Statistical inference for structured models

Part II: Example 3 (Human population models). Statistical methodology

Marc Hoffmann, Université Paris-Dauphine PSL

Spring school SFB 1294, March 2019

Today's program

- Example 3: Human population models
- Adaptive nonparametric estimation
  - Lepski's principle: soft heuristics
  - The Goldenshluger-Lepski method without pain $\rightsquigarrow$ afternoon discussion session/Lecture III

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) .

Example 3 (Lecture I continued): Human population models

Statistical experiments and methodology

Adaptive nonparametric estimation

Nonparametric estimation

Paradigmatic examples

1. Cell division: blue growth-fragmentation models
   - Age-structured models and the renewal equation
   - Size-structured models
2. General bifurcating models
3. Human population models for demography
   - Cohort effects in human mortality
   - Towards nonlinearity
4. Models of interacting neurons $\rightsquigarrow$ Lecture IV
   - Spikes models
   - Hawkes models
5. More nonlinear models in a mean-field limit $\rightsquigarrow$ Lecture IV

Motivation: improving mortality estimates

- Mortality table = mortality rates for several age classes (with length one or several years), at several periods in time (usually each year)
- Mortality tables $\rightsquigarrow$ age shape of mortality and dynamics over time

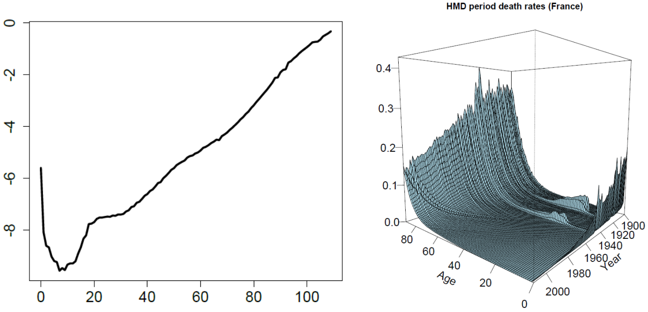

Figure: *Left:* Mortality rate, 2008, France, as a function of age (log-scale).*Right:* Mortality table by age and time

A (very) brief history of demographics

- The first mortality table appeared in 1662 by John Graunt $\rightsquigarrow$ estimation of death probabilities as a function of age.
- 1865: graphical formalizations of life trajectories within a population by Lexis and his contemporaries.

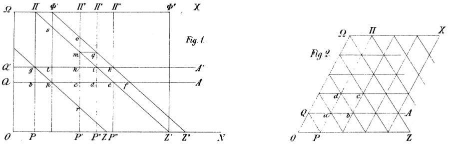

Figure: Examples of the so-called 'Lexis Diagram'

- The first demographers understood that it is crucial to (i) keep a non-homogeneous picture and (ii) the measurement of the mortality rate depends on an underlying population dynamics.

Recent awareness about anomalies

- Cohort effects have long fascinated demographers.

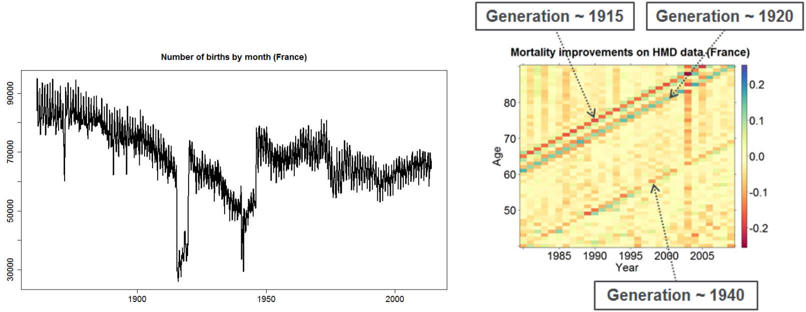

Figure: *Left:* monthly births in France. *Right:* artefact (?) cohort effects in mortality improvements from crude tables of the Human Mortality Database.

- Richards (2008) suggested that cohort effects could be artefacts caused by anomalies in the calculation of death rates due to shocks in birth patterns.

Recent awareness about anomalies

- Cairns *et al.* (2016) confirm Richards' conclusions with England and Wales data completed on monthly fertility data.
- Boumezoued (2016) aggregates the (HMD) database and the (HFD) database and suggests that these anomalies are universal.

