \ t \in [0,T], \\
    \mathcal{L}(X^{1,N}_0, \ldots, X^{N,N}_0) = \mu_0 \times \cdots \times \mu_0,
\end{cases} \]

where

- \( W^i := (W^i_t)_{t \in [0,T]}, i = 1, \ldots, N \), are independent standard Brownian motions,
- \( X^{1,N}_0, \ldots, X^{N,N}_0 \) are independent of \( W^1, \ldots, W^N \),
- \( \mu^N_t := N^{-1} \sum_{i=1}^N \delta_{X^i_t,N} \),
- \( b : \mathbb{R} \times \mathcal{P}_2(\mathbb{R}) \times \Theta_1 \to \mathbb{R}, \ a : \mathbb{R} \times \mathcal{P}_2(\mathbb{R}) \times \Theta_2 \to \mathbb{R} \), where \( \mathcal{P}_2(\mathbb{R}) \) is the set of all probability measures on \( \mathbb{R} \) with finite 2nd moments endowed with the Wasserstein 2-metric \( W_2^* \), \( \Theta_j \subset \mathbb{R}^{p_j} \) is compact, convex for some \( p_j \in \mathbb{N}, j = 1, 2 \),
- \( \theta_j \in \Theta_j^\circ, j = 1, 2 \), are parameters of interest.

\[ W_2^*(\mu, \nu) := \inf_{\pi \in \Pi(\mu, \nu)} (\int_{\mathbb{R} \times \mathbb{R}} |x - y|^2 2\pi(dx, dy))^{1/2}, \]

where \( \Pi(\mu, \nu) \) denotes the set of all probability measures on \( \mathbb{R} \times \mathbb{R} \) with marginals \( \mu, \nu \).
The associated McKean-Vlasov SDE (introduced by McKean 1966):

\[
\begin{aligned}
d\bar{X}_t &= b(\bar{X}_t, \bar{\mu}_t, \theta_1)dt + a(\bar{X}_t, \bar{\mu}_t, \theta_2)dW_t, \quad t \in [0, T], \\
\bar{\mu}_0 &= \mu_0,
\end{aligned}
\]  

(1)

where

- \( W \) is a standard Brownian motion,
- \( \bar{X}_0 \) is independent of \( W \),
- \( \bar{\mu}_t := \mathcal{L}(\bar{X}_t) \).
The associated McKean-Vlasov SDE (introduced by McKean 1966):

\[
\begin{aligned}
    d\tilde{X}_t &= b(\tilde{X}_t, \bar{\mu}_t, \theta_1)dt + a(\tilde{X}_t, \bar{\mu}_t, \theta_2)dW_t, \quad t \in [0, T], \\
    \bar{\mu}_0 &= \mu_0,
\end{aligned}
\]

where

- \( W \) is a standard Brownian motion,
- \( \tilde{X}_0 \) is independent of \( W \),
- \( \bar{\mu}_t := \mathcal{L}(\tilde{X}_t) \).

The propagation of chaos property: if \( \tilde{X}^i \) satisfies (1) with \( W = W^i \), \( \tilde{X}_0 = X_0^{i,N} \), \( i = 1, \ldots, N \), under A1, A2 (to be stated later), then

\[
\mathbb{E}[|X_t^{i,N} - \tilde{X}_t^i|^2] \leq CN^{-1/2}, \quad i = 1, \ldots, N, t \in [0, T].
\]

This and other probabilistic properties: Sznitman 1984, 1991, Méléard 1996, Fernandez, Méléard 1997, Tanaka, Hitsuda 1981, Malrieu 2001, Bolley, Guillin, Villani 2007, Cattiaux, Guillin, Malrieu 2008, ...
2. Problem

Statistical inference:

▶ parametric: Kasonga 1990, Löcherbach 2002, Bishwal 2011, Giesecke, Schwenkler, Sirignano 2020, Genon-Catalot, Larédo 2021a, 2021b, 2022+, Chen 2021, Sharrock, Kantas, Parpas, Pavliotis 2021+, Della Maestra, Hoffmann 2022+

