Deterministic models and statistical aspects

M. Doumic-Jauffret, P. Reynaud-Bouret

INRIA Rocquencourt, Nice

Journées PDMP, 26-28 Mars 2012, Marne-La-Vallée

Asymptotics of the PDE

(equal mitosis)

Size-Structured Population Equation (asymptotics)

$$\begin{cases} \kappa \dfrac{\partial}{\partial x}\big(g(x)N(x)\big) + \lambda N(x) = \mathcal{L}\big(BN\big)(x), \\ B(0)N(0) = 0, \qquad \int N(x)dx = 1, \end{cases}$$

where

- for any real-valued function $x \rightsquigarrow \varphi(x)$, $\mathcal{L}\big(\varphi\big)(x) := 4\varphi(2x) - \varphi(x)$.
- $\kappa = \lambda \frac{\int_{\mathbb{R}_+} xN(x)dx}{\int_{\mathbb{R}_+} g(x)N(x)dx}$.

The inverse problem

Under the previous differential equation, we consider the inverse problem of finding B given a "noisy" version of N.

The inverse problem

Under the previous differential equation, we consider the inverse problem of finding B given a "noisy" version of N.

- Practical: biologists take a sample of, say, plankton in a lake, and they look at the respective size of the cells.

The inverse problem

Under the previous differential equation, we consider the inverse problem of finding B given a "noisy" version of N.

- Practical: biologists take a sample of, say, plankton in a lake, and they look at the respective size of the cells. Then they perform a preprocessing, by, say a kernel estimator. This is N_ϵ.

The inverse problem

Under the previous differential equation, we consider the inverse problem of finding B given a "noisy" version of N.

- Practical: biologists take a sample of, say, plankton in a lake, and they look at the respective size of the cells. Then they perform a preprocessing, by, say a kernel estimator. This is N_ϵ. (probably more approximation than that).

The inverse problem

Under the previous differential equation, we consider the inverse problem of finding B given a "noisy" version of N.

- Practical: biologists take a sample of, say, plankton in a lake, and they look at the respective size of the cells. Then they perform a preprocessing, by, say a kernel estimator. This is N_ϵ. (probably more approximation than that).
- Analytical point of view: N_ϵ is a noisy version of N, less regular than N (it is likely that no derivative exists) and $\|N - N_\epsilon\|_2 \leq \epsilon$. (see Perthame, Zubelli, etc)

The inverse problem

Under the previous differential equation, we consider the inverse problem of finding B given a "noisy" version of N.

- Practical: biologists take a sample of, say, plankton in a lake, and they look at the respective size of the cells. Then they perform a preprocessing, by, say a kernel estimator. This is N_ϵ. (probably more approximation than that).
- Analytical point of view: N_ϵ is a noisy version of N, less regular than N (it is likely that no derivative exists) and $\|N - N_\epsilon\|_2 \leq \epsilon$. (see Perthame, Zubelli, etc)
- Statistical point of view: we observe a n-sample $X_1, ..., X_n$ of iid variables with density N.

Pro and Con

Analytical point of view

Pro: taking into account maybe more approximations (but not all), results true for any N_ϵ.

Pro and Con

Analytical point of view

Pro: taking into account maybe more approximations (but not all), results true for any N_ϵ.
Con: N_ϵ is probably differentiable. If there are numerical methods which adapt to the regularity of N (discrepancy principle), they need to know ϵ.

Pro and Con

Analytical point of view

Pro: taking into account maybe more approximations (but not all), results true for any N_ϵ.
Con: N_ϵ is probably differentiable. If there are numerical methods which adapt to the regularity of N (discrepancy principle), they need to know ϵ.

Statistical point of view

Pro: Framework close to what biologists do, true inverse problem. We can adapt to the regularity, noise is given by the sample size.

Pro and Con

Analytical point of view

Pro: taking into account maybe more approximations (but not all), results true for any N_ϵ.
Con: N_ϵ is probably differentiable. If there are numerical methods which adapt to the regularity of N (discrepancy principle), they need to know ϵ.

Statistical point of view

Pro: Framework close to what biologists do, true inverse problem. We can adapt to the regularity, noise is given by the sample size.
Con: We only take one approximation into account and assume that we have access to the sample. Results true in expectation.

Pro and Con

Analytical point of view

Pro: taking into account maybe more approximations (but not all), results true for any N_ϵ.
Con: N_ϵ is probably differentiable. If there are numerical methods which adapt to the regularity of N (discrepancy principle), they need to know ϵ.

Statistical point of view

Pro: Framework close to what biologists do, true inverse problem. We can adapt to the regularity, noise is given by the sample size.
Con: We only take one approximation into account and assume that we have access to the sample. Results true in expectation.

The statistical problem

In the previous example(s) and more generally in inverse problems through PDE based on densities, we need "most of the time" to find

- a density estimate

The statistical problem

In the previous example(s) and more generally in inverse problems through PDE based on densities, we need "most of the time" to find

- a density estimate
- an estimate of the (nth) derivative of this density

The statistical problem

In the previous example(s) and more generally in inverse problems through PDE based on densities, we need "most of the time" to find

- a density estimate
- an estimate of the (nth) derivative of this density
- in an L^p (usually L^2) sense

To do so, we observe "a n sample", ie iid variables....

