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:

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

$$Z^N \to \xi \text{ limiting object,}$$

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

- Objective: recover f from the observation of Z^N (or a proxy $\mathcal{Z}^N$ 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

$$(Z_t)_{0\le t\le T} \quad \text{or} \quad Z_T$$

- $Z_t = \sum_{i=1}^{N_t} \delta_{A_i(t)}$ with $g(t,\cdot) = \mathbb{E}[Z_t^N]$.
- Heuristically $Z_T \approx g(T,\cdot)$ when T is large.
- $N = \mathbb{E}[\langle Z_T, \mathbf{1}\rangle] \to \infty$ as $T \to \infty$.

Observation scheme

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

$$\mathcal{T}_T = \{u\in\mathbb{T}, b_u \leq T\} = \mathring{\mathcal{T}}_T \cup \partial\mathcal{T}_T,$$
$$\mathring{\mathcal{T}}_T = \{u\in\mathbb{T}, d_u \leq T\},$$
$$\partial\mathcal{T}_T = \{u\in\mathbb{T}, b_u \leq T < d_u\}.$$

- 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 $\mathring{\mathcal{T}}_T \cup \partial \mathcal{T}_T$

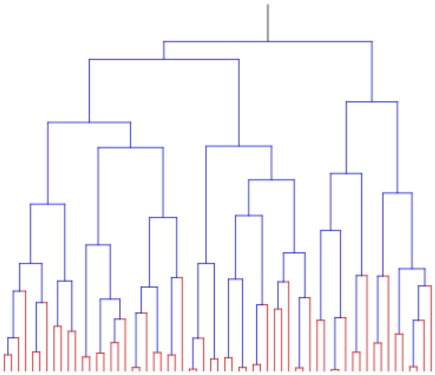


Figure: *A sample path of $Z_t(da)_{0 \leq t \leq T}$ with $B(a) = a^2$ and $T = 7$.*

Estimation of $B(a)$ from $\mathcal{T}_T$

- Many-to-one formula: For nice test functions φ:

$$\boxed{\mathbb{E}\Big[\sum_{u \in \mathring{\mathcal{T}}_T} \varphi(\zeta_u)\Big] = \int_0^T e^{\lambda_B s}\mathbb{E}\big[\varphi(\chi(s))H_B\big(\chi(s)\big)\big]\,ds}$$

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

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

- λ_B: Malthus parameter, related to χ 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 $\mathcal{T}_T$

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

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

- Rate of convergence: $\left(e^{\lambda_B T}\right)^{1/2} = N^{1/2}$ 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 $u, v \in \mathring{\mathcal{T}}_T$ + 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 $\widehat{\lambda}_T$ of λ_B.
- Set

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

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

$$\boxed{N^{\alpha/(2\alpha+1)}\Big(\widehat{B}^T_{\widehat{h}^T(\alpha)}(a)-B(a)\Big)}$$

  uniformly over $\mathcal{B}\cap\mathcal{H}^\alpha$ for appropriate $\mathcal{B}$.
- The rate is nearly minimax.
- Open problem: we do not have adaptation, for lack of concentration inequalities.

What if data are taken from $\partial \mathcal{T}_T$ solely?

- By another many-to-one formula, we have for good test functions φ

$$|\partial \mathcal{T}_T|^{-1} \sum_{u \in \partial \mathcal{T}_T} \varphi(\zeta_u) \xrightarrow{\mathbb{P}} 2\lambda_B \int_0^\infty \varphi(a) e^{\lambda_B a} \frac{f_B(a)}{B(a)} da$$
$$= 2\lambda_B \int_0^\infty \varphi(a) e^{\lambda_B a} e^{-\int_0^a B(s) ds} da.$$

- 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

$$\boxed{N_B^{\alpha/(2\alpha+3)}}$$

but not $N^{\alpha/(2\alpha+1)}$!.