▶ nonparametric: Della Maestra, Hoffmann 2022, Belomestny, Pilipauskaitė, Podolskij 2022+, Belomestny, Podolskij, Zhou 2022+
2. Problem

Statistical inference:

- parametric: Kasonga 1990, Löcherbach 2002, Bishwal 2011, Giesecke, Schwenkler, Sirignano 2020, Genon-Catalot, Larédo 2021a, 2021b, 2022+, Chen 2021, Sharrock, Kantas, Parpas, Pavliotis 2021+, Della Maestra, Hoffmann 2022+

- nonparametric: Della Maestra, Hoffmann 2022, Belomestny, Pilipauskaitė, Podolskij 2022+, Belomestny, Podolskij, Zhou 2022+

Problem of this talk: estimation of the parameter \( \theta := (\theta_1, \theta_2) \) from observations

\[
X_{j\Delta_n}^{i,N}, \; i = 1, \ldots, N, \; j = 1, \ldots, n,
\]

with \( \Delta_n := T/n \), where \( T \) is fixed and \( N, n \to \infty \).
We define an estimator of $\theta = \theta_0$ as

$$\hat{\theta}_n^N \in \arg \min_{\theta \in \Theta} S_n^N(\theta),$$

where

$$S_n^N(\theta) := \sum_{0<i \leq N} \sum_{0 \leq j < n} \left\{ \frac{(X_{(j+1)\Delta_n}^{i,N} - X_{j\Delta_n}^{i,N} - b(X_{j\Delta_n}^{i,N}, \mu_{j\Delta_n}^N, \theta_1) \Delta_n)^2}{a^2(X_{j\Delta_n}^{i,N}, \mu_{j\Delta_n}^N, \theta_2) \Delta_n} + \log a^2(X_{j\Delta_n}^{i,N}, \mu_{j\Delta_n}^N, \theta_2) \right\}$$

uses a Gaussian quasi-likelihood which originates from the Euler scheme:

$$X_{(j+1)\Delta_n}^{i,N} \approx X_{j\Delta_n}^{i,N} + b(X_{j\Delta_n}^{i,N}, \mu_{j\Delta_n}^N, \theta_1) \Delta_n$$

$$+ a(X_{j\Delta_n}^{i,N}, \mu_{j\Delta_n}^N, \theta_2)(W_{(j+1)\Delta_n}^i - W_{j\Delta_n}^i).$$

The Euler scheme contrast in case of the classical SDE: Florens-Zmirou 1989, Yoshida 1992, Kessler 1997
3. Assumptions

A1 $\mu_0$ has finite $k$-th moment for all $k \in \mathbb{N}$.

A2 For all $\theta$ there exists a $C > 0$ such that all $(x, \mu), (y, \nu)$,

$$|b(x, \mu, \theta_1) - b(y, \nu, \theta_1)| + |a(x, \mu, \theta_2) - a(y, \nu, \theta_2)| \leq C(|x - y| + W_2(\mu, \nu)).$$
3. Assumptions

A1 \( \mu_0 \) has finite \( k \)-th moment for all \( k \in \mathbb{N} \).

A2 For all \( \theta \) there exists a \( C > 0 \) such that all \( (x, \mu), (y, \nu), \)
\[
|b(x, \mu, \theta_1) - b(y, \nu, \theta_1)| + |a(x, \mu, \theta_2) - a(y, \nu, \theta_2)| \leq C(|x - y| + W_2(\mu, \nu)).
\]

A3 There exists a \( C > 0 \) such that \( a^2(x, \mu, \theta_2) > C \) for all \( (x, \mu, \theta_2) \).

A4 (i) \( b(x, \mu, \cdot) \in C^3(\Theta_1) \), \( a(x, \mu, \cdot) \in C^3(\Theta_2) \) for all \( (x, \mu) \) and their partial derivatives up to order 3 have polynomial growth in \( (x, \mu) \) uniformly in \( \theta \).
3. Assumptions

A1 $\mu_0$ has finite $k$-th moment for all $k \in \mathbb{N}$.