The statistical problem

In the previous example(s) and more generally in inverse problems through PDE based on densities, we need "most of the time" to find

- a density estimate
- an estimate of the (nth) derivative of this density
- in an L^p (usually L^2) sense

To do so, we observe "a n sample", ie iid variables....
At the end, I will mention "other" possible settings

1 Classical methods

1. Classical methods

2. Adaptive methods

How to estimate a density N?

- Kernel methods: the closest to filtering methods.

How to estimate a density N?

- Kernel methods: the closest to filtering methods.
- Projection methods: histogram, wavelet, Fourier basis.

How to estimate a density N?

- Kernel methods: the closest to filtering methods.
- Projection methods: histogram, wavelet, Fourier basis.
- Others ....

Kernel methods

Given K a kernel (L^1, symmetric), we set $K_h(x) = \frac{1}{h} K\left(\frac{x}{h}\right)$ and

$$\hat{N}_h(x) := \frac{1}{n} \sum_{i=1}^{n} K_h(x - X_i)$$

Kernel methods

Given K a kernel (L^1, symmetric), we set $K_h(x) = \frac{1}{h}K\left(\frac{x}{h}\right)$ and

$$\hat{N}_h(x) := \frac{1}{n}\sum_{i=1}^{n} K_h(x - X_i)$$

Bias-Variance decomposition

$$\mathbb{E}\left[\left\|N - \hat{N}_h\right\|_2\right] \leq \|N - K_h \star N\|_2 + \frac{1}{\sqrt{nh}}\|K\|_2,$$

where $K_h \star N = \mathbb{E}(\hat{N}_h)$

Kernel methods

Given K a kernel (L^1, symmetric), we set $K_h(x) = \frac{1}{h}K\left(\frac{x}{h}\right)$ and

$$\hat{N}_h(x) := \frac{1}{n}\sum_{i=1}^{n} K_h(x - X_i)$$

Bias-Variance decomposition

$$\mathbb{E}\left[\left\|N - \hat{N}_h\right\|_2\right] \leq \|N - K_h \star N\|_2 + \frac{1}{\sqrt{nh}}\|K\|_2,$$

where $K_h \star N = \mathbb{E}(\hat{N}_h)$

Advantages : defined on $\mathbb{R}$ ($\mathbb{R}^d$), $\int \hat{N}_h = 1$ and if K positive, true density

Kernel methods

Given K a kernel (L^1, symmetric), we set $K_h(x) = \frac{1}{h} K\left(\frac{x}{h}\right)$ and

$$\hat{N}_h(x) := \frac{1}{n} \sum_{i=1}^{n} K_h(x - X_i)$$

Bias-Variance decomposition

$$\mathbb{E}\left[\left\| N - \hat{N}_h \right\|_2\right] \leq \| N - K_h \star N \|_2 + \frac{1}{\sqrt{nh}} \| K \|_2,$$

where $K_h \star N = \mathbb{E}(\hat{N}_h)$

Advantages : defined on $\mathbb{R}$ ($\mathbb{R}^d$), $\int \hat{N}_h = 1$ and if K positive, true density

Problem : find a good h.

Projection methods

Let $\Phi = \{\phi\}$ be an orthonormal family for L^2 (wavelet, Fourier if on segment etc).

Projection methods

Let $\Phi = \{\phi\}$ be an orthonormal family for L^2 (wavelet, Fourier if on segment etc).

$$\hat{N}_\Phi := \sum_{\phi \in \Phi} \hat{\beta}_\phi \phi \text{ with } \hat{\beta}_\phi := \frac{1}{n} \sum_{i=1}^{n} \phi(X_i).$$

Bias-Variance decomposition

$$\mathbb{E}\left[\left\|N - \hat{N}_\Phi\right\|_2^2\right] = \|N - \Pi_\Phi(N)\|_2^2 + \frac{1}{n} \sum_{\phi \in \Phi} \mathrm{Var}(\hat{\beta}_\phi),$$

where $\Pi_\Phi(N)$ the orthogonal projection of N on $\mathrm{Vect}\{\Phi\}$.

Projection methods

Let $\Phi = \{\phi\}$ be an orthonormal family for L^2 (wavelet, Fourier if on segment etc).

$$\hat{N}_\Phi := \sum_{\phi\in\Phi} \hat{\beta}_\phi \phi \text{ with } \hat{\beta}_\phi := \frac{1}{n}\sum_{i=1}^{n} \phi(X_i).$$

Bias-Variance decomposition

$$\mathbb{E}\left[\left\|N - \hat{N}_\Phi\right\|_2^2\right] = \|N - \Pi_\Phi(N)\|_2^2 + \frac{1}{n}\sum_{\phi\in\Phi} \mathrm{Var}(\hat{\beta}_\phi),$$

where $\Pi_\Phi(N)$ the orthogonal projection of N on $\mathrm{Vect}\{\Phi\}$.