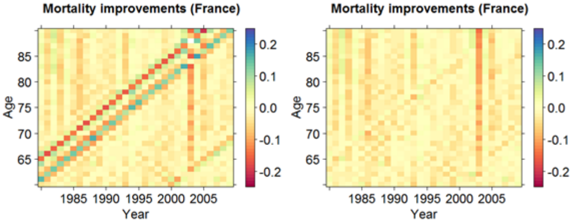

Figure: Mortality rates estimates before (*left*) and after (*right*) correction from Boumezoued (2016).

Example 3: identification of the objects of interest

- $F, B : \mathbb{R}_+ \times \mathbb{R}_+ \to \mathbb{R}_+$ model parameters.
- $F(t, a)$: fertility rate of the population with age a at time t.
- $B(t, a)$: mortality rate of the population with age a at time t.
- $Z_0 \in \mathcal{M}_F$: initial age distribution of the population at time $t = 0$.

Example 3: evolution equation

- $\big(A_i(t)\big)_{1\le i\le N_t}$ = all the ages of the population at time t.
- $Z_t = \sum_{i=1}^{N_t} \delta_{A_i(t)}$ with $N_t = \langle Z_t, \mathbf{1}\rangle$.
- Associated SDE

$$
\begin{aligned}
Z_t = \; & \tau_t Z_0 \\
& + \int_0^t \sum_{i\le\langle Z_{s-},\mathbf{1}\rangle} \int_{0\le\theta\le F(s,a_i(Z_{s-}))} \delta_{t-s} Q(ds,di,d\theta) \\
& - \int_0^t \sum_{i\le\langle Z_{s-},\mathbf{1}\rangle} \int_{0\le\theta\le B(s,a_i(Z_{s-}))} \delta_{a_i(Z_{s-})+t-s} \widetilde{Q}(ds,di,d\theta)
\end{aligned}
$$

- $Q, \widetilde{Q}$ independent Poisson random measures, intensity $ds\big(\sum_{k\ge1}\delta_k(di)\big)d\theta$.

Microscopic evolution equation

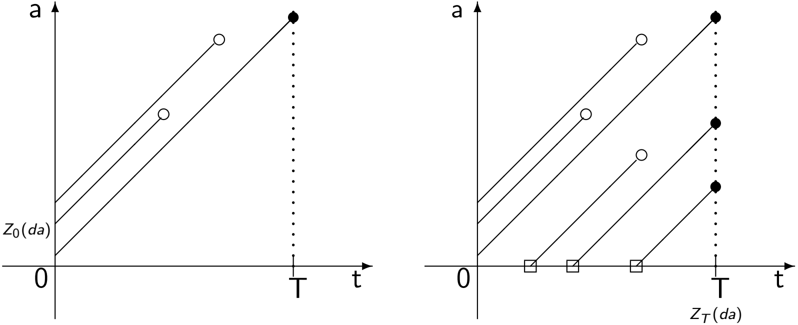

Figure: *Left*: Sample path of $Z_0(da)$ and its evolution without births.
Right: Sample path of $(Z_t(da), t \in [0, T])$.

Large population limit

- Large population limit approach: $Z_0 \rightsquigarrow Z_0(N)$, $N \geq 1$, with $\langle Z_0(N), \mathbf{1} \rangle \approx N$.
- N reminiscent of a (large) population size.
- Renormalisation: $Z_t \rightsquigarrow Z_t/N =: Z_t^N$ (thus $Z_0^N = N^{-1} Z_0(N)$) yields

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

Example 3: large population limit

- $N \to \infty$ abstract asymptotic parameter.
- Reminiscent of a population size : $\langle NZ_t^N, \mathbf{1} \rangle \approx N$ 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

$$\begin{cases} \frac{\partial}{\partial t} g(t,a) + \frac{\partial}{\partial a} g(t,a) + B(t,a) g(t,a) = 0, \\ g(0,a) = g_0(a), \quad g(t,0) = \int_{\mathbb{R}_+} F(t,a) g(t,a) da. \end{cases}$$

Example 3: large population limit

We can identify the following objects

- $N \to \infty$ is arbitrary, reminiscent of the population size $\langle Z_t^N, \mathbf{1} \rangle$ for every $t \in [0, T]$.
- Z^N is $(Z_t^N)_{0 \le t \le T}$ and we observe $\mathcal{Z}^N = Z^N$.
- f is any of the functions $(t, a) \mapsto g(t, a), F(t, a)$ or $B(t, a)$.
- $\mathcal{H}^N$ and $\mathcal{H}$ are the SDE and the McKendrick & Von Foerster equation.

