Statistical inference for structured models

Part IV: Estimation with bias sampling and proxy experiments. Large population models. Further models

Marc Hoffmann, Université Paris-Dauphine PSL

Spring school SFB 1294, March 2019
Informal structure of the study

► **Statistical setting:** We have (i) data $Z^N$ and (ii) a parameter of interest $f$. Asymptotics are taken as $N \to \infty$.

► **Structure of the problem:**

\[
\mathcal{H}_N(Z^N) = 0 \quad \text{for some SDE } \mathcal{H}_N,
\]

\[
Z^N \to \xi \quad \text{limiting object},
\]

\[
\mathcal{H}(\xi, f) = 0 \quad \text{for some PDE } \mathcal{H}.
\]

► **Objective:** recover $f$ from the observation of $Z^N$ (or a proxy $Z^N$ of $Z^N$).
Today’s program

- **Bias sampling** for growth-fragmentation models
  - Age model: many-to-one formulas.
  - Size models steady-state approximation.
- Human population models and **nonlinear extensions**
- **Nonlinear models** and open questions
  - Models of interacting neurons
  - More nonlinear models in a mean-field limit
Bias sampling
    Age dependent model
    Size model: estimation at a large fixed time in a proxy model

Large population models

Nonlinear extensions, open questions
    Models of interacting neurons
    More non-linear models in a mean-field limit
Age dependent division rate $B(a)$

- The associated deterministic model is

$$\begin{cases} 
\partial_t g(t, a) + \partial_a g(t, a) + B(a)g(t, a) = 0 \\
g(0, a) = g_0(a), \ g(t, 0) = 2 \int_0^{\infty} B(a)g(t, a)da.
\end{cases}$$

- We are interested in recovering $a \mapsto B(a)$ from data $(Z_t)_{0 \leq t \leq T}$ or $Z_T$

$$Z_t = \sum_{i=1}^{N_t} \delta_{A_i(t)} \text{ with } g(t, \cdot) = \mathbb{E}[Z_t^N].$$

- Heuristically $Z_T \approx g(T, \cdot)$ when $T$ is large.

- $N = \mathbb{E}[\langle Z_T, 1 \rangle] \to \infty$ as $T \to \infty$. 
Observation scheme

- We observe \((Z_t)_{0 \leq t \leq T}\) or \(Z_T\).
- Tree representation:

\[
\mathcal{T}_T = \{ u \in \mathbb{T}, b_u \leq T \} = \mathcal{\hat{T}}_T \cup \partial \mathcal{T}_T,
\]

\[
\mathcal{\hat{T}}_T = \{ u \in \mathbb{T}, d_u \leq T \},
\]

\[
\partial \mathcal{T}_T = \{ u \in \mathbb{T}, b_u \leq T < d_u \}.
\]

- We have the correspondence

\[
\begin{cases}
(Z_t)_{0 \leq t \leq T} \leftrightarrow \{ \zeta_u^T = \min(d_u, T) - b_u, u \in \mathcal{T}_T \}, \\
Z_T \leftrightarrow \{ \zeta_u^T, u \in \partial \mathcal{T}_T \}.
\end{cases}
\]

- Additional difficulty: bias selection.
- Recovering strategy: many-to-one formulae.
Observation schemes $\hat{T}_T \cup \partial T_T$

Figure: A sample path of $Z_t(da)_{0 \leq t \leq T}$ with $B(a) = a^2$ and $T = 7$. 
Estimation of $B(a)$ from $\mathcal{T}_T$

- **Many-to-one formula**: For nice test functions $\varphi$:

$$
\mathbb{E}\left[ \sum_{u \in \mathcal{T}_T} \varphi(\zeta_u) \right] = \int_0^T e^{\lambda_B s} \mathbb{E}\left[ \varphi(\chi(s)) H_B(\chi(s)) \right] ds
$$

- $\chi(t)$: a tagged branch picked at random on the tree.
- We have $\mathbb{E}[|\mathcal{T}_T|] \sim \kappa_B e^{\lambda_B T}$ and thus

$$
N = N_B \approx e^{\lambda_B T} \text{ depends on } B \text{ itself!}
$$

- $\lambda_B$: Malthus parameter, related to $\chi$ and $H_B$.
- $H_B(a)$ explicit: $f_{H_B}(a) = 2e^{-\lambda_B a}f_B(a)$.
- We have all the ingredients needed for a law of large numbers.
Estimation of $B(a)$ from $\mathcal{T}_T$

- $f_B(a) = B(a) \exp\left(-\int_0^\infty B(s)ds\right)$.
- Law of large numbers

\[
\frac{1}{|\mathcal{T}_T|} \sum_{u \in \mathcal{T}_T} \varphi(\zeta_u) \xrightarrow{\mathbb{P}} \int_0^\infty \varphi(a) 2e^{\lambda_B a} f_B(a) da
\]

- Rate of convergence: $(e^{\lambda_B T})^{1/2} = \mathcal{N}^{1/2}$ in probability.
- Rate heavily parameter dependent.
- Proof: establish rates of convergence in the many-to-one formula for test functions on forks $\varphi(\zeta_u, \zeta_v)$ for $u, v \in \mathcal{T}_T$ + geometric ergodicity.
- We meet the same difficulties as for BMC models.
Estimation of $B(a)$ from $\mathcal{T}_T$