The age dependent model, simulated data

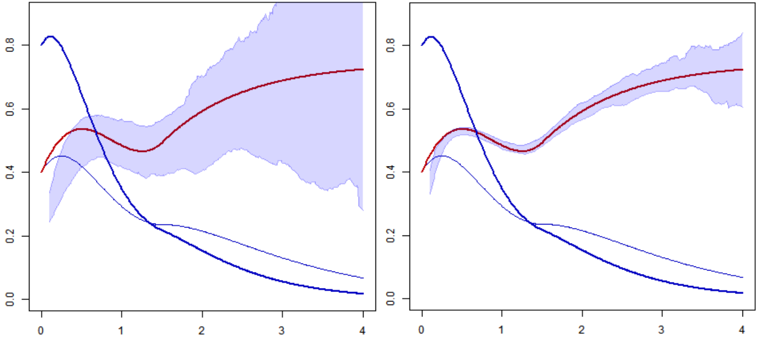


Figure: *Reconstruction of* B *over* $\mathcal{D} = [0.1, 4]$ *with* 95%*-level confidence bands constructed over* $M = 100$ *Monte-Carlo trees. In bold red line:* $x \rightsquigarrow B(x)$*; in bold blue line:* f_{H_B}*; in blue line:* f_B*. Left:* $T = 15$*. Right:* $T = 23$*.*

Size dependent division rate $B(x)$

- The associated deterministic model is

$$\begin{cases} \partial_t g(t,x) + \partial_x\big(\kappa(x)g(t,x)\big) + B(x)g(t,x) = 4B(2x)g(t,2x) \\ g(0,x) = g_0(x),\ g(t,0) = 0. \end{cases}$$

- 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.
- $N = \mathbb{E}[\langle Z_T, \mathbf{1}\rangle] \to \infty$ 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 ($\kappa(x) = \tau x$)

$$\partial_t g(t,x) + \partial_x\big(\tau x g(t,x)\big) + B(x)g(t,x) = 4B(2x)g(t,2x)$$

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

$$\boxed{\partial_x\big(\tau x\, N(x)\big) + \big(\lambda + B(x)\big)N(x) = 4B(2x)N(2x)}.$$

- 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" $X_1, \dots, X_n$ of cell sizes.
  2. Postulate the approximation
  $$\mathbb{P}(X_1 \in dx_1, \dots, X_n \in dx_n) \approx \otimes_{i=1}^{n} N(x_i) dx_i.$$
  If $n \to \infty$ but $n \ll N$, hope for a chaos propagation property.
  3. Recover B through the representation
  $$L(N) = \mathfrak{L}(BN),$$
  or
  $$\boxed{B = \frac{\mathfrak{L}^{-1} L(N)}{N}}$$
  with
  $$L(\varphi)(x) = \partial_x\big(\tau x \varphi(x)\big) + \lambda \varphi(x),$$
  $$\mathfrak{L}(\varphi)(x) = 4\varphi(2x) - \varphi(x).$$
- The operator $L(\cdot)$ has ill-posedness degree of order 1. The operator $\mathfrak{L}$ is "nicer".

Growth-fragmentation: a word of conclusion

data	Size model	Age model
proxy model	$n^{-\alpha/(2\alpha+3)}$ + adaptation	irrelevant
$\partial \mathcal{T}_T$	?	$(e^{\lambda_B T})^{-\alpha/(2\alpha+3)}$
genealogical	$n^{-\alpha/(2\alpha+1)}$ + adaptation	$n^{-\alpha/(2\alpha+1)}$ + adaptation
$\mathring{\mathcal{T}}_T$	?	$(e^{\lambda_B T})^{-\alpha/(2\alpha+1)}$

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

- $b, \mu : \mathbb{R}_+ \times \mathbb{R}_+ \to \mathbb{R}_+$ 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 $t \in [0, T]$:

$$
\begin{aligned}
Z_t^N = \; & \tau_t Z_0^N \\
& + N^{-1} \int_0^t \sum_{i \leq \langle Z_{s-}^N, \mathbf{1} \rangle} \int_{0 \leq \theta \leq b(s, a_i(Z_{s-}^N))} \delta_{t-s}(da) Q_1(ds, di, d\theta) \\
& - N^{-1} \int_0^t \sum_{i \leq \langle Z_{s-}^N, \mathbf{1} \rangle} \int_{0 \leq \theta \leq \mu(s, a_i(Z_{s-}^N))} \delta_{a_i(Z_{s-})+t-s}(da) Q_2(ds, di, d\theta)
\end{aligned}
$$