$\frac{1}{n}\sum_{\phi\in\Phi} \mathrm{Var}(\hat{\beta}_\phi) \leq \frac{|\Phi|}{n} \sup_{\mathbb{R}}(N)$

Projection methods

Let $\Phi = \{\phi\}$ be an orthonormal family for L^2 (wavelet, Fourier if on segment etc).

$$\hat{N}_\Phi := \sum_{\phi \in \Phi} \hat{\beta}_\phi \phi \text{ with } \hat{\beta}_\phi := \frac{1}{n} \sum_{i=1}^{n} \phi(X_i).$$

Bias-Variance decomposition

$$\mathbb{E}\left[\left\| N - \hat{N}_\Phi \right\|_2^2\right] = \| N - \Pi_\Phi(N) \|_2^2 + \frac{1}{n} \sum_{\phi \in \Phi} \mathrm{Var}(\hat{\beta}_\phi),$$

where $\Pi_\Phi(N)$ the orthogonal projection of N on $\mathrm{Vect}\{\Phi\}$.

$\frac{1}{n} \sum_{\phi \in \Phi} \mathrm{Var}(\hat{\beta}_\phi) \leq \frac{|\Phi|}{n} \sup_{\mathbb{R}}(N)$

$|\Phi|$ (dimension) $\leftrightarrow 1/h$ (in $\mathbb{R}^d$, $1/h^d$).

But the "variance term" here depends on $N \rightarrow$ find a good Φ !

Projection methods

Let $\Phi = \{\phi\}$ be an orthonormal family for L^2 (wavelet, Fourier if on segment etc).

$$\hat{N}_\Phi := \sum_{\phi \in \Phi} \hat{\beta}_\phi \phi \text{ with } \hat{\beta}_\phi := \frac{1}{n} \sum_{i=1}^{n} \phi(X_i).$$

Bias-Variance decomposition

$$\mathbb{E}\left[\left\| N - \hat{N}_\Phi \right\|_2^2\right] = \| N - \Pi_\Phi(N) \|_2^2 + \frac{1}{n} \sum_{\phi \in \Phi} \mathrm{Var}(\hat{\beta}_\phi),$$

where $\Pi_\Phi(N)$ the orthogonal projection of N on $\mathrm{Vect}\{\Phi\}$.

$\frac{1}{n} \sum_{\phi \in \Phi} \mathrm{Var}(\hat{\beta}_\phi) \leq \frac{|\Phi|}{n} \sup_{\mathbb{R}}(N)$
$|\Phi|$ (dimension) $\leftrightarrow 1/h$ (in $\mathbb{R}^d$, $1/h^d$).
But the "variance term" here depends on $N \rightarrow$ find a good Φ !
Also problem when infinite family $\rightarrow$ usually finite support.

The "old" Lepski's method (1)

Monotonicity

If

- K has m vanishing moments, $m \geq s$
- N is with regularity s (Hölder, Sobolev, ...)

then

- Bias : $\|N - K_h \star N\|_2 \leq Ch^s$ increases with h, C depends on Hölder norm of N and K
- Variance : $C(nh)^{-1/2}$ decreases with h.

Hence optimum in $h_s \simeq n^{-\frac{1}{2s+1}}$ and optimal (minimax) rate in $\phi(s) = n^{-\frac{s}{2s+1}}$.

The "old" Lepski's method (2)

- family of $\mathcal{H} = \{h_k = h_{s_k}\}$ for $s_k = a + k(lnn)^{-1} \in [a, b]$ $(m > s)$
- If $l < k$, then $\|K_{h_k} \star N - K_{h_l} \star N\|_2 \leq \Box\phi(h_l)$

Hence

The "old Lepski's" method

$$\hat{k} = \max\{k \geq 0/\forall l < k, \ \|\hat{N}_{h_k} - \hat{N}_{h_l}\| \leq C\phi(h_l)\}$$

If C good (and generally depends on N) and if N is of regularity s_{k_0} (unknown to the user) then rate in $\phi(h_{k_0})$. (adaptivity in the minimax sense).
Remark : numerous variants .... (see Lepski, Spokoiny, 97 etc ...)

The "old" Lepski's method (3)

Problems :

- Procedure not data driven
- only aim is rate : purely asymptotic point of view, no "oracle" inequality, nothing said if K has not enough vanishing moments (for instance K positive).
- What if no monotonocity ? what if choice on K too ?

Model selection

Family of Φ and want to choose.