- We can find a fast converging preliminary estimator $\hat{\lambda}_T$ of $\lambda_B$.
- Set

$$
\hat{B}_h^T(a) = \frac{|\hat{\mathcal{T}}_T|^{-1} \sum_{u \in \hat{\mathcal{T}}_T} \frac{1}{2} e^{\hat{\lambda}_T \zeta_u} K_h(a - \zeta_u)}{1 - |\hat{\mathcal{T}}_T|^{-1} \sum_{u \in \hat{\mathcal{T}}_T} \frac{1}{2} e^{\hat{\lambda}_T \zeta_u} 1\{\zeta_u \leq a\}}
$$

- For $h = \hat{h}^T(\alpha) = \left( \exp(\hat{\lambda} T) \right)^{-1/(2\alpha+1)}$, we have the weak boundedness of

$$
\mathcal{N}^{\alpha/(2\alpha+1)} \left( \hat{B}_h^T(\alpha)(a) - B(a) \right)
$$

uniformly over $B \cap \mathcal{H}^\alpha$ for appropriate $B$.

- The rate is nearly minimax.

- Open problem: we do not have adaptation, for lack of concentration inequalities.
What if data are taken from $\partial T_T$ solely?

- By another many-to-one formula, we have for good test functions $\varphi$

  $$|\partial T_T|^{-1} \sum_{u \in \partial T_T} \varphi(\zeta_u) \xrightarrow{\text{P}} 2\lambda_B \int_0^\infty \varphi(a) e^{\lambda_B a} \frac{f_B(a)}{B(a)} da$$

  $$= 2\lambda_B \int_0^\infty \varphi(a) e^{\lambda_B a} e^{-\int_0^a B(s) ds} da.$$

- We still have a $N^{1/2}$-rate of convergence (in probability).
- We retrieve an ill-posed problem of order 1, leading to convergence rate

  \[ N^{\alpha/(2\alpha+3)} \]

  but not $N^{\alpha/(2\alpha+1)}$. 
The age dependent model, simulated data

Figure: Reconstruction of $B$ over $\mathcal{D} = [0.1, 4]$ with 95%-level confidence bands constructed over $M = 100$ Monte-Carlo trees. In bold red line: $x \sim B(x)$; in bold blue line: $f_{HB}$; in blue line: $f_B$. Left: $T = 15$. Right: $T = 23$. 
Size dependent division rate $B(x)$

- The associated deterministic model is

$$
\begin{align*}
\partial_t g(t, x) + \partial_x (\kappa(x) g(t, x)) + B(x) g(t, x) &= 4B(2x)g(t, 2x) \\
g(0, x) &= g_0(x), \; g(t, 0) = 0.
\end{align*}
$$

- We are interested in recovering $x \mapsto B(x)$ from terminal data

$$Z_T \longleftrightarrow \partial \mathcal{T}_T \; \text{solely.}$$

- $Z_t = \sum_{i=1}^{N_t} \delta_{X_i(t)}$ with $g(t, \cdot) = \mathbb{E}[Z_t^N]$.

- Heuristically $Z_T \approx g(T, \cdot)$ when $T$ is large.

- $N = \mathbb{E}[\langle Z_T, 1 \rangle] \to \infty$ as $T \to \infty$.

- This is too difficult!
Alternate strategy: “if the data don’t fit, change the data!”

- Represent the solution of the transport-fragmentation equation in a stationary regime.
- Obtain a reconstruction formula for $B(x)$ via this representation in terms of the steady-state or stationary density of the model.
- Postulate a proxy model where one observes exactly a drawn from the stationary density.
- Transfer standard nonparametric estimation techniques in this setting.
Solution by stable distribution

- Start with the transport-fragmentation equation \((\kappa(x) = \tau x)\)

\[
\partial_t g(t, x) + \partial_x (\tau x g(t, x)) + B(x)g(t, x) = 4B(2x)g(t, 2x)
\]

- Ansatz: \(g(t, x) = e^{\lambda t} N(x)\) \((\lambda = \lambda_B\): Malthus parameter).

\[
\partial_x (\tau x N(x)) + (\lambda + B(x)) N(x) = 4B(2x)N(2x).
\]

- Steady-state approximation: \(g(T, x) \approx e^{\lambda T} N(x)\) when \(T \to \infty\) with explicit (fast) rates of convergence.

- Interpretation: \(N(x)\) stationary size distribution of a cell in a stationary regime.
A proxy statistical model

- Yields a strategy for the nonparametric estimation of $B$:
  1. Extract from $Z_T$ a “sample” $X_1, \ldots, X_n$ of cell sizes.
  2. Postulate the approximation

$$\mathbb{P}(X_1 \in dx_1, \ldots, X_n \in dx_n) \approx \otimes_{i=1}^{n} N(x_i) dx_i.$$

If $n \to \infty$ but $n \ll N$, hope for a chaos propagation property.