A2 For all $\theta$ there exists a $C > 0$ such that all $(x, \mu), (y, \nu)$,

$$|b(x, \mu, \theta_1) - b(y, \nu, \theta_1)| + |a(x, \mu, \theta_2) - a(y, \nu, \theta_2)| \leq C(|x - y| + W_2(\mu, \nu)).$$

A3 There exists a $C > 0$ such that $a^2(x, \mu, \theta_2) > C$ for all $(x, \mu, \theta_2)$.

A4 (i) $b(x, \mu, \cdot) \in C^3(\Theta_1), \ a(x, \mu, \cdot) \in C^3(\Theta_2)$ for all $(x, \mu)$ and their partial derivatives up to order 3 have polynomial growth in $(x, \mu)$ uniformly in $\theta$.

(ii) For all $\theta$ there exist $C > 0, k, l = 0, 1, \ldots$ such that for all $r_1 + r_2 = 1, 2, h_1, h_2 = 1, \ldots, p_1, \tilde{h}_1, \tilde{h}_2 = 1, \ldots, p_2$ and $(x, \mu), (y, \nu)$,

$$|\partial_{\theta_1, h_1}^{r_1} \partial_{\theta_1, h_2}^{r_2} b(x, \mu, \theta_1) - \partial_{\theta_1, h_1}^{r_1} \partial_{\theta_1, h_2}^{r_2} b(y, \nu, \theta_1)|$$

$$+ |\partial_{\theta_2, \tilde{h}_1}^{r_1} \partial_{\theta_2, \tilde{h}_2}^{r_2} a(x, \mu, \theta_2) - \partial_{\theta_2, \tilde{h}_1}^{r_1} \partial_{\theta_2, \tilde{h}_2}^{r_2} a(y, \nu, \theta_2)|$$

$$\leq C(|x - y| + W_2(\mu, \nu))(1 + |x|^k + |y|^k + W_2^l(\mu, \delta_0) + W_2^l(\nu, \delta_0)).$$
A5 For all $\varepsilon > 0$,

$$\inf_{\theta \in \Theta: \|\theta - \theta_0\| \geq \varepsilon} I(\theta) > 0 \quad \text{and} \quad \inf_{\theta_2 \in \Theta_2: \|\theta_2 - \theta_0\| \geq \varepsilon} J(\theta_2) > J(\theta_0, 2),$$

where $I : \Theta \to \mathbb{R}$, $J : \Theta_2 \to \mathbb{R}$ are defined as

$$I(\theta) := \int_0^T \int_{\mathbb{R}} \frac{(b(x, \bar{\mu}_t, \theta_1) - b(x, \bar{\mu}_t, \theta_0, 1))^2}{a^2(x, \bar{\mu}_t, \theta)} \bar{\mu}_t(dx)dt,$$

$$J(\theta_2) := \int_0^T \int_{\mathbb{R}} \left( \frac{a^2(x, \bar{\mu}_t, \theta_0, 2)}{a^2(x, \bar{\mu}_t, \theta_2)} + \log a^2(x, \bar{\mu}_t, \theta_2) \right) \bar{\mu}_t(dx)dt.$$
A6 The main-diagonal block elements of $\Sigma(\theta_0): = \text{diag}(\Sigma^{(1)}(\theta_0), \Sigma^{(2)}(\theta_0))$ are $p_j \times p_j$ invertible $\Sigma^{(j)}(\theta_0) = (\Sigma_{kl}^{(j)}(\theta_0))$, $j = 1, 2$, where

$$
\Sigma_{kl}^{(1)}(\theta_0) := 2 \int_0^T \int_{\mathbb{R}} \frac{\partial_{\theta_1,k} b(x, \bar{\mu}_t, \theta_0,1) \partial_{\theta_1,l} b(x, \bar{\mu}_t, \theta_0,1)}{a^2(x, \bar{\mu}_t, \theta_0,2)} \bar{\mu}_t(dx)dt,
$$

$$
\Sigma_{kl}^{(2)}(\theta_0) := \int_0^T \int_{\mathbb{R}} \frac{\partial_{\theta_2,k} a^2(x, \bar{\mu}_t, \theta_0,2) \partial_{\theta_2,l} a^2(x, \bar{\mu}_t, \theta_0,2)}{a^4(x, \bar{\mu}_t, \theta_0,2)} \bar{\mu}_t(dx)dt.
$$