From Z^N to a statistical experiment.

- We have a stochastic model $(Z_t)_{1\leq t\leq T}$, as a time evolving point measure where either
  - $\langle Z_T, \mathbf{1}\rangle$ is large when T is large, T deterministic or random (stopping time).
  - T is fixed but $Z_t = Z_t^N$ depends on a renormalisation parameter N and $\langle Z_t^N, \mathbf{1}\rangle$ is large for every t when N is large.
- $N \to \infty$ asymptotic parameter.
- We write Z^N for $(Z_t)_{0\leq t\leq T(N)}$ or $(Z_t^N)_{0\leq t\leq T}$.
- We extract from Z^N an observation $\mathcal{Z}^N$.

The experiment generated by $\mathcal{Z}^N$

- $\mathcal{Z}^N$ generates (a sequence of) statistical experiment

$$\left\{\mathbb{P}^N_{B,k}, B \in \mathcal{B}, k \in \mathcal{K}\right\}_{N \geq 1}$$

  B: parameter of interest, k nuisance parameter (possibly known, usually functional).
- $B : [0, \infty) \times [0, \infty) \to [0, \infty)$ belongs to a functional class.
- We need a methodology for recovering B non-parametrically.

Nonparametric estimation

- Experiment: $\mathcal{E}^N = \{\mathbb{P}^N_{B,k}, B \in \mathcal{B}, k \in \mathcal{K}\}$.
- Objective: recover $B(t,a)$ with $B \in \mathcal{B}$ from data $\mathcal{Z}^N$.
- $\widehat{B}^N(t,a) = \widehat{B}^N\big(\mathcal{Z}^N, (t,a)\big)$ estimator of $B(t,a)$.
- Reconstruction criterion

$$\boxed{\mathcal{R}^N(\widehat{B}^N(t,a), \mathcal{B}) = \sup_{B \in \mathcal{B}, k \in \mathcal{K}} \mathbb{E}^N_{B,k}\Big[\big(\widehat{B}^N(t,a) - B(t,a)\big)^2\Big]}$$

- $v_N \to 0$ is an admissible rate of of convergence for estimating $B(t,a)$ over $\mathcal{B}$ if there exists $\widehat{B}^N(t,a)$ such that

$$\sup_N v_N^{-2} \mathcal{R}^N(\widehat{B}^N(t,a), \mathcal{B}) < \infty.$$

Nonparametric estimation

- Sometimes, we only require the (weaker) tightness of

$$\left(v_N^{-1}\left(\widehat{B}^N(t,a) - B(t,a)\right)\right)_{N\geq 1},$$

uniformly in $B \in \mathcal{B}$, meaning

$$\sup_{B\in\mathcal{B},k\in\mathcal{K}} \mathbb{P}^N_{B,k}\left(v_N^{-1}\left|\widehat{B}^N(t,a) - B(t,a)\right| \geq K\right) \to 0, \quad K \to \infty.$$

- $\widehat{B}^N_\star(t,a)$ is minimax optimal if

$$\mathcal{R}^N(\widehat{B}^N(t,a),\mathcal{B}) \approx \inf_F \mathcal{R}^N(F,\mathcal{B}) \;\text{ as }\; N \to \infty,$$

infimum taken over all estimators F of $B(t,a)$ from $\mathcal{Z}^N$.

Example 3 (Lecture I continued): Human population models

Statistical experiments and methodology

Adaptive nonparametric estimation

Nonparametric estimation

Nonparametric estimation in density estimation

- Let us consider an apparently different problem: estimate a probability distribution $g(t,a)dtda$ from a (IID) drawn

$$\mathcal{Z}^N \leftrightarrow (T_1, A_1), \dots, (T_N, A_N).$$

- Statistical objective: pointwise estimation of $g(t,a)$.
- Assumption: $g \in L^\infty_{\mathrm{loc}} +$ local smoothness properties.
- Anisotropic Hölder space $\mathcal{H}^{\alpha,\beta}$:

$$g \in \mathcal{H}^{\alpha,\beta} \iff \begin{cases} t \mapsto g(t,a) \in \mathcal{H}^\alpha, & \forall a, \\ a \mapsto g(t,a) \in \mathcal{H}^\beta, & \forall t, \end{cases}$$

where $\mathcal{H}^s$ is the usual (univariate) Hölder space.
$\Big(x \mapsto f(x) \in \mathcal{H}^s, s = n + \{s\}, n \text{ integer}, 0 < \{s\} \leq 1$ iff
$\|f\|_{L^\infty} + \sup_{x,y} \frac{|f^{(n)}(y) - f^{(n)}(x)|}{|y-x|^{\{s\}}} < \infty.\Big)$