3. Recover $B$ through the representation

$$L(N) = \mathcal{L}(BN),$$

or

$$B = \frac{\mathcal{L}^{-1}L(N)}{N}$$

with

$$L(\varphi)(x) = \partial_x(\tau x \varphi(x)) + \lambda \varphi(x),$$

$$\mathcal{L}(\varphi)(x) = 4\varphi(2x) - \varphi(x).$$

- The operator $L(\cdot)$ has ill-posedness degree of order 1. The operator $\mathcal{L}$ is “nicer”.
Growth-fragmentation: a word of conclusion

<table>
<thead>
<tr>
<th>data</th>
<th>Size model</th>
<th>Age model</th>
</tr>
</thead>
<tbody>
<tr>
<td>proxy model</td>
<td>$n^{-\alpha/(2\alpha+3)}$ + adaptation</td>
<td>irrelevant</td>
</tr>
<tr>
<td>$\partial T_T$</td>
<td>?</td>
<td>$(e^{\lambda B T})^{-\alpha/(2\alpha+3)}$</td>
</tr>
<tr>
<td>genealogical</td>
<td>$n^{-\alpha/(2\alpha+1)}$ + adaptation</td>
<td>$n^{-\alpha/(2\alpha+1)}$ + adaptation</td>
</tr>
<tr>
<td>$\hat{T}_T$</td>
<td>?</td>
<td>$(e^{\lambda B T})^{-\alpha/(2\alpha+1)}$</td>
</tr>
</tbody>
</table>
Bias sampling
  Age dependent model
  Size model: estimation at a large fixed time in a proxy model

Large population models

Nonlinear extensions, open questions
  Models of interacting neurons
  More non-linear models in a mean-field limit
Construction of the microscopic model

- \( b, \mu : \mathbb{R}_+ \times \mathbb{R}_+ \rightarrow \mathbb{R}_+ \) model parameters.

- \( b(t, a) \): fertility rate of the population with age \( a \) at time \( t \).
  \( \mu(t, a) \): mortality rate of the population with age \( a \) at time \( t \).

- \( Z_0 \) random variable with value in \( \mathcal{M}_F \), the set of finite point measures on \( \mathbb{R}_+ \): initial age distribution of the population at time \( t = 0 \).
Microscopic evolution equation

- Evolution equation for $t \in [0, T]$:

$$Z^N_t = \tau_t Z^N_0$$

$$+ N^{-1} \int_0^t \sum_{i \leq \langle Z^N_s, 1 \rangle} \int_{0 \leq \theta \leq b(s, a_i(Z^N_s))} \delta_{t-s}(da)Q_1(ds, di, d\theta)$$

$$- N^{-1} \int_0^t \sum_{i \leq \langle Z^N_s, 1 \rangle} \int_{0 \leq \theta \leq \mu(s, a_i(Z^N_s))} \delta_{a_i(Z^N_s) + t-s}(da)Q_2(ds, di, d\theta)$$

- $Q_i$: two independent random Poisson measures on $\mathbb{R}_+ \times \mathbb{N} \times \mathbb{R}_+$ with intensity $dt \left( \sum_{k \geq 1} \delta_k(di) \right) d\theta$. 
Microscopic evolution equation

Figure: Left: Sample path of $NZ_0^N(da)$ with $N = 3$ and its evolution without births. Right: Sample path of $(NZ_t^N(da), t \in [0, T])$.
Large population limit

- $N \to \infty$ abstract asymptotic parameter.
- Reminiscent of a population size: $\langle NZ_t^N, 1 \rangle \approx N$ for every $t \in [0, T]$.
- $T$ is fixed throughout!
- If $Z_0^N \approx g_0(a)da$, then $Z_t^N(da) \approx \xi_t(da) = g(t, a)da$.
- $g(t, a)$ weak solution to the McKendrick & Von Foerster equation

\[
\begin{cases}
\frac{\partial}{\partial t} g(t, a) + \frac{\partial}{\partial a} g(t, a) + \mu(t, a) g(t, a) = 0, \\
g(0, a) = g_0(a), \quad g(t, 0) = \int_{\mathbb{R}^+} b(t, a) g(t, a)da.
\end{cases}
\]
Identifiability of the parameters

- Under a suitable approximation $Z_0^N \approx \phi \sim \text{identification} \text{ of } \phi$.
- Need to understand how $Z_0^N \approx \phi$ propagates to $Z_t^N \approx g(t, \cdot)$ for $t \in [0, T]$.

**Claim:** Under “suitable propagation”, we can identify $g$ from $Z^N$.

**Claim:** Likewise, we can identify $\mu$ from $Z^N$.

- We cannot identify $b$ from $Z^N$ for lack of injectivity of $b \mapsto g$. 
First estimators

- **Statistical objective:** estimate $g(t, a)$ and $\mu(t, a)$ from data $(Z^N_t, t \in [0, T])$.

- **First kernel estimator of $g(t, a)$:**

  \[
  \hat{g}^{\text{prel}}_h(t, a) = \int_0^T \int_{\mathbb{R}^+} K_h(t - s, a - u)Z^N_s(du).
  \]

- We will see that both bias and variance of $\hat{g}^{\text{prel}}_h(t, a)$ behave poorly!
First estimator of the mortality rate $\mu$

- Extract from $Z^N$ the mortality process

\[
\Gamma^N(dt, da) = \sum_{k \geq 1} \delta(T^N_k, A^N_k),
\]

$(T^N_k, A^N_k) =$ (time of death, age at death) of the $k$-th occurrence of mortality.

- First kernel estimator of $\mu$:

\[
\hat{\mu}_{h}^{\text{prel}}(t, a) = \frac{\int_0^T \int_{\mathbb{R}^+} K_h(t - s, a - u) \Gamma^N(ds, du)}{\hat{g}(t, a)}
\]

given an estimator of $g(t, a)$ of $\hat{g}(t, a)$.