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

Microscopic evolution equation

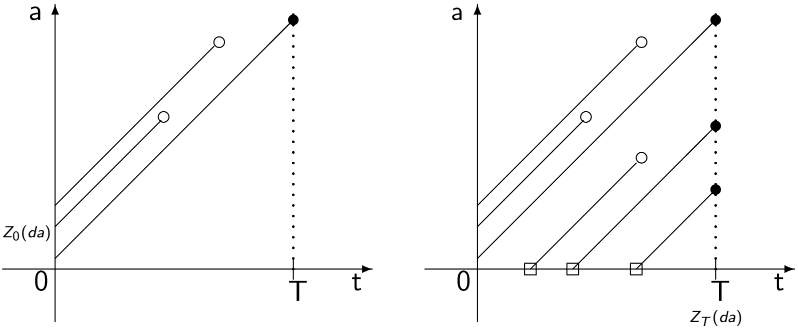


Figure: *Left*: Sample path of $NZ_0^N(da)$ with $N = 3$ and its evolution without births. *Right:* Sample path of $(NZ_t^N(da), t \in [0, T])$.

Large population limit

- $N \to \infty$ abstract asymptotic parameter.
- Reminiscent of a population size : $\langle N Z_t^N, \mathbf{1} \rangle \approx N$ 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 $Z_0^N \approx \phi \rightsquigarrow$ identification of ϕ.
- 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 μ 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 $(Z_t^N, t \in [0,T])$.
- First kernel estimator of $g(t,a)$:

$$\boxed{\widehat{g}_{\boldsymbol{h}}^{\texttt{prel}}(t,a) = \int_0^T \int_{\mathbb{R}_+} K_{\boldsymbol{h}}(t-s, a-u) Z_s^N(du).}$$

- We will see that both bias and variance of $\widehat{g}_{\boldsymbol{h}}^{\texttt{prel}}(t,a)$ behave poorly!

First estimator of the mortality rate μ

- Extract from Z^N the mortality process

$$\Gamma^N(dt, da) = \sum_{k\geq 1} \delta_{(T_k^N, A_k^N)},$$

$(T_k^N, A_k^N) = (\text{time of death}, \text{age at death})$ of the k-th occurence of mortality.

- First kernel estimator of μ:

$$\widehat{\mu}_{\boldsymbol{h}}^{\texttt{prel}}(t,a) = \frac{\int_0^T \int_{\mathbb{R}_+} K_{\boldsymbol{h}}(t-s, a-u)\Gamma^N(ds, du)}{\widehat{g}(t,a)}$$

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

- bias of $\widehat{\mu}_{\boldsymbol{h}}^{\texttt{prel}}(t,a)$ behaves poorly + inherits of the possible defects of $\widehat{g}(t,a)$.

Hölder regularity of the limit

- Look for the regularity of g as the solution of the Mc Kendrick & Von Foerster equation:

$$\begin{cases} \frac{\partial}{\partial t} g(t,a) + \frac{\partial}{\partial a} g(t,a) + \mu(t,a) g(t,a) = 0 \\ g(0,a) = \phi(a), \quad g(t,0) = \int_{\mathbb{R}_+} b(t,a) g(t,a) da. \end{cases}$$

- 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

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

  and

$$g\mathbf{1}_{\{a>t\}} \in \mathcal{H}^{\min(\gamma+1,\delta),\max(\min(\gamma,\delta+1),\delta)}.$$