A7 There are some $\tilde{a}, K \in C^2(\mathbb{R}^2)$ whose partial derivatives of order 1, 2 have polynomial growth such that for all $(x, \mu)$,

$$
a(x, \mu, \theta_0,2): = \tilde{a} \left( x, \int_{\mathbb{R}} K(x, y) \mu(dy) \right).
$$
4. Examples

(i) The Kuramoto model: $N$ oscillators defined by $N$ angles $X_{t}^{i,N}$ (defined modulo $2\pi$) evolving over $[0,T]$ according to

$$dX_{t}^{i,N} = -\frac{\theta_1}{N} \sum_{j=1}^{N} \sin(X_{t}^{i,N} - X_{t}^{j,N}) dt + \theta_2 dW_{t}^{i}.$$

Here

$$b(x, \mu, \theta_1) = -\theta_1 \int_{\mathbb{R}} \sin(x - y) \mu(\text{d}y), \quad a(x, \mu, \theta_2) = \theta_2.$$
4. Examples

(i) The Kuramoto model: $N$ oscillators defined by $N$ angles $X_{t,N}^{i,N}$ (defined modulo $2\pi$) evolving over $[0, T]$ according to

$$dX_{t,N}^{i,N} = -\frac{\theta_1}{N} \sum_{j=1}^{N} \sin(X_{t,N}^{i,N} - X_{t,N}^{j,N})dt + \theta_2 dW_{t}^{i}.$$ 

Here

$$b(x, \mu, \theta_1) = -\theta_1 \int_{\mathbb{R}} \sin(x - y)\mu(dy), \quad a(x, \mu, \theta_2) = \theta_2.$$ 

(ii) A model for opinion dynamics:

$$dX_{t,N}^{i,N} = -\frac{1}{N} \sum_{j=1}^{N} \varphi(|X_{t,N}^{i,N} - X_{t,N}^{j,N}|, \theta_1)(X_{t,N}^{i,N} - X_{t,N}^{j,N})dt + \theta_2 dW_{t}^{i},$$

where the influence function $\varphi(x, \theta_1)$, which acts on the “difference of opinions” between agents, is an infinitely differentiable approximation of $\theta_1, 1_{[0, \theta_1, 2]}(x), x \in \mathbb{R}.$
(iii) The interacting particle system

\[
dX_{t}^{i,N} = \left( \theta_{1,1} + \frac{\theta_{1,2}}{N} \sum_{j=1}^{N} X_{t}^{j,N} - \theta_{1,3}X_{t}^{i,N} \right) dt + \theta_{2} \sqrt{1 + (X_{t}^{i,N})^2} dW_{t}^{i}
\]

for \( \theta_{1,2} = 0 \) reduces to \( N \) independent samples of a special case of the Pearson diffusion, which has applications in finance.

(iv) The mean field limit of

\[
dX_{t}^{i,N} = \left( \theta_{1,1} + \frac{\theta_{1,2}}{N} \sum_{j=1}^{N} X_{t}^{j,N} - \theta_{1,3}X_{t}^{i,N} \right) dt + \left( \theta_{2,1} + \theta_{2,2} \sqrt{\frac{1}{N} \sum_{j=1}^{N} (X_{t}^{j,N})^2} \right) dW_{t}^{i}
\]

is a time-inhomogeneous Ornstein-Uhlenbeck process. See Kasonga 1990 for \( \theta_{1,1} = \theta_{2,2} = 0 \).
5. Main results & proofs

Theorem (Consistency)
Assume A1-A5 without A4(ii). Then
\[ \hat{\theta}_n^N \xrightarrow{P} \theta_0 \quad \text{as } N, n \to \infty. \]

Theorem (Asymptotic normality)
Assume A1-A7. If \( N \Delta_n \to 0 \) then
\[
\left( \sqrt{N}(\hat{\theta}_{n,1}^N - \theta_{0,1}), \sqrt{N/\Delta_n}(\hat{\theta}_{n,2}^N - \theta_{0,2}) \right) \xrightarrow{L} \mathcal{N}(0, 2(\Sigma(\theta_0))^{-1})
\quad \text{as } N, n \to \infty.
\]
Technical results: Assume A1-A2.