- Bias of $\hat{\mu}_{h}^{\text{prel}}(t, a)$ behaves poorly + inherits of the possible defects of $\hat{g}(t, a)$. 
Hölder regularity of the limit

- Look for the regularity of $g$ as the solution of the McKendrick & Von Foerster equation:

$$\frac{\partial}{\partial t} g(t, a) + \frac{\partial}{\partial a} g(t, a) + \mu(t, a) g(t, a) = 0$$

$$g(0, a) = \phi(a), \quad g(t, 0) = \int_{\mathbb{R}^+} b(t, a) g(t, a) da.$$

- Assumption: $b \in \mathcal{H}^{\alpha, \beta}, \mu \in \mathcal{H}^{\gamma, \delta}, \phi \in \mathcal{H}^{\nu}$ for some $\alpha > 0, \beta > 0, \gamma > 0, \delta > 0, \nu \gg 1$.

- Theorem
  
  We have

$$g^1_{\{a < t\}} \in \mathcal{H}^{\min(\alpha, \beta, \gamma + 1, \delta), \min(\alpha, \beta, \gamma + 1, \delta)}$$

and

$$g^1_{\{a > t\}} \in \mathcal{H}^{\min(\gamma + 1, \delta), \max(\min(\gamma, \delta + 1), \delta)}.$$
Hölder regularity of the limit

Figure: $g \in \mathcal{H}^{\min(\alpha, \beta, \gamma+1, \delta), \min(\alpha, \beta, \gamma+1, \delta)}$ on $\mathcal{D}_L$ and $g \in \mathcal{H}^{\min(\gamma+1, \delta), \max(\min(\gamma, \delta+1), \delta)}$ on $\mathcal{D}_U$. 
Hölder regularity of the limit

- “Improve” the smoothness of $g \sim \text{change of coordinates.}$
- With $\varphi(t, a) = (t, t - a)$, we have

\[ \mathcal{D}_L \xrightarrow{\varphi} \tilde{\mathcal{D}}_L = \mathcal{D}_L \text{ and } \mathcal{D}_U \xrightarrow{\varphi} \tilde{\mathcal{D}}_U = \{ a' < 0, 0 \leq t \leq T \}, \]

- Define $\tilde{g}$ via

\[ g(t, a) = \tilde{g} \circ \varphi(t, a) \]

- Theorem

We have (with $\tilde{g}(t, a') = g(t, t - a')$)

\[ \tilde{g}(t, a')1_{\{0 < a' < t\}} \in \mathcal{H}^{\min(\gamma + 1, \delta + 1), \min(\alpha, \beta, \gamma + 1, \delta)} \]

and

\[ \tilde{g}(t, a')1_{\{a' < 0\}} \in \mathcal{H}^{\min(\gamma + 1, \delta + 1), \max(\min(\gamma, \delta + 1), \delta)}. \]
Reconstruction of $g$

- Estimate $g$ on $\mathcal{D}_U = \{(t,a) \in (0, T) \times (0, a_{\text{max}}), t < a\}$ and $\mathcal{D}_L = \{(t,a), 0 < a < t < T\}$ separately.

\[ \hat{g}_{\text{prel}}(t,a) = \int_0^T \int_{\mathbb{R}_+} K_h(t-s, a-u) Z_s^N(du)ds. \]

- We estimate $g(t,a)$ in the direction suggested by $\tilde{g}(t,a)$ in order to benefit from its smoothness:

\[ \hat{g}_{\text{inter}}(t,a) = \int_0^T \int_{\mathbb{R}_+} K_h(t-s, (t-s) - (a-u)) Z_s^N(du)ds. \]
Reconstruction of $d$ via $\hat{g}_{N,h}$ and the process $\Gamma^N$

We also estimate $\mu(t, a) = \mu(t, a)g(t, a)/g(t, a)$ in the direction suggested by $\tilde{g}(t, a)$:

$$\hat{\mu}_{N,h,h'}^\text{inter}(t, a) = \frac{\int_0^T \int_{\mathbb{R}_+} K_h(t-s,(t-s)-(a-u)) \Gamma_s^N(du)}{\hat{g}_{N,h}^\text{inter}(t, a)}.$$
Stochastic error analysis for $g$

- We now look for a control $g_h(t, a) \approx \hat{g}_{N,h}^{\text{inter}}(t, a)$, with

$$g_h(t, a) = \int_0^T \int_{\mathbb{R}_+} K_h(t - s, (t - s) - (a - u)) g(s, u) duds.$$  

- We have

$$\hat{g}_{N,h}^{\text{inter}}(t, a) = \int_0^T \int_{\mathbb{R}_+} K_h(t - s, (t - s) - (a - u)) Z_s^N(du) ds$$

- We need

$$Z_s^N(du) ds \approx g(s, u) duds$$

in an appropriate sense (related to $K_h$) as $N \to \infty$. 
Toward a coherence property

- How does a suitable assumption on \( \text{dist}(Z^N_0, \xi_0) \) propagates to \( \text{dist}(Z^N_t, \xi_t) \) as \( N \to \infty \)? For which \( \text{dist}(\cdot, \cdot) \)? (coherence)

- Introduce a pseudo-distance related to a weight function \( \psi \in L^\infty(\mathbb{R}) \). For a suitable class of functions \( \mathcal{F} \) let

\[
\mathbb{W}_\psi(\mu, \nu) = \sup_{\varphi \in \mathcal{F}} \left| \int_{\mathbb{R}^+} \psi(a) \varphi(a) (\mu(da) - \nu(da)) \right|.
\]

- For instance, if \( \mathcal{F} \) consists of 1-Lipschitz functions, reminiscent of a weighted Wasserstein-1 distance in the degenerate case \( \psi = 1 \).
Toward a coherence property

- **Assume** $\mathbb{W}_\psi(Z_0^N, \xi_0) \lesssim w_N$ for some (small) $w_N$.