Hölder regularity of the limit

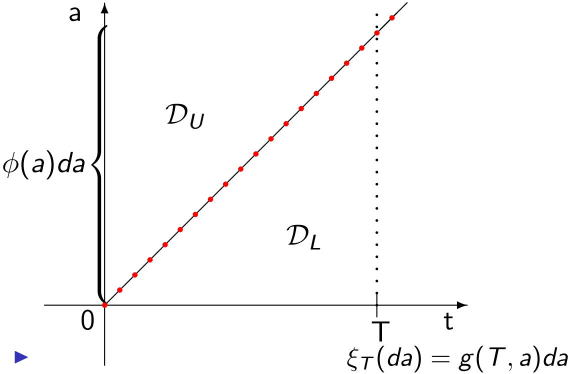


Figure: $g \in \mathcal{H}^{\min(\alpha,\beta,\gamma+1,\delta),\min(\alpha,\beta,\gamma+1,\delta)}$ on $\mathcal{D}_L$ and $g \in \mathcal{H}^{\min(\gamma+1,\delta),\max(\min(\gamma,\delta+1),\delta)}$ on $\mathcal{D}_U$.

Hölder regularity of the limit

- "Improve" the smoothness of g $\rightsquigarrow$ change of coordinates.
- With $\varphi(t,a) = (t, t-a)$, we have

$$\mathcal{D}_L \xrightarrow{\varphi} \widetilde{\mathcal{D}}_L = \mathcal{D}_L \text{ and } \mathcal{D}_U \xrightarrow{\varphi} \widetilde{\mathcal{D}}_U = \{a' < 0, 0 \le t \le T\},$$

- Define $\widetilde{g}$ via

$$\boxed{g(t,a) = \widetilde{g} \circ \varphi(t,a)}$$

- Theorem

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

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

and

$$\widetilde{g}(t,a')\mathbf{1}_{\{a'<0\}} \in \mathcal{H}^{\min(\gamma+1,\delta+1),\max(\min(\gamma,\delta+1),\delta)}.$$

Reconstruction of g

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

$$\widehat{g}_{\boldsymbol{h}}^{\texttt{prel}}(t,a) = \int_0^T \int_{\mathbb{R}_+} K_{\boldsymbol{h}}\left(t-s, a-u\right) Z_s^N(du)ds.$$

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

$$\boxed{\widehat{g}_{N,\boldsymbol{h}}^{\texttt{inter}}(t,a) = \int_0^T \int_{\mathbb{R}_+} K_{\boldsymbol{h}}\big(t-s, (t-s)-(a-u)\big) Z_s^N(du)ds.}$$

Reconstruction of d via $\widehat{g}_{N,\boldsymbol{h}}$ and the process Γ^N

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

$$\widehat{\mu}^{\texttt{inter}}_{N,\boldsymbol{h},h'}(t,a) = \frac{\int_0^T \int_{\mathbb{R}_+} K_{\boldsymbol{h}}\big(t-s,(t-s)-(a-u)\big)\Gamma^N_s(du)}{\widehat{g}^{\texttt{inter}}_{N,h}(t,a)}.$$

Stochastic error analysis for g

- We now look for a control $g_{\boldsymbol{h}}(t,a) \approx \widehat{g}^{\mathtt{inter}}_{N,\boldsymbol{h}}(t,a)$, with

$$g_{\boldsymbol{h}}(t,a) = \int_0^T \int_{\mathbb{R}_+} K_{\boldsymbol{h}}\big(t-s,(t-s)-(a-u)\big) g(s,u)\,du\,ds.$$

- We have

$$\widehat{g}^{\mathtt{inter}}_{N,\boldsymbol{h}}(t,a) = \int_0^T \int_{\mathbb{R}_+} K_{\boldsymbol{h}}\big(t-s,(t-s)-(a-u)\big) Z_s^N(du)\,ds$$

- We need

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

  in an appropriate sense (related to $K_{\boldsymbol{h}}$) as $N \to \infty$.

Toward a coherence property

- How does a suitable assumption on $\mathrm{dist}(Z_0^N, \xi_0)$ propagates to $\mathrm{dist}(Z_t^N, \xi_t)$ as $N \to \infty$? For which $\mathrm{dist}(\cdot, \cdot)$? (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

$$\boxed{\mathbb{W}_\psi(\mu, \nu) = \sup_{\varphi \in \mathcal{F}} \Big| \int_{\mathbb{R}_+} \psi(a)\varphi(a)\big(\mu(da) - \nu(da)\big) \Big|}.$$

- 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) δ_N that controls the error propagation.
- For $\delta_N \lesssim w_N$, we say that we have a coherence property.