Then for all $k \in \mathbb{N}, N \in \mathbb{N}, i = 1, \ldots, N, s, t \in [0, T]$ it holds

$$\sup_{t \in [0,T]} \mathbb{E}[|X_{t}^{i,N}|^{k}] \leq C, \quad \mathbb{E}[|X_{t}^{i,N} - X_{s}^{i,N}|^{k}] \leq C|t - s|^{k/2}.$$
Technical results: Assume A1-A2.

Then for all $k \in \mathbb{N}$, $N \in \mathbb{N}$, $i = 1, \ldots, N$, $s, t \in [0, T]$ it holds

$$
\sup_{t \in [0, T]} \mathbb{E}[|X_{t}^{i, N}|^k] \leq C, \quad \mathbb{E}[|X_{t}^{i, N} - X_{s}^{i, N}|^k] \leq C|t - s|^{k/2}.
$$

Moreover, assume that $f : \mathbb{R} \times \mathcal{P}_l(\mathbb{R}) \rightarrow \mathbb{R}$ satisfies for some $C > 0$, $k, l = 0, 1, \ldots$ and all $(x, \mu), (y, \nu)$,

$$
|f(x, \mu) - f(y, \nu)| 
\leq C(|x - y| + W_{2}(\mu, \nu))(1 + |x|^k + |y|^k + W_{l}^{l}(\mu, \delta_0) + W_{l}^{l}(\nu, \delta_0))
$$

and $(x, t) \mapsto f(x, \bar{\mu}_t)$ is in $L^1(\mathbb{R} \times [0, T], \bar{\mu}_t(dx)dt)$. Then as $N, n \rightarrow \infty$,

$$
\frac{\Delta_n}{N} \sum_{0 < i \leq N} \sum_{0 \leq j < n} f(X_{j \Delta_n}^{i, N}, \mu_{j \Delta_n}^N) \overset{\mathbb{P}}{\rightarrow} \int_0^T \int_{\mathbb{R}} f(x, \bar{\mu}_t)\bar{\mu}_t(dx)dt.
$$
Proof of consistency:

Assume A1-A4(i). Then

$$\sup_{\theta \in \Theta} \left| \frac{\Delta_n}{N} S_n^N (\theta) - J(\theta_2) \right| = o_P(1),$$

$$\sup_{\theta \in \Theta} \left| \frac{1}{N} (S_n^N (\theta) - S_n^N (\theta_0,1,\theta_2)) - I(\theta) \right| = o_P(1).$$

(2)
Proof of consistency:

▶ Assume A1-A4(i). Then

\[
\sup_{\theta \in \Theta} \left| \frac{\Delta_n}{N} S_n^N (\theta) - J(\theta_2) \right| = o_P(1),
\]

\[
\sup_{\theta \in \Theta} \left| \frac{1}{N} (S_n^N (\theta) - S_n^N (\theta_0, 1, \theta_2)) - I(\theta) \right| = o_P(1). \tag{2}
\]

▶ A5 implies that for every \( \varepsilon > 0 \) there exists \( \eta > 0 \) such that

\[
\{ \| \hat{\theta}_{n,1} - \theta_0, 1 \| \geq \varepsilon \} \subseteq \{ I(\hat{\theta}_n^N) > \eta \},
\]

where the definition of \( \hat{\theta}_n^N \) and (2) imply

\[
I(\hat{\theta}_n^N) \leq I(\hat{\theta}_n^N) - \frac{1}{N} (S_n^N (\hat{\theta}_n^N) - S_n^N (\theta_0, 1, \hat{\theta}_n^N)) = o_P(1).
\]

The proof is similar in the case of \( \hat{\theta}_n^N, 2 \).
Proof of asymptotic normality: By mean value theorem and definition of $\hat{\theta}_n^N$, 

$$-\nabla_\theta S_n^N(\theta_0) = (\hat{\theta}_n^N - \theta_0) \int_0^1 \nabla^2_\theta S_n^N(\theta_0 + s(\hat{\theta}_n^N - \theta_0)) ds.$$ 