- **Seek** a bound of the form

  $$\mathbb{W}_\psi(?) (Z_t^N, \xi_t) \overset{P}{\lesssim} w_N + \delta_N \quad \text{for} \quad t \in [0, T]$$

  for some (small) $\delta_N$ that controls the error propagation.

- **For** $\delta_N \lesssim w_N$, we say that we have a coherence property.
Assumption: (Initial approximation): For some \( p \geq 2 \)

\[
\mathbb{E} \left[ W_\psi (Z_0^N, \xi_0)^p \right] \lesssim |\psi|_\infty^{p/2} |\psi|_1^{p/2} w_N^p
\]

with \( w_N \to 0 \) as \( N \to \infty \).

If \( Z_0^N = N^{-1} \sum_{i=1}^N \delta_{A_i} \) for IID \( A_i \), we expect \( w_N \approx N^{-1/2} \).
Coherence property

- $\mathcal{N}(\mathcal{F}, |\cdot|_\infty, \epsilon)$ minimal number of $\epsilon$-balls in $|\cdot|_\infty$ norm necessary to cover $\mathcal{F}$.

- **Assume:** $\int_0^1 \log (1 + \mathcal{N}(\mathcal{F}, |\cdot|_\infty, \epsilon)) d\epsilon < \infty$ + ‘some’ stability for $\mathcal{F}$.

**Theorem (Coherence property)**

*We have for all $t \in [0, T]*

\[
\mathbb{E}\left[\mathcal{W}_{\psi(t-\cdot)}(Z_t^N, \xi_t)^p\right] \lesssim |\psi|_\infty^{p/2} |\psi|_1^{p/2} w_N^p \vee N^{-p/2}
\]
Stochastic error analysis for $g$

- With $G = K^{(1)}(\cdot - t)$ and $H = K^{(2)}(\cdot -(t - a))$:

  $$\left| \hat{g}_{N,h}(t, a) - g_{h}(t, a) \right|$$

  $$= \left| \int_{0}^{T} G_{h_{1}}(s) \int_{\mathbb{R}_{+}} H_{h_{2}}(s - u) (Z_{s}^{N}(du) - g(s, u)) \, ds \right|$$

  $$\leq \int_{0}^{T} \left| G_{h_{1}}(s) \right| W_{H_{h_{2}}(s-\cdot)}(Z_{s}^{N}, \xi_{s}) \, ds$$

- Using the coherence property we get $\forall (t, a) \in D_{L} \cup D_{U}$

  $$\mathbb{E} \left[ \left| \hat{g}_{N,h}(t, a) - g_{h}(t, a) \right|^{2} \right] \lesssim w_{N}^{2} \vee N^{-1} \frac{|K^{(1)}|_{2}^{2}|K^{(2)}|_{\infty}|K^{(2)}|_{1}}{h_{1} h_{2}}.$$  

- Appended with the previous bias control
Convergence rates

- Anisotropic rate $v(t, a)^{-1}$

$$= \begin{cases} 
\min(\gamma + 1, \delta + 1)^{-1} + (\min(\alpha, \beta, \gamma + 1, \delta)^{-1} & \text{on } D_L(t, a) \\
\min(\gamma + 1, \delta + 1)^{-1} + (\max(\min(\gamma, \delta + 1), \delta)^{-1} & \text{on } D_U(t, a).
\end{cases}$$

Theorem

We have for pointwise (non-adaptive) optimisation of $h$:

$$\sup_{b, \mu, \phi, (t, a)} \mathbb{E} \left[ (\hat{g}_{N,h}^{\text{inter}}(t, a) - g(t, a))^2 \right] \lesssim (w_N^2 \lor N^{-1})^{2v(t,a)/(2v(t,a)+1)}.$$

- Supremum in $(t, a)$ over compacts of $D_L \cup D_U$ and in $(b, \mu, \phi)$ over (balls) of Hölder classes

- This result is not optimal!
Optimal estimation of $g$ (and subsequently $\mu$)

- The stochastic error for $\hat{g}_N,h$ is stable as $h_1 \to 0$!
- $G_{h_1}(t - \cdot) = G_{h_1=0}(t - \cdot) = \delta_t$ works! Estimating $g(t, \cdot)$ is a univariate problem, for each $t \in [0, T]$.
- This is no longer true for statistics based on $\Gamma_h(dt, da)$: need a bivariate anisotropic estimator for estimating $\mu(t, a)$ together with a choice of direction dictated by $\tilde{g}$.
- Final estimators

$$
\hat{g}_{N,h}^{\text{fin}}(t, a) = \int_{\mathbb{R}^+} K_h(a - u) Z_t^N(du)
$$

and

$$
\hat{\mu}_{N,h,h}^{\text{fin}}(t, a) = \frac{\int_0^T \int_{\mathbb{R}^+} K_h(t - s, (t - s) - (a - u)) \Gamma_s^N(du)}{\hat{g}_{N,h}^{\text{fin}}(t, a)}.
$$
Convergence rates for $\hat{g}_{n,h}^{\text{fin}}$