Toward a coherence property

- Assumption: (Initial approximation): For some $p \geq 2$

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

  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 ϵ-balls in $|\cdot|_\infty$ norm necessary to cover $\mathcal{F}$.
- Assume: $\int_0^1 \log\left(1 + \mathcal{N}(\mathcal{F}, |\cdot|_\infty, \epsilon)\right) d\epsilon < \infty$ + 'some' stability for $\mathcal{F}$.

Theorem (Coherence property)

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

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

Stochastic error analysis for g

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

$$
\begin{aligned}
&|\widehat{g}_{N,\boldsymbol{h}}(t,a) - g_{\boldsymbol{h}}(t,a)| \\
=&\Big| \int_0^T G_{h_1}(s) \int_{\mathbb{R}_+} H_{h_2}(s-u)\big(Z_s^N(du) - g(s,u)\big)\, ds \Big| \\
\leq &\int_0^T |G_{h_1}(s)| \mathbb{W}_{H_{h_2}(s-\cdot)}(Z_s^N, \xi_s)\, ds
\end{aligned}
$$

- Using the coherence property we get $\forall (t,a) \in \mathcal{D}_L \cup \mathcal{D}_U$

$$
\boxed{\mathbb{E}\left[\left|\widehat{g}_{N,\boldsymbol{h}}(t,a) - g_{\boldsymbol{h}}(t,a)\right|^2\right] \lesssim w_N^2 \vee N^{-1} \frac{|K^{(1)}|_2^2 |K^{(2)}|_\infty |K^{(2)}|_1}{h_1 h_2}}.
$$

- Appended with the previous bias control

Convergence rates

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

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

Theorem

We have for pointwise (non-adaptive) optimisation of $\boldsymbol{h}$*:*

$$\sup_{b,\mu,\phi,(t,a)} \mathbb{E}\Big[\big(\widehat{g}^{\,\mathtt{inter}}_{N,\boldsymbol{h}}(t,a) - g(t,a)\big)^2\Big] \lesssim (w_N^2 \vee N^{-1})^{2v(t,a)/(2v(t,a)+1)}.$$

- Supremum in (t,a) over compacts of $\mathcal{D}_L \cup \mathcal{D}_U$ and in (b,μ,ϕ) over (balls) of Hölder classes
- This result is not optimal!

Optimal estimation of g (and subsequently μ)

- The stochastic error for $\widehat{g}^{\,\text{inter}}_{N,\boldsymbol{h}}$ 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 $\Gamma^N(dt, da)$: need a bivariate anisotropic estimator for estimating $\mu(t,a)$ together with a choice of direction dictated by $\widetilde{g}$.
- Final estimators

$$\widehat{g}^{\,\mathtt{fin}}_{N,h}(t,a) = \int_{\mathbb{R}_+} K_{\boldsymbol{h}}(a-u)\, Z^N_t(du)$$

and

$$\widehat{\mu}^{\,\mathtt{fin}}_{N,\boldsymbol{h},h}(t,a) = \frac{\int_0^T \int_{\mathbb{R}_+} K_{\boldsymbol{h}}\big(t-s, (t-s)-(a-u)\big)\Gamma^N_s(du)}{\widehat{g}^{\,\mathtt{fin}}_{N,h}(t,a)}.$$

Convergence rates for $\widehat{g}^{\mathtt{fin}}_{n,h}$

- Our (univariate) rate estimation for g:

$$v_1^\star(t,a) = \min\{\alpha, \beta, \gamma+1, \delta\}\mathbf{1}_{\mathcal{D}_L(t,a)} + \max(\min(\gamma, \delta+1), \delta)\mathbf{1}_{\mathcal{D}_U(t,a)}.$$