- Least-square contrast : $\gamma(f) = -2/n \sum_{i=1}^{n} f(X_i) + \int f^2$ also log likelihood...
- Penalized criterion : $\gamma(\hat{N}_\Phi) + \mathrm{pen}(\Phi)$ to minimize on the family

Remarks :

- classically on bounded support : best Willett and Nowak method (2007, penalized log likelihood + cart + piecewise polynomial)
- Estimation of the variance also possible, oracle inequalities available.
- Estimate classically non positive $\rightarrow$ clipped version
- Time consuming (except WN)

Thresholding rules

ONB $\{\phi_\lambda \lambda \in \Lambda\}$

- $\hat{N} = \sum_{\lambda \in \Gamma} \hat{\beta}_\lambda \mathbf{1}_{|\hat{\beta}_\lambda| \geq t} \phi_\lambda$
- same thing as Model selection with $\Phi \subset \Gamma$ and $\mathrm{pen}(\Phi) = |\Phi| t^2$
- easy to compute
- Version on $\mathbb{R}$! (Reynaud-Bouret, Rivoirard, Tuleau-Malot 2011), Oracle inequalities etc ...
- Still if you want positivity, it is not very smooth (either Haar/ histograms or clipping)

Goldenshluger and Lepski's method

Set for any x and any $h, h' > 0$,

$$\hat{N}_{h,h'}(x) := (K_h \star \hat{N}_{h'})(x) = \frac{1}{n}\sum_{i=1}^{n}(K_h \star K_{h'})(x - X_i),$$

Goldenshluger and Lepski's method

Set for any x and any $h, h' > 0$,

$$\hat{N}_{h,h'}(x) := (K_h \star \hat{N}_{h'})(x) = \frac{1}{n}\sum_{i=1}^{n}(K_h \star K_{h'})(x - X_i),$$

"Estimator" of the bias term

$$A(h) := \sup_{h' \in \mathcal{H}} \left\{ \|\hat{N}_{h,h'} - \hat{N}_{h'}\|_2 - \frac{\chi}{\sqrt{nh'}}\|K\|_2 \right\}_+$$

where, given $\varepsilon > 0$, $\chi := (1+\varepsilon)(1+\|K\|_1)$.

Goldenshluger and Lepski's method

Set for any x and any $h, h' > 0$,

$$\hat{N}_{h,h'}(x) := (K_h \star \hat{N}_{h'})(x) = \frac{1}{n}\sum_{i=1}^{n}(K_h \star K_{h'})(x - X_i),$$

"Estimator" of the bias term

$$A(h) := \sup_{h' \in \mathcal{H}} \left\{ \|\hat{N}_{h,h'} - \hat{N}_{h'}\|_2 - \frac{\chi}{\sqrt{nh'}}\|K\|_2 \right\}_+$$

where, given $\varepsilon > 0$, $\chi := (1+\varepsilon)(1+\|K\|_1)$.

$$\hat{h} := \arg\min_{h \in \mathcal{H}} \left\{ A(h) + \frac{\chi}{\sqrt{nh}}\|K\|_2 \right\} \quad \text{and} \quad \hat{N} := \hat{N}_{\hat{h}}.$$

...Uniform bounds ...

GL's oracle inequality

Oracle inequality

If $\mathcal{H} = \{1/\ell \ / \ \ell = 1, ..., \ell_{max}\}$ and if $\ell_{max} = \delta n$, if moreover $\|N\|_\infty < \infty$,
then for any $q \geq 1$,

$$\mathbb{E}\left(\|\hat{N} - N\|_2^{2q}\right) \leq \square_q \chi^{2q} \inf_{h\in\mathcal{H}} \left\{\|K_h \star N - N\|_2^{2q} + \frac{\|K\|_2^{2q}}{(hn)^q}\right\} + \square_{q,\varepsilon,\delta,\|K\|_2,\|K\|_1,\|N\|_\infty} \frac{1}{n^q}.$$

Remark : toy version. One can do it in higher dimension, choose the bandwidth according to direction, choose (under assumptions) the kernel, continuum of bandwidths etc (see the three recent papers of Goldenshluger and Lepski)

More ad hoc rules that work remarkably well in practice

- Silverman 86 : either assume it is "almost gaussian" or cross validation (see also V-fold cross-validation Arlot, Lerasle work in progress)
- Abramson 82 : for point wise estimation $h(x) \sim N(x)^{-1/2}$ or other formula .... See also Giné and Sang (09).
- Sain et Scott (96) bandwidth moved locally ... Based on cross-validation ...

What exists ?

Most of it in white noise models (but equivalence possible),

- Possible to estimate simultaneously a signal and its derivative, by the derivatives of the estimate. Use of Fourier transform (Hall Patil 95, Efromovich 98). Nothing adaptive as far as I know. on a finite interval !
- Local polynomials : Estimate in one point x_0 the curve by local polynomials. Coefficients of higher order estimate the derivatives. Possibility to do adaptation (Fan Gijbels 95, Spokoiny 98). Need to find a bandwidth in an adaptive way, see also ad hoc Lepski's method.
- Wavelet approaches via inverse problems : Abramovich Silverman (98, thresholding), Cai (02, block thresholding) on a finite interval !