- Our (univariate) rate estimation for $g$:

$$\nu_1^*(t,a) = \min\{\alpha, \beta, \gamma+1, \delta\} 1_{D_L(t,a)} + \max(\min(\gamma, \delta+1), \delta) 1_{D_U(t,a)}.$$ 

**Theorem**

We have, $\forall (t,a) \in D_L \cup D_U$, for pointwise (non-adaptive) optimisation of $h$:

$$\sup_{b,\mu,\phi,(t,a)} \mathbb{E}\left[ (\hat{g}_{N,h}^{\text{fin}}(t,a) - g(t,a))^2 \right] \preceq (w_N^2 \lor N^{-1})^{2\nu_1^*(t,a)/(2\nu_1^*(t,a)+1)}.$$ 

- Minimax lower bound: $N^{-2 \min(\gamma,\delta)/(2 \min(\gamma,\delta)+1)}$.

- Minimax optimality: on $D_U$ if $\delta \leq \gamma \leq \delta + 1$ and on $D_L$ if $\delta - 1 \leq \gamma \leq \delta$ and $\delta \geq \gamma$. 
Convergence rates for $\hat{\mu}^{\text{fin}}_{N,h,h}$

- Our (bivariate) rate estimation for $\mu$: $v_2^*(t,a)$

$$= \begin{cases} 
\min(\gamma, \delta)^{-1} + \min(\alpha, \beta, \gamma + 1, \delta)^{-1} & \text{on } \mathcal{D}_L \\
\min(\gamma, \delta)^{-1} + \delta^{-1} & \text{on } \mathcal{D}_U.
\end{cases}$$

**Theorem**

We have, $\forall (t,a) \in \mathcal{D}_L \cup \mathcal{D}_U$, for pointwise (non-adaptive) optimisation of $h$:

$$\sup_{b,\mu,\phi}(t,a) \mathbb{E}\left[\left(\hat{\mu}^{\text{fin}}_{N,h,h}(t,a) - \mu(t,a)\right)^2\right] \lesssim (w_N^2 \lor N^{-1})^{2v_2^*(t,a)/(2v_2^*(t,a)+1)}.$$}

- Minimax lower bound: $N^{-2s(\gamma, \delta)/(2s(\gamma, \delta)+1)}$ with $s(\gamma, \delta)^{-1} = \gamma^{-1} + \delta^{-1}$.
- Minimax optimality: If $\gamma \leq \delta$ on $\mathcal{D}_U$ and if $\gamma \leq \delta \leq \gamma + 1$ on $\mathcal{D}_L$. 
Toward smoothness adaptation

Let

\[ W_\psi (\xi, \zeta) = \sup_{\varphi \in \mathcal{F}} \left| \int_0^T \int_{\mathbb{R}_+} \psi(s, s-u) \varphi(s, u) (\xi_s(du) - \zeta_s(du)) \right| ds. \]

Theorem

Under a proper modification of the initial approximation at \( t = 0 \), we have, with \( \xi^N = \Gamma^N \) (resp. \( Z^N \)) and \( \zeta = \mu g \) (resp. \( g \))

\[
P\left( W_\psi (\xi^N, \zeta) \geq C w_N \wedge N^{-1/2} (\|\psi\|_\infty \|\psi\|_1)^{1/2} + u \right) \leq \varepsilon_N(\psi, u)
\]

with \( \varepsilon_N(\psi, u) = C' (e^{C''Nu^2} (\|\psi\|_\infty \|\psi\|_1)^{-1} - 1)^{-1} \).

yields proper tools to study the deviation of \( \hat{g}^\text{fin}_{N,h}(t, a) - g_h(t, a) \) and \( \hat{\mu}^\text{fin}_{N,h}(t, a) - g_h(t, a) \rightsquigarrow \) adaptation.
Oracle inequalities

▶ Goldenschluger-Lepski → data driven bandwidth $\hat{h}_N$ and $\hat{h}_N$.

Theorem (Oracle inequality)
We have, for any $(t, a) \in D_L \cup D_U$

$$\mathbb{E}\left[ (\hat{f}_N(t, a) - f(t, a))^2 \right] \leq C \inf_{\kappa} \mathbb{E}\left[ (\hat{f}_{N,\kappa}(t, a) - f(t, a))^2 \right] + \delta_N,$$

with $\hat{f}_N = \hat{g}_{N, h_N}^{\text{fin}}$ (resp. $\hat{\mu}_{N, h, h_N}^{\text{fin}}$) and $f = g$ (resp. $\mu$) and $\kappa = h$ (resp. $(h, \mu)$), where $\delta_N = O(N^{-1})$ up to a constant depending on $b_{\text{max}}, \mu_{\text{max}}, T, \phi$.

▶ Adaptation over appropriate domains according to the preceding results.
Some numerical illustration

\[ \mu(t, a) = 4 \times 10^{-4} \exp(8 \times 10^{-3} a), \quad b =, \quad \phi(a)da \sim \mathcal{N}(60, 20^2) \]
conditioned upon \([0, 120]\).

**Figure:** Unknown \(g\). X-axis: time (0 to 100 years), Y-axis: age (0 to 120 years).
Some numerical illustration

- $N = 10^3, 5 \times 10^3, 10^4, 2 \times 10^4, 5 \times 10^4, 10^5$ over $10$ MC samples.
- $K^{(1)} = K^{(2)} = \text{Gaussian kernel}.$
- Calibration parameters... !