Theorem

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

$$\sup_{b,\mu,\phi,(t,a)} \mathbb{E}\Big[\big(\widehat{g}^{\mathtt{fin}}_{N,\boldsymbol{h}}(t,a) - g(t,a)\big)^2\Big] \lesssim (w_N^2 \vee N^{-1})^{2v_1^\star(t,a)/(2v_1^\star(t,a)+1)}.$$

- Minimax lower bound: $N^{-2\min(\gamma,\delta)/(2\min(\gamma,\delta)+1)}$.
- Minimax optimality: on $\mathcal{D}_U$ if $\delta \leq \gamma \leq \delta + 1$ and on $\mathcal{D}_L$ if $\delta - 1 \leq \gamma \leq \delta$ and $\delta \geq \gamma$.

Convergence rates for $\widehat{\mu}^{\mathtt{fin}}_{N,\boldsymbol{h},h}$

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

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

Theorem

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

$$\boxed{\sup_{b,\mu,\phi,(t,a)} \mathbb{E}\Big[\big(\widehat{\mu}^{\mathtt{fin}}_{N,\boldsymbol{h},h}(t,a) - \mu(t,a)\big)^2\Big] \lesssim (w_N^2 \vee N^{-1})^{2v_2^\star(t,a)/(2v_2^\star(t,a)+1)}}.$$

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

Toward smoothness adaptation

- Let

$$\mathbb{W}_\psi(\xi_\cdot,\zeta_\cdot) = \sup_{\varphi\in\mathcal{F}} \Big| \int_0^T \int_{\mathbb{R}_+} \psi(s,s-u)\varphi(s,u)\big(\xi_s(du)-\zeta_s(du)\big)\Big| ds.$$

Theorem

Theo 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)

$$\boxed{P\Big(\mathbb{W}_\psi(\xi^N,\zeta) \geq C w_N \wedge N^{-1/2}(\|\psi\|_\infty\|\psi\|_1)^{1/2}+u\Big) \leq \varepsilon_N(\psi,u)}$$

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

- yields proper tools to study the deviation of $\widehat{g}^{\mathtt{fin}}_{N,h}(t,a)-g_h(t,a)$ and $\widehat{\mu}^{\mathtt{fin}}_{N,\boldsymbol{h}}(t,a)-g_{\boldsymbol{h}}(t,a) \rightsquigarrow$ adaptation.

Oracle inequalities

- Goldenschluger-Lepski $\rightsquigarrow$ data driven bandwidth $\widehat{h}_N$ and $\widehat{\boldsymbol{h}}_N$.

Theorem (Oracle inequality)

We have, for any $(t,a) \in \mathcal{D}_L \cup \mathcal{D}_U$

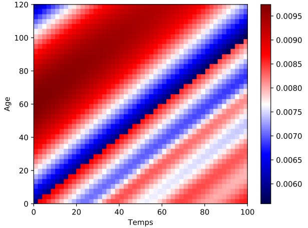
$$\mathbb{E}\Big[\big(\widehat{f}_N(t,a) - f(t,a)\big)^2\Big] \leq C \inf_{\kappa} \mathbb{E}\Big[\big(\widehat{f}_{N,\kappa}(t,a) - f(t,a)\big)^2\Big] + \delta_N,$$

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

- Adaptation over appropriate domains according to the preceding results.

Some numerical illustration

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



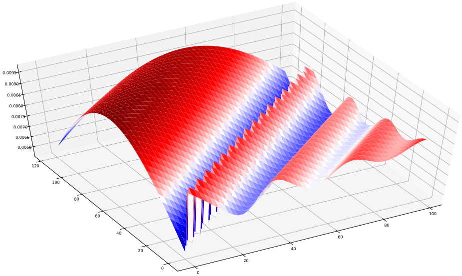


Figure: Unknown g. X-axis: time (0 to 100 years), Y-axis: age (0 to 120 years).