Estimation of $D = \frac{\partial}{\partial x}\big(g(x)N(x)\big)$

If K is differentiable, $\int K = 1$ and $\int |K'|^2 < \infty$.

$$\hat{D}_h(x) := \frac{1}{n}\sum_{i=1}^{n} g(X_i)K_h'(x - X_i)$$

Estimation of $D = \frac{\partial}{\partial x}\big(g(x)N(x)\big)$

If K is differentiable, $\int K = 1$ and $\int |K'|^2 < \infty$.

$$\hat{D}_h(x) := \frac{1}{n}\sum_{i=1}^{n} g(X_i)K_h'(x - X_i)$$

Estimation of $D = \frac{\partial}{\partial x}\big(g(x)N(x)\big)$

If K is differentiable, $\int K = 1$ and $\int |K'|^2 < \infty$.

$$\hat{D}_h(x) := \frac{1}{n}\sum_{i=1}^{n} g(X_i)K_h'(x - X_i)$$

Bias-Variance decomposition:

$$\mathbb{E}\left(\left\|D - \hat{D}_h\right\|_2\right) \le \|D - K_h \star D\|_2 + \frac{1}{\sqrt{nh^3}}\|g\|_\infty \|K'\|_2.$$

GL's trick

$\hat{D}_{h,h'}(x) := \frac{1}{n}\sum_{i=1}^{n} g(X_i)(K_h \star K_{h'})'(x - X_i)$,

$$\tilde{A}(h) := \sup_{h' \in \tilde{\mathcal{H}}} \left\{ \|\hat{D}_{h,h'} - \hat{D}_{h'}\|_2 - \frac{\tilde{\chi}}{\sqrt{nh'^3}}\|g\|_\infty \|K'\|_2 \right\}_+ ,$$

where, given $\tilde{\varepsilon} > 0$, $\tilde{\chi} := (1 + \tilde{\varepsilon})(1 + \|K\|_1)$.

Estimation of $D = \frac{\partial}{\partial x}\big(g(x)N(x)\big)$

If K is differentiable, $\int K = 1$ and $\int |K'|^2 < \infty$.

$$\hat{D}_h(x) := \frac{1}{n}\sum_{i=1}^{n} g(X_i)K'_h(x - X_i)$$

GL's trick

$\hat{D}_{h,h'}(x) := \frac{1}{n}\sum_{i=1}^{n} g(X_i)(K_h \star K_{h'})'(x - X_i)$,

$$\tilde{A}(h) := \sup_{h' \in \tilde{\mathcal{H}}} \left\{ \|\hat{D}_{h,h'} - \hat{D}_{h'}\|_2 - \frac{\tilde{\chi}}{\sqrt{nh'^3}}\|g\|_\infty \|K'\|_2 \right\}_+ ,$$

where, given $\tilde{\varepsilon} > 0$, $\tilde{\chi} := (1 + \tilde{\varepsilon})(1 + \|K\|_1)$.

Finally, we estimate D by using $\hat{D} := \hat{D}_{\tilde{h}}$ with

$$\tilde{h} := \mathrm{argmin}_{h \in \tilde{\mathcal{H}}} \left\{ \tilde{A}(h) + \frac{\tilde{\chi}}{\sqrt{nh^3}}\|g\|_\infty \|K'\|_2 \right\}.$$

Result for the derivative D

Oracle inequality for D

If $\tilde{\mathcal{H}} = \{1/\ell \ / \ \ell = 1, \dots, \ell_{max}\}$ and if $\ell_{max} = \sqrt{\delta' n}$, if moreover $\|N\|_\infty$ and $\|g\|_\infty < \infty$, then for any $q \geq 1$,

$$\mathbb{E}\left(\|\hat{D} - D\|_2^{2q}\right) \leq \square_q \tilde{\chi}^{2q} \inf_{h \in \tilde{\mathcal{H}}} \left\{ \|K_h \star D - D\|_2^{2q} + \left[\frac{\|g\|_\infty \|K'\|_2}{\sqrt{nh^3}}\right]^{2q} \right\}$$

$$+\square_{q,\tilde{\varepsilon},\delta',\|K'\|_2,\|K\|_1,\|K'\|_1,\|N\|_\infty,\|g\|_\infty} \frac{1}{n^q}.$$

The informal problem and the PDE translation for size-structured population

- A cell grows.
- Depending on its size x, the cell has a certain chance to divide itself in 2 offsprings, ie 2 cells of size $x/2$.
- We are interesting by the evolution of the whole population of cells, each of them having this behavior.

The informal problem and the PDE translation for size-structured population

- A cell grows.
- Depending on its size x, the cell has a certain chance to divide itself in 2 offsprings, ie 2 cells of size $x/2$.
- We are interesting by the evolution of the whole population of cells, each of them having this behavior.

Size-Structured Population Equation (finite time)

$$\begin{cases} \frac{\partial}{\partial t}(n(t,x)) + \kappa \frac{\partial}{\partial x}(g(x)n(t,x)) + B(x)n(t,x) = 4B(2x)n(t,2x), \\ n(t,x=0) = 0, \quad t > 0 \\ n(0,x) = n_0(x), \quad x \geq 0. \end{cases}$$

- $n(t,x)$ the "amount" of cells with size x ($\neq$ density),
- g the "qualitative" growth rate of one cell: linear is $g = 1$ ...
- B is the division rate, which depends on the size

Asymptotics of the PDE

It can be shown (Perthame Ryzhik 2005 for instance) that

- $n(t,.)$ grows exponentially fast ie $I_t = \int n(t,x)dx$ asymptotically proportional to $e^{\lambda t}$,
- the renormalized $n(t,x)/I_t$ tends to a density N, which satisfies

Size-Structured Population Equation (asymptotics)

$$\begin{cases} \kappa \dfrac{\partial}{\partial x}(g(x)N(x)) + \lambda N(x) = \mathcal{L}(BN)(x), \\ B(0)N(0) = 0, \qquad \int N(x)dx = 1, \end{cases}$$

where N step D step κ step L step H step B step

- for any real-valued function $x \rightsquigarrow \varphi(x)$, $\mathcal{L}(\varphi)(x) := 4\varphi(2x) - \varphi(x)$.
- $\kappa = \lambda \dfrac{\int_{\mathbb{R}_+} xN(x)dx}{\int_{\mathbb{R}_+} g(x)N(x)dx}$.