Furthermore, we introduce 

$$M_n^N := \text{diag} \left( N^{-\frac{1}{2}} 1_{p_1}, \left( \frac{\Delta_n}{N} \right)^{\frac{1}{2}} 1_{p_2} \right),$$ 

$$\Sigma_n^N(\theta) := M_n^N \nabla^2_\theta S_n^N(\theta) M_n^N.$$
Steps:

If $N\Delta_n \to 0$ then

$$\nabla_\theta S_n^N (\theta_0) M_n^N = \sum_{0<j\leq n} \xi_{n,j} \xrightarrow{L} \mathcal{N}(0, 2\Sigma(\theta_0))$$

follows from

$$\sum_{0<j\leq n} \mathbb{E}_{(j-1)\Delta_n} [\xi_{n,j}] \xrightarrow{P} 0$$

and a central limit theorem for martingale difference triangular arrays (Theorem 3.2 in Hall, Heyde 1980):

$$\sum_{0<j\leq n} \mathbb{E}_{(j-1)\Delta_n} [\xi_{n,j} (\xi_{n,j})^\top] \xrightarrow{P} 2\Sigma(\theta_0), \quad \sum_{0<j\leq n} \mathbb{E}_{(j-1)\Delta_n} [|\xi_{n,j,h}|^4] \xrightarrow{P} 0,$$

where $h = 1, \ldots, p_1 + p_2$ and $\mathbb{E}_t [\cdot]$ denotes the conditional expectation with respect to $\sigma(\{ X_0^i, N, (W_s^i)_{s \in [0,t]}, i = 1, \ldots, N\})$. 
Steps:

▶ If $N\Delta_n \to 0$ then

$$\nabla_{\theta} S_n^N (\theta_0) M_n^N = \sum_{0<j \leq n} \xi_{n,j} \overset{\mathcal{L}}{\to} \mathcal{N}(0, 2\Sigma(\theta_0))$$

follows from

$$\sum_{0<j \leq n} \mathbb{E}_{(j-1)\Delta_n} [\xi_{n,j}] \overset{\mathbb{P}}{\to} 0$$

and a central limit theorem for martingale difference triangular arrays (Theorem 3.2 in Hall, Heyde 1980):

$$\sum_{0<j \leq n} \mathbb{E}_{(j-1)\Delta_n} [\xi_{n,j} (\xi_{n,j})^\top] \overset{\mathbb{P}}{\to} 2\Sigma(\theta_0),$$

$$\sum_{0<j \leq n} \mathbb{E}_{(j-1)\Delta_n} [|\xi_{n,j,h}|^4] \overset{\mathbb{P}}{\to} 0,$$

where $h = 1, \ldots, p_1 + p_2$ and $\mathbb{E}_t [\cdot]$ denotes the conditional expectation with respect to $\sigma(\{X_{i,0}^N, (W_i^s)_{s \in [0,t]}, i = 1, \ldots, N\})$.

▶ $\Sigma_n^N (\theta_0) \overset{\mathbb{P}}{\to} \Sigma(\theta_0)$,

▶ $\sup_{s \in [0,1]} \|\Sigma_n^N (\theta_0 + s(\hat{\theta}^N_n - \theta_0)) - \Sigma_n^N (\theta_0)\| \overset{\mathbb{P}}{\to} 0$. 
6. Open problems

- Efficiency
- To improve $N \Delta_n \to 0$
- $X_{t}^{i,N} \in \mathbb{R}^{d}$
- Presence of jumps
- Non or semi-parameteric estimation
References

- Amorino, C., Heidari, A., Pilipauskaitė, V., Podolskij, M. (2022+). Parameter estimation of discretely observed interacting particle systems. Preprint. arXiv:2208.11965.

- Florens-Zmirou, D. (1989). Approximate discrete-time schemes for statistics of diffusion processes. Statistics: A Journal of Theoretical and Applied Statistics, 20(4), 547-557.

- Kessler, M. (1997). Estimation of an ergodic diffusion from discrete observations. Scandinavian Journal of Statistics, 24(2), 211-229.

Merci pour votre attention