Some numerical illustration

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

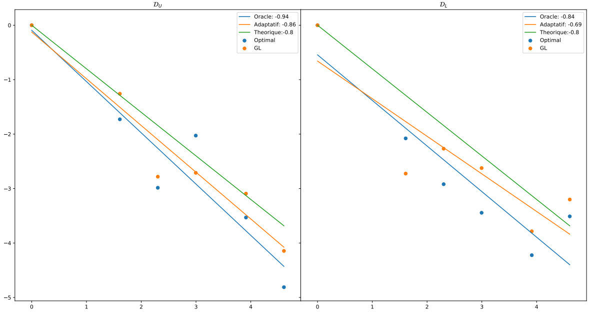


Figure: Rate estimation of $g(t, a)$. $(t, a) = (40, 60) \in \mathcal{D}_U$ (left) and $(t, a) = (60, 90) \in \mathcal{D}_L$ (right). Green = True, Blue = Oracle, Red = estimator via GL.

Some numerical illustration

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

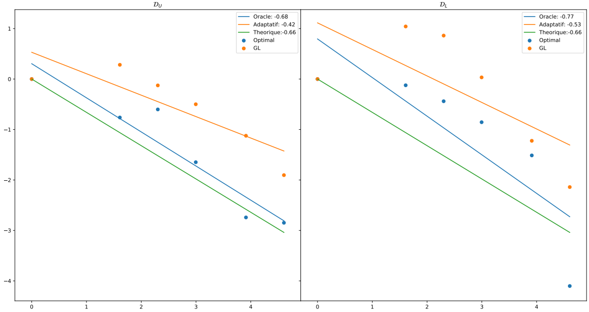


Figure: Rate estimation of $\mu(t,a)$. $(t,a) = (40,60) \in \mathcal{D}_U$ (left) and $(t,a) = (60,90) \in \mathcal{D}_L$ (right) Green = True, Blue = Oracle, Red = estimator via GL.

Conclusion: needed improvements

s

- 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

$$\begin{cases} \frac{\partial}{\partial t} g(t,a) + \frac{\partial}{\partial a}\big(v(a) g(t,a)\big) + \\ + \Big(\mu(t,a) + \int_{\mathbb{R}_+} U(a,y) g(t,y) dy\Big) g(t,a) = 0, \\ g(0,a) = \phi(a), \;\; g(t,0) = \int_{\mathbb{R}_+} b(t,a) g(t,a) da. \end{cases}$$

- 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.
- De Masi *et al.* (2015), Löcherbach and Fournier (2015) following De Masi and Galvez (2013).
- 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

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

Associated SDE

$$
\begin{aligned}
Z^N_t = &\ \phi_{Z^N_0}(t) \\
&+\frac{1}{N}\int_0^t \sum_{i=1}^N \int_{0\leq\theta\leq B(u_i(Z^N_{s-}))} \left(\delta_{\phi_0(t-s)} - \delta_{\phi_{u_i(Z^N_{s-})}(t-s)}\right) Q^i(ds,d\theta) \\
&+\frac{1}{N}\int_0^t \sum_{i=1,j\neq i}^N \int_{\theta\leq B(u_j(Z_{s-}))} \left(\delta_{\phi_{u_i(Z^N_{s-})+N^{-1}}(t-s)} - \delta_{\phi_{u_i(Z^N_{s-})}(t-s)}\right) Q^j(ds,d\theta).
\end{aligned}
$$

- $(Q^i)_{1\leq i\leq N}$ 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 $\frac{d}{dt}\phi_x(t) = \kappa(\phi_x(t), Z_t^N)dt$, with mean-reverting

$$\kappa(x, Z_t^N) = -\lambda(x - Z_t^N), \quad \lambda \geq 0.$$

- 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 evolution equation

$$\begin{cases} \partial_t g + \left(\langle g(t,\cdot), B\rangle - \lambda u\right)\partial_u g + (B(u) - \lambda)g = 0, \\ g(0,u) = g_0(u), \quad g(t,0) = \frac{\langle g(t,\cdot), B\rangle}{\langle B + \lambda\cdot, g(t,\cdot)\rangle}. \end{cases}$$