Estimation of λ and κ

SSPE

Estimation of λ and κ

SSPE λ is estimated via another (or simultaneous experiment).

Estimation of λ and κ

SSPE λ is estimated via another (or simultaneous experiment).

Assumption on $\hat{\lambda}$

There exist some $q > 1$ such that

- $\varepsilon_\lambda = \mathbb{E}[|\sqrt{n}(\hat{\lambda} - \lambda)|^q] < \infty$,
- $R_\lambda = \mathbb{E}(\hat{\lambda}^{2q}) < \infty$.

Estimation of λ and κ

SSPE λ is estimated via another (or simultaneous experiment).

Assumption on $\hat{\lambda}$

There exist some $q > 1$ such that

- $\varepsilon_\lambda = \mathbb{E}[|\sqrt{n}(\hat{\lambda} - \lambda)|^q] < \infty$,
- $R_\lambda = \mathbb{E}(\hat{\lambda}^{2q}) < \infty$.

Let $c > 0$,

$$\hat{\kappa} = \hat{\lambda} \frac{\sum_{i=1}^n X_i}{\sum_{i=1}^n g(X_i) + c}.$$

Oracle inequality for the estimation of $H = BN$

We establish an oracle inequality for $H = BN$ which is true under all previous assumptions.

Theorem

$$\mathbb{E}\left[\left\|\hat{H} - H\right\|_{2,T}^{q}\right] \leq C\left\{E_D + E_N + E_\lambda + E_{\mathcal{L}} + n^{-\frac{q}{2}}\right\}$$

Oracle inequality for the estimation of $H = BN$

We establish an oracle inequality for $H = BN$ which is true under all previous assumptions.

Theorem

$$\mathbb{E}\left[\left\|\hat{H} - H\right\|_{2,T}^{q}\right] \leq C\left\{E_D + E_N + E_\lambda + E_{\mathcal{L}} + n^{-\frac{q}{2}}\right\}$$

with

- $E_D = \sqrt{R_\lambda} \inf_{h \in \tilde{\mathcal{H}}} \left\{\|K_h \star D - D\|_2^q + \left(\frac{\|g\|_\infty \|K'\|_2}{\sqrt{nh^3}}\right)^q\right\}$

Oracle inequality for the estimation of $H = BN$

We establish an oracle inequality for $H = BN$ which is true under all previous assumptions.

Theorem

$$\mathbb{E}\left[\left\|\hat{H} - H\right\|_{2,T}^{q}\right] \leq C\left\{E_D + E_N + E_\lambda + E_{\mathcal{L}} + n^{-\frac{q}{2}}\right\}$$

with

- $E_D = \sqrt{R_\lambda}\inf_{h\in\tilde{\mathcal{H}}}\left\{\|K_h \star D - D\|_2^q + \left(\frac{\|g\|_\infty \|K'\|_2}{\sqrt{nh^3}}\right)^q\right\}$
- $E_N = \inf_{h\in\mathcal{H}}\left\{\|K_h \star N - N\|_2^q + \left(\frac{\|K\|_2}{\sqrt{nh}}\right)^q\right\}$

Oracle inequality for the estimation of $H = BN$

We establish an oracle inequality for $H = BN$ which is true under all previous assumptions.

Theorem

$$\mathbb{E}\left[\left\|\hat{H} - H\right\|_{2,T}^{q}\right] \leq C\left\{E_D + E_N + E_\lambda + E_{\mathcal{L}} + n^{-\frac{q}{2}}\right\}$$

with

- $E_D = \sqrt{R_\lambda} \inf_{h \in \tilde{\mathcal{H}}} \left\{\|K_h \star D - D\|_2^q + \left(\frac{\|g\|_\infty \|K'\|_2}{\sqrt{nh^3}}\right)^q\right\}$
- $E_N = \inf_{h \in \mathcal{H}} \left\{\|K_h \star N - N\|_2^q + \left(\frac{\|K\|_2}{\sqrt{nh}}\right)^q\right\}$
- $E_\lambda = \varepsilon_\lambda n^{-\frac{q}{2}}$

Oracle inequality for the estimation of $H = BN$

We establish an oracle inequality for $H = BN$ which is true under all previous assumptions.