Figure: Rate estimation of $g(t, a)$. $(t, a) = (40, 60) \in D_U$ (left) and $(t, a) = (60, 90) \in D_L$ (right). Green = True, Blue = Oracle, Red = estimator via GL.
Some numerical illustration

- $N = 10^3, 5 \times 10^3, 10^4, 2 \times 10^4, 5 \times 10^4, 10^5$ over 10 MC samples.
- $K^{(1)} = K^{(2)} = \text{Gaussian kernel}.$
- Calibration parameters... !

Figure: Rate estimation of $\mu(t, a)$. $(t, a) = (40, 60) \in D_U$ (left) and $(t, a) = (60, 90) \in D_L$ (right). Green = True, Blue = Oracle, Red = estimator via GL.
Conclusion: needed improvements

- Complete minimax optimality (\(\rightsquigarrow\) shed light on the anisotropic structure).
- Study the birth rate estimation \(b(t, a)\) (inverse problem) \(\rightsquigarrow\) ill-posed. Modification of the problem.
- Generalisation to other transports and some interactions?
Generalisations: arbitrary transport + interactions

Can we extend our results to dynamics of the form

\[
\frac{\partial}{\partial t} g(t, a) + \frac{\partial}{\partial a} (v(a) g(t, a)) + \\
+ (\mu(t, a) + \int_{\mathbb{R}^+} U(a, y) g(t, y) dy) g(t, a) = 0,
\]

\[
g(0, a) = \phi(a), \quad g(t, 0) = \int_{\mathbb{R}^+} b(t, a) g(t, a) da.
\]

In particular, can we build consistent tests for detecting the presence of an interaction?
Bias sampling

- Age dependent model
- Size model: estimation at a large fixed time in a proxy model

Large population models

Nonlinear extensions, open questions

- Models of interacting neurons
- More non-linear models in a mean-field limit
A model of interacting neurons

- Modelling the evolution of the electrical potentials of a system of $N$ spiking neurons.


- Each neuron spikes randomly with rate $B(u)$ depending on the membrane potential $u$ of the neuron.
  1. At spiking time,
     - Spiking membrane is reset to a resting potential (here $u = 0$).
     - Action of chemical synapses increases the potential of other neurons by $N^{-1}$.
  2. Action of electrical synapses synchronises the potentials of the system.

- We model the distribution of membrane potentials of a system of $N$ neurons through time.
Example 4: a model of interacting neurons

- $(U_i(t))_{1 \leq t \leq N} = \text{the membrane potentials at time } t$.
- $Z^N_t = N^{-1} \sum_{i=1}^{N} \delta U_i(t)$.

Associated SDE

\[
Z^N_t = \phi Z^N_0(t) \\
+ \frac{1}{N} \int_0^t \sum_{i=1}^{N} \int_{0 \leq \theta \leq B(u_i(Z^N_s -))} (\delta \phi_0(t-s) - \delta \phi_{u_i(Z^N_s)}(t-s)) Q^i(ds, d\theta) \\
+ \frac{1}{N} \int_0^t \sum_{i=1, j \neq i}^{N} \int_{\theta \leq B(u_j(Z^N_s -))} (\delta \phi_{u_i(Z^N_s -) + N-1}(t-s) - \delta \phi_{u_i(Z^N_s)}(t-s)) Q^j(ds, d\theta).
\]

- $(Q^i)_{1 \leq i \leq N} = \text{independent Poisson measures, intensity } ds \otimes d\theta$.
- $\phi \sum_i \delta u_i(t) = \sum_i \delta \phi_{u_i}(t)$. 

Example 4: a model of interacting neurons

- Mean-field limit $N \to \infty$.
- Example 4.1: The simplest case when synchronization is ignored: $\phi_u(t) = u$ for every $t \geq 0$.
- If $Z_0^N \approx g_0(u)du$, then $Z_t^N(du) \approx \xi_t(du) = g(t, u)du$.
- $g(t, u)$ weak solution to the nonlinear evolution equation
  \[
  \begin{cases}
  \frac{\partial}{\partial t} g(t, u) + \langle g(t, \cdot), B \rangle \frac{\partial}{\partial u} g(t, u) + B(u)g(t, u) = 0, \\
  g(0, u) = g_0(u), \quad g(t, 0) = 1.
  \end{cases}
  \]
- The nonlinearity in the limiting model reflects the interactions of the individuals.
Example 4.2: a model of interacting neurons with stochastic flow

► Case of a stochastic flow \( \frac{d}{dt} \phi_x(t) = \kappa(\phi_x(t), Z^N_t) dt \), with mean-reverting

\[
\kappa(x, Z^N_t) = -\lambda(x - Z^N_t), \quad \lambda \geq 0.
\]

► If \( Z_N^0 \approx g_0(u) du \), then \( Z_N^t(du) \approx \xi_t(du) = g(t, u) du \).
► \( g(t, u) \) weak solution to the evolution equation

\[
\left\{ \begin{align*}
\partial_t g + \langle g(t, \cdot), B \rangle - \lambda u \rangle \partial_u g + (B(u) - \lambda)g = 0, \\
g(0, u) = g_0(u), \quad g(t, 0) = \frac{\langle g(t, \cdot), B \rangle}{\langle B + \lambda \cdot g(t, \cdot) \rangle}.
\end{align*} \right.
\]
Example 4.1 and 4.2: identification of the objects of interest

We can identify the following objects

- $N \to \infty$.
- $Z^N$ is $(Z^N_t)_{0 \leq t \leq T}$ and we observe $\mathcal{Z}^N = Z^N$ or a uniform sample of size $n \ll N$ extracted from $Z^N$.
- $f$ is $(t, u) \mapsto g(t, u)$ or $x \mapsto B(u)$.
- $\mathcal{H}^N$ and $\mathcal{H}$ are the SDE and the nonlinear transport evolution equation.