Preparation: anisotropic estimation

- Kernel reconstruction: Pick a smooth and compactly supported product kernel K

$$K(t,a) = K^{(1)}(t)K^{(2)}(a).$$

- L^1-normalisation: for $\boldsymbol{h} = (h_1, h_2), h_i > 0$:

$$K_{\boldsymbol{h}}(t,a) = (h_1 h_2)^{-1} K^{(1)}\big(h_1^{-1} t\big) K^{(2)}\big(h_2^{-2} a\big).$$

- Kernel estimation

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

where $\mathcal{Z}^N(ds,du) = N^{-1}\sum_{i=1}^N \delta_{(T_i,A_i)}(ds,du)$.

Nonparametric estimation in density estimation

- Error analysis: standard bias + variance decomposition.
- Bias analysis:

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

- Assume $g \in \mathcal{H}^{\alpha,\beta}$. Then

$$\boxed{\big|g(t,a) - g_{\boldsymbol{h}}(t,a)\big| \lesssim |g|_{\mathcal{H}^{\alpha,\beta}} \big(h_1^{\alpha\wedge(L+1)} + h_2^{\beta\wedge(L+1)}\big)}$$

$\big(L =$ order of the kernel: $\int x^\ell K(x)dx = \mathbf{1}_{\{\ell=0\}}$ for $\ell = 0,\ldots,L.\big)$

- Remark: different (equivalent) choices for $|g|_{\mathcal{H}^{\alpha,\beta}}$.

Nonparametric estimation in density estimation

- Variance analysis:

$$\mathrm{Var}\big(\widehat{g}_{N,\boldsymbol{h}}(t,a)\big) \leq N^{-1}\int_0^T\int_{\mathbb{R}_+} K_{\boldsymbol{h}}(t-s,a-u)^2 g(s,u)dsdu$$
$$\leq N^{-1}\|K_{\boldsymbol{h}}\|_{L^2}^2|g|_{L^\infty_{\mathrm{loc}}} = |K|_2^2|g|_{L^\infty_{\mathrm{loc}}} N^{-1}h_1^{-1}h_2^{-1}.$$

- Window optimisation $\boldsymbol{h}=\boldsymbol{h}^\star$ yields error bound

$$\boxed{\sup_g \mathbb{E}\Big[\big(\widehat{g}^N_{\boldsymbol{h}^\star}(t,a)-g(t,a)\big)^2\Big] \lesssim N^{-2s(\alpha,\beta)/(2s(\alpha,\beta)+1)}}$$

with effective smoothness

$$\frac{1}{s(\alpha,\beta)} = \frac{1}{\alpha}+\frac{1}{\beta}.$$

- Supremum over (local) Hölder balls, minimax optimality.

Towards adaptive estimation

We have established

$$\mathbb{E}\Big[\big(\widehat{g}^N_{\boldsymbol{h}^\star}(t,a) - g(t,a)\big)^2\Big] \lesssim \big(K_{h_1,h_2} \star g(t,a) - g(t,a)\big)^2 + \Big(\frac{1}{\sqrt{Nh_1 h_2}}\Big)^2$$
$$=: \mathbb{B}_{\boldsymbol{h}}(g) + \mathbb{V}^N_{\boldsymbol{h}}$$

- Oracle estimation: look for $\boldsymbol{h} = \widehat{h}(\mathcal{Z}^N)$ so that

$$\mathbb{E}\big[\big(\widehat{g}^N_{\boldsymbol{h}^\star}(t,a) - g(t,a)\big)^2\big] \lesssim \inf_{\boldsymbol{h}\in\mathcal{H}} \big(\mathbb{B}_{\boldsymbol{h}}(g) + \mathbb{V}^N_{\boldsymbol{h}}\big).$$

- Need $\mathcal{H}$ rich enough so that it can mimick the optimal bandwidth $\boldsymbol{h}^\star$ if $g \in \mathcal{H}^{\alpha,\beta}$.
- If (α,β) unknown $\rightsquigarrow$ adaptive estimation $\rightsquigarrow$ Lepski's principle.