Theorem

$$\mathbb{E}\left[\left\|\hat{H} - H\right\|_{2,T}^{q}\right] \leq C\left\{E_D + E_N + E_\lambda + E_{\mathcal{L}} + n^{-\frac{q}{2}}\right\}$$

with

- $E_D = \sqrt{R_\lambda}\inf_{h\in\tilde{\mathcal{H}}}\left\{\|K_h \star D - D\|_2^q + \left(\frac{\|g\|_\infty \|K'\|_2}{\sqrt{nh^3}}\right)^q\right\}$
- $E_N = \inf_{h\in\mathcal{H}}\left\{\|K_h \star N - N\|_2^q + \left(\frac{\|K\|_2}{\sqrt{nh}}\right)^q\right\}$
- $E_\lambda = \varepsilon_\lambda n^{-\frac{q}{2}}$
- $E_{\mathcal{L}} = \left((\|N\|_{\mathcal{W}^1} + \|gN\|_{\mathcal{W}^2})\frac{T}{\sqrt{k}}\right)^q$

Rate of convergence for the estimation of B

here We finally set $\hat{B} = \hat{H}/\hat{N}$ and $\tilde{B} = \max(\min(\hat{B}, \sqrt{n}), -\sqrt{n})$.

Rate of convergence for the estimation of B

here We finally set $\hat{B} = \hat{H}/\hat{N}$ and $\tilde{B} = \max(\min(\hat{B}, \sqrt{n}), -\sqrt{n})$.
If $B \in \mathcal{W}_s$ $(s > 1/2)$ and $g \in \mathcal{W}_{s+1}$, then (under suitable assumptions and enough moments for the kernel) $N \in \mathcal{W}_{s+1}$.

Rate of convergence for the estimation of B

here We finally set $\hat{B} = \hat{H}/\hat{N}$ and $\tilde{B} = \max(\min(\hat{B}, \sqrt{n}), -\sqrt{n})$.
If $B \in \mathcal{W}_s$ ($s > 1/2$) and $g \in \mathcal{W}_{s+1}$, then (under suitable assumptions and enough moments for the kernel) $N \in \mathcal{W}_{s+1}$.

Theorem

one can choose a family of $\mathcal{H}$ and $\mathcal{H}'$ independent of s such that for any compact $[a, b]$ of $[0, T]$ (under technical assumptions),

$$\mathbb{E}\left[\left\|(\tilde{B} - B)1_{[a,b]}\right\|_2^q\right] = O\left(n^{-\frac{qs}{2s+3}}\right).$$

Why is it the good rate?(1)

In the deterministic set-up

- we observe $N_\epsilon = N + \epsilon\zeta$, with $\|\zeta\|_2 \leq 1$ and

$$BN = \mathcal{L}^{-1}\left(\kappa\partial_x\big(g(x)N(x)\big) + \lambda N(x)\right).$$

- Since $\mathcal{L}^{-1}$ is continuous and the recovery of $\partial_x N$ is a more difficult inverse problem than the recovery of N, hence the ill-posedness is only due to ∂N (degree of ill-posedness $=1$)
- Hence if $N \in \mathcal{W}^s$, error in $\epsilon^{\frac{s}{s+1}}$.

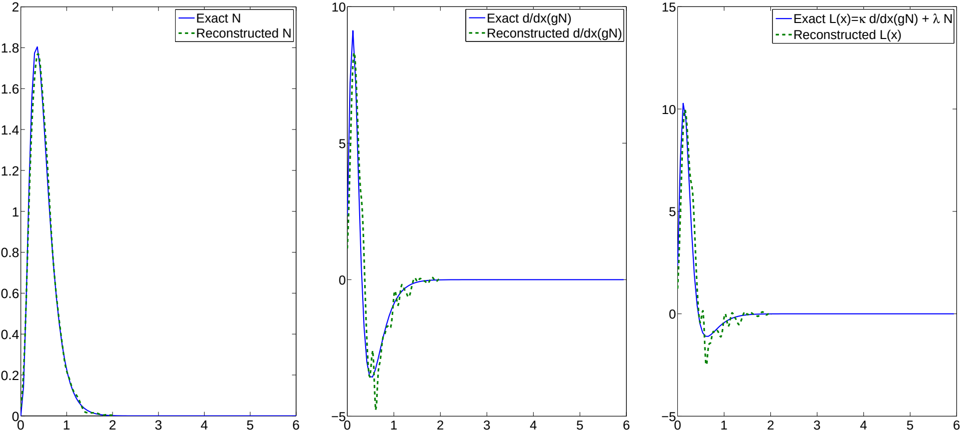
Why is it the good rate?(2)