Observation schemes
Interaction between particles can play at various levels. We elaborate briefly on three more examples.

Example 3.2: Birth-and-death processes with population dependent death rate.

Example 5: Interacting Hawkes processes.

Example 6: The McKean-Vlasov model and the effect of diffusion.
Example 3.2: nonlinear death rate in population models

- In Example 3, we replace the death rate $B(t, a)$ by

\[
B(t, a, Z^N_t) = B(t, a) + \int_{\mathbb{R}^+} U(a, a') Z^N_t(da')
\]

for some kernel $U : \mathbb{R}^+ \times \mathbb{R}^+ \to \mathbb{R}^+$.

- The kernel $U$ accounts for some population dependent pressure on the death rate.

- If $Z^N_0 \approx g_0(a)da$, then $Z^N_t(da) \approx g(t, a)da$.

- $g(t, a)$ weak solution of the nonlinear evolution equation

\[
\begin{cases}
(\partial_t + \partial_a)g(t, a) + (B(t, a) + \int_{\mathbb{R}^+} U(a, a')g(t, a'))g(t, a) = 0, \\
g(0, u) = g_0(u), \quad g(t, 0) = \int_0^\infty b(t, a)g(t, a)da.
\end{cases}
\]
We consider a system of point processes interacting through their jump intensities.

Point process: \( N_t = \sum_{i \geq 1} 1_{\{T_i \leq t\}} \) where

\[
T_0 = 0 \leq T_1 < T_2 < \ldots < T_i < \ldots \quad \text{jump times}
\]

Simplest example: Poisson process with intensity \( \lambda > 0 \):
- The \( T_i - T_{i-1} \) are independent and \( \text{Exp}(\lambda) \) distributed.
- Alternative representation:

\[
N_t = \int_0^t \int_{0 \leq \theta \leq \lambda} Q(ds, d\theta)
\]

\( Q \): Poisson random measure with intensity \( ds \otimes d\theta \).
Example 5: univariate Hawkes processes

- **Nonlinear Hawkes processes**: replace $\lambda$ by a random past dependent stochastic intensity

$$\lambda_t = h(\lambda + \int_0^{t-} \varphi(t-s)dN_s),$$

- $h : \mathbb{R} \to \mathbb{R}_+$ ($h(x) = x$: linear Hawkes processes.)
- $\varphi : \mathbb{R} \to \mathbb{R}$ causal interacting kernel: $\text{Supp}(\varphi) \subset \mathbb{R}_+$.

- **Interpretation**: with $\mathcal{F}_t = \sigma(N_s, s \leq t),$

$$\mathbb{P}(N_{t+dt} - N_t \geq 1 | \mathcal{F}_{t-}) = \lambda_t dt.$$

- **Alternative representation as a SDE**:

$$N_t = \int_0^t \int_0^{s-} \int_0^{h(\lambda + \int_0^{s-} \varphi(s-u)dN_u)} Q(ds, d\theta).$$
Example 5: interacting Hawkes processes

- System of nonlinear Hawkes processes: defined by the family of SDE: for $i = 1, \ldots, N,$

$$N^i_t = \int_0^t \int_{0 \leq \theta \leq h} h(\lambda + N^{-1} \sum_{j=1}^N \int_{s-\theta}^s \varphi(s-u) dN^j_u) Q^i(ds, d\theta),$$

$Q^i$ ind. Poisson, intens. $ds \otimes d\theta$.

- $Z^N_t = N^{-1} \sum_{i=1}^N \delta_{N^i_t}$.

- Mean-field limit: $Z^N_t(ds) \approx g(t, ds)$ as $N \to \infty$.

- $g$ is a weak solution of

$$\begin{cases} 
\partial_t g(t, s) + h(\int_0^t \varphi(t-u) dm_u)(g(t, s) - g(t, s-1)) = 0, \\
\delta_0(ds) = g(0, s), \\
m_t = \int_0^t h(\int_0^s \varphi(s-u) dm_u) ds.
\end{cases}$$
Example 6: McKean-Vlasov model

- System of $N$ interacting diffusion processes:

$$dX_t^i = -b(X_t^i)dt - N^{-1} \sum_{j=1}^N U(X_t^i - X_t^j)dt + \sigma dB_t^i, \quad i = 1, \ldots, N,$$

$B_t^i$ ind. Brownian motions.

- $Z_t^N = N^{-1} \sum_{i=1}^N \delta_{N_i}.$

- Mean-field limit: if $Z_0^N(dx) \approx g_0(dx)$, then $Z_t^N(dx) \approx g(t, x)dx$.

- $g(t, s)$ is a weak solution to the McKean-Vlasov equation

$$\begin{cases} 
\partial_t g(t, x) + \partial_x g(b + U * g) = \frac{\sigma^2}{2} \partial^2_x g, \\
g(0, x) = g_0(dx). 
\end{cases}$$
THANK YOU FOR YOUR ATTENTION!
Some references