Example 4.1 and 4.2: identification of the objects of interest

We can identify the following objects

- $N \to \infty$.
- Z^N is $(Z_t^N)_{0 \leq t \leq T}$ 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

More non-linear models in a large population model

- 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

$$B(t, a, Z_t^N) = B(t, a) + \int_{\mathbb{R}_+} U(a, a') Z_t^N(da')$$

  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 $Z_0^N \approx g_0(a)da$, then $Z_t^N(da) \approx g(t, a)da$.
- $g(t, a)$ weak solution of the nonlinear evolution equation

$$\begin{cases} (\partial_t + \partial_a) g(t, a) + \left(B(t, a) + \int_{\mathbb{R}_+} U(a, a') g(t, a')\right) g(t, a) = 0, \\ g(0, u) = g_0(u), \;\; g(t, 0) = \int_0^\infty b(t, a) g(t, a) da. \end{cases}$$

Example 5: interacting Hawkes processes

- We consider a system of point processes interacting through their jump intensities.
- Point process: $N_t = \sum_{i \geq 1} \mathbf{1}_{\{T_i \leq t\}}$ 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 $\mathrm{Exp}(\lambda)$ 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 λ by a random past dependent stochastic intensity

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

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

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

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

- Alternative representation as a SDE:

$$N_t = \int_0^t \int_{0 \leq \theta \leq h\left(\lambda + \int_0^{s-} \varphi(s-u)dN_u\right)} Q(ds, d\theta).$$

Example 5: interacting Hawkes processes

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

$$N_t^i = \int_0^t \int_{0 \le \theta \le h\left(\lambda + N^{-1} \sum_{j=1}^N \int_0^{s-} \varphi(s-u) dN_u^j\right)} Q^i(ds, d\theta),$$

  Q^i ind. Poisson, intens. $ds \otimes d\theta$.
- $Z_t^N = N^{-1} \sum_{i=1}^N \delta_{N_t^i}$.
- Mean-field limit: $Z_t^N(ds) \approx g(t, ds)$ as $N \to \infty$.
- g is a weak solution of

$$\begin{cases} \partial_t g(t,s) + h\left(\int_0^t \varphi(t-u) dm_u \right) \left(g(t,s) - g(t,s-1) \right) = 0, \\ g(0,s) = \delta_0(ds), m_t = \int_0^t h\left(\int_0^s \varphi(s-u) dm_u \right) ds. \end{cases}$$

Example 6: McKean-Vlasov model

- System of N interacting diffusion processes :

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

  B_t^i ind. Brownian motions.
- $Z_t^N = N^{-1}\sum_{i=1}^{N} \delta_{N_t^i}$.
- 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

$$\begin{cases} \partial_t g(t,x) + \partial_x g\left(b + U \star g\right) = \frac{\sigma^2}{2}\partial_x^2 g, \\ g(0,x) = g_0(dx). \end{cases}$$

THANK YOU FOR YOUR ATTENTION!

Some references

- A. Boumezoued, M.H. and P. Jeunesse. *Nonparametric adaptive inference of birth and death models in a large population limit*, Preprint, 2019. arXiv:1903.00673.
- M.H., A. Marguet. *Statistical estimation in a randomly structured branching population*. *SPA*, 2019. arXiv:1712.04404
- S.V. Bitseki Penda, M.H. and A. Olivier. *Adaptive estimation for bifurcating Markov chains*. *Bernoulli*, 23, 3598–3637, 2017. arXiv:1509.03119
- M.H. and A. Olivier. *Nonparametric estimation of the division rate of an age dependent branching process*. *SPA*, 126, 1433-1471, 2016. arXiv:1412.5936.
- M. Doumic, M.H., N. Krell and L. Robert. *Statistical estimation of a growth-fragmentation model observed on a genealogical tree*. *Bernoulli*, 21, 1760–1799, 2015. arXiv:1210.3240.
- M. Doumic, M.H., P. Reynaud-Bouret and V. Rivoirard. *Nonparametric estimation of the division rate of a size-structured population*. *SINUM*, 50, 925-950, 2012. arXiv:1103.4226