In the n-sample set-up

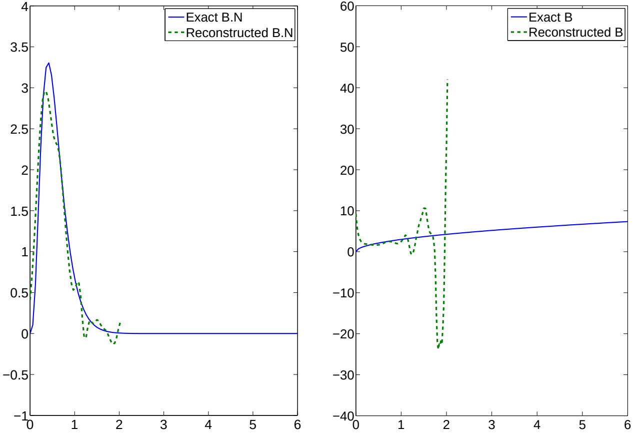
- problem well approximated by $N_\epsilon = N + \epsilon\mathbb{B}$ with $\mathbb{B}$ Gaussian white noise and $\epsilon = n^{-1/2}$.
- $\mathbb{B}$ is not in $\mathbb{L}_2$ but in $\mathcal{W}^{-1/2}$,
- Hence one needs to integrate ie $Z_\epsilon = \mathcal{I}^{1/2}N + \epsilon\mathcal{I}^{1/2}\mathbb{B}$ to have a noise in $\mathbb{L}_2$.
- Hence $Z_\epsilon = \mathcal{I}^{3/2}(\partial N) + \epsilon\mathcal{I}^{1/2}\mathbb{B}$ is of degree of ill-posedness 3/2.
- Hence if $N \in \mathcal{W}^s$, error in $\epsilon^{\frac{s}{s+3/2}} = n^{-\frac{s}{2s+3}}$.

Simulations

n=5000, Gaussian kernel, $B = 3\sqrt{x}$, $g = 1$.



Simulations



What if data not iid ?

- data = all the times of division + all the sizes : work in progress Doumic, Hoffmann, Krell etc : Kernel possible, no adaptation
- data = irreducible stationary Markov chain : Claire Lacour (and co) adaptive estimate of stationary density and transition density (on finite interval)
- An analogue to Talagrand for Markov chain : Adamczak 08
- Chaos propagation and control ?
- Berbee's lemma, mixing properties and being almost independent ?

For this particular PDE problem

- Calibration and numerical optimization of the GL's method

For this particular PDE problem

- Calibration and numerical optimization of the GL's method
- To take into account noise in the measurements: Replace observations X_i with $X_i + Z_i$

For this particular PDE problem

- Calibration and numerical optimization of the GL's method
- To take into account noise in the measurements: Replace observations X_i with $X_i + Z_i$
- Extensions to fit with a more realistic biological model:

For this particular PDE problem

- Calibration and numerical optimization of the GL's method
- To take into account noise in the measurements: Replace observations X_i with $X_i + Z_i$
- Extensions to fit with a more realistic biological model:
  - The division law is given by a kernel $k(x, y)$:

$$... = 2\int_x^{\infty} B(y)k(x,y)n(t,y)dy - B(x)n(t,x),$$

Division of the cell of size y into 2 cells of size x and $y - x$ with probability density$=k(x,y)$. Equal mitosis: $k(x,y) = \delta_{x=\frac{y}{2}}$, so $2\int_x^{\infty} B(y)k(x,y)n(t,y)dy = 4B(2x)n(t,2x)$

For this particular PDE problem

- Calibration and numerical optimization of the GL's method
- To take into account noise in the measurements: Replace observations X_i with $X_i + Z_i$
- Extensions to fit with a more realistic biological model:
  - The division law is given by a kernel $k(x, y)$:

$$... = 2\int_x^{\infty} B(y)k(x,y)n(t,y)dy - B(x)n(t,x),$$

    Division of the cell of size y into 2 cells of size x and $y - x$ with probability density$=k(x, y)$. Equal mitosis: $k(x, y) = \delta_{x=\frac{y}{2}}$, so $2\int_x^{\infty} B(y)k(x,y)n(t,y)dy = 4B(2x)n(t,2x)$
  - Construct a microscopic stochastic system (PDMP) that matches with the PDE's approximation and that take advantage of richer observation schemes (Probabilistic works in progress studied by B. Cloez, V. Bansaye, M. Doumic, M. Hoffmann, N. Krell, T. Lepoutre, L. Robert,...)

References

Doumic, M. and Gabriel, P. (2010) *Eigenelements of a General Aggregation-Fragmentation Model*. Math. Models Methods Appl. Sci. 20(5), 757–783.

Doumic, M., Hoffmann, M., Reynaud-Bouret, P. and Rivoirard, V. (2011) Nonparametric estimation of the division rate of a size-structured population. To appear in SIAM J. Numer. Anal.

Doumic, M., Perthame, B. and Zubelli, J. (2009) *Numerical Solution of an Inverse Problem in Size-Structured Population Dynamics*. Inverse Problems, 25, 25pp.

Goldenshluger, A. and Lepski, O. (2009) *Uniform bounds for norms of sums of independent random functions* arXiv:0904.1950.

Goldenshluger, A. and Lepski, O. (2011) *Bandwidth selection in kernel density estimation: oracle inequalities and adaptive minimax optimality*. Ann. Statist. 39(3), 1608–1632.

Perthame, B. (2007) Transport equations in biology. In *Frontiers in Mathematics*, Frontiers in Mathematics. Birckhauser.

Perthame, B. and Ryzhik, L. (2005) *Exponential decay for the fragmentation or cell-division equation,* J. of Diff. Eqns, 210, 155–177 .

Perthame, B. and Zubelli, J. P. (2007) *On the inverse problem for a size-structured population model,* Inverse Problems, 23(3), 1037–1052.