Nonparametric estimation of the intensity function from indirect point process observations

Martin Kroll
Universität Mannheim

1. Model

We consider the nonparametric estimation of an intensity $\lambda : [0,1) \to [0,\infty)$ from samples

$$N_1, \dots, N_n \qquad \text{i.i.d.} \sim \mathcal{L}(N) \tag{1}$$

and

$$Y_1, \dots, Y_m \qquad \text{i.i.d.} \sim \mathcal{L}(Y) \tag{2}$$

for a random variable Y with *unknown* density function f and

$$N = \sum_i \delta_{x_i+\varepsilon_i-\lfloor x_i+\varepsilon_i \rfloor} \tag{3}$$

where $\widetilde{N} = \sum_i \delta_{x_i}$ is a Poisson point process (PPP) with intensity λ and ε_i i.i.d. $\sim f$.

- ▷ In analogy to deconvolution problems the considered model can be traced back to an indirect sequence space model and is an example of a *statistical inverse problem*.
- ▷ Due to the inpendence assumption on the errors ε_i in (3), the observed point processes $N_1, \dots, N_n$ are again Poisson with intensity

$$\ell(t) := \int_0^1 \lambda((t-\varepsilon) - \lfloor t-\varepsilon \rfloor) f(\varepsilon) d\varepsilon, \qquad t \in [0,1).$$

- ▷ The paper [Big+13] considers a similar model (with Cox instead of Poisson observations) and proposes a non-linear hard thresholding estimator in case of a *known* and *ordinary smooth* error density f.
- ▷ Our approach is mainly inspired by the one used in the article [JS13] which is adapted to deal with point process observations.

2. Methodology

The construction of our estimator is motivated by the Fourier series representation of $\lambda \in \mathbb{L}^2$ in terms of the *complex trigonometric basis* $\{\mathbf{e}_j\}_{j\in\mathbb{Z}}$ with $\mathbf{e}_j(t) = \exp(2\pi i j t)$:

$$\lambda = \sum_{j\in\mathbb{Z}} [\lambda]_j \mathbf{e}_j \qquad \text{where} \qquad [g]_j = \int_0^1 g(t)\mathbf{e}_j(-t)dt \quad \text{for} \quad g \in \mathbb{L}^2.$$

- ▷ By the convolution theorem, $[\ell]_j = [\lambda]_j \cdot [f]_j$ for all $j \in \mathbb{Z}$.
- ▷ Unbiased estimators of $[\ell]_j$ and $[f]_j$ are easily obtained:

$$\widehat{[\ell]}_j := \frac{1}{n}\sum_{i=1}^n \int_0^1 \mathbf{e}_j(-t) dN_i(t) \qquad \text{and} \qquad \widehat{[f]}_j := \frac{1}{m}\sum_{i=1}^m \mathbf{e}_j(-Y_i).$$

- ▷ We consider the orthogonal series estimator $\widehat{\lambda}_k$ defined via

$$\widehat{\lambda}_k := \sum_{0\le|j|\le k} \frac{\widehat{[\ell]}_j}{\widehat{[f]}_j} \mathbb{1}_{\Omega_j} \mathbf{e}_j,$$

where $\Omega_j = \{|\widehat{[f]}_j|^2 \ge m^{-1}\}$ and $k \in \mathbb{N}_0$ is a tuning parameter the choice of which crucially determines the performance of the estimator.

3. Minimax theory

We consider the $\mathbb{L}^2$-*ellipsoid*

$$\Lambda_\gamma^r := \{\lambda \in \mathbb{L}^2 : \lambda \ge 0 \text{ and } \sum_{j\in\mathbb{Z}} \gamma_j |[\lambda]_j|^2 \le r\}, \qquad r \ge 0,$$

and the $\mathbb{L}^2$-*hyperrectangle*

$$\mathcal{F}_\alpha^d := \{f \in \mathbb{L}^2 : f \ge 0, [f]_0 = 1 \text{ and } d^{-1} \le |[f]_j|^2 / \alpha_j \le d\}, \qquad d \ge 1.$$

The sequences $\{\gamma_j\}_{j\in\mathbb{Z}}$, $\{\alpha_j\}_{j\in\mathbb{Z}}$ are assumed to be strictly positive and symmetric. The decay of $\{\alpha_n\}_{n\in\mathbb{N}_0}$ towards 0 determines the *degree of ill-posedness* of the inverse problem. Consider the quantities

$$\Psi_n := \min_{k\in\mathbb{N}_0} \max\left\{\frac{1}{\gamma_k}, \sum_{0\le|j|\le k} \frac{1}{n\alpha_j}\right\} \qquad \text{and} \qquad \Phi_m := \max_{j\in\mathbb{N}} \left\{\frac{1}{\gamma_j} \min\left\{1, \frac{1}{m\alpha_j}\right\}\right\}.$$

Theorem 1. *Under some mild technical assumptions it holds that*

$$\inf_{\widetilde{\lambda}} \sup_{\lambda\in\Lambda_\gamma^r} \sup_{f\in\mathcal{F}_\alpha^d} \mathbb{E}[\|\widetilde{\lambda} - \lambda\|^2] \gtrsim \Psi_n + \Phi_m,$$

where the infimum is taken over all estimators $\widetilde{\lambda}$ based on the observations in (1) *and* (2)*.*

From a minimax point of view the optimal choice of the tuning parameter k is given by

$$k_n^* := \operatorname*{argmin}_{k\in\mathbb{N}_0} \max\left\{\frac{1}{\gamma_k}, \sum_{0\le|j|\le k} \frac{1}{n\alpha_j}\right\}.$$

Theorem 2. *Under some mild technical assumptions it holds that*

$$\sup_{\lambda\in\Lambda_\gamma^r} \sup_{f\in\mathcal{F}_\alpha^d} \mathbb{E}[\|\widehat{\lambda}_{k_n^*} - \lambda\|^2] \lesssim \Psi_n + \Phi_m.$$

Illustration of rates

Typical choices for the sequence γ:

(pol): $\gamma_0 = 1$ and $\gamma_j = |j|^{2p}$ for all $j \ne 0$, $p \ge 0$ (*Sobolev ellipsoid*)

(exp): $\gamma_j = \exp(2p|j|)$ for all $j \in \mathbb{Z}$, $p \ge 0$ (*Class of analytic functions*)

Typical choices for the sequence α:

(pol): $\alpha_0 = 1$ and $\alpha_j = |j|^{-2a}$ for all $j \ne 0$, $a \ge 0$ (f *ordinary smooth*)

(exp): $\alpha_j = \exp(-2a|j|)$ for all $j \in \mathbb{Z}$, $a \ge 0$ (f *supersmooth*)

Under slight assumptions on the parameters a and p we get the following rates:

γ	α	Ψ_n	Φ_m
(pol)	(pol)	$n^{-\frac{2p}{2a+1}}$	$m^{-\frac{p\wedge a}{a}}$
(exp)	(pol)	$(\log n)^{2a+1} \cdot n^{-1}$	m^{-1}
(pol)	(exp)	$(\log n)^{-2p}$	$(\log m)^{-2p}$
(exp)	(exp)	$n^{-\frac{p}{p+a}}$	$m^{-p/a}$ if $a \ge p$; m^{-1} if $a < p$

4. Adaptive estimation

Put

$$\widehat{\Delta}_k := \max_{0\le j\le k} |\widehat{[f]}_j|^{-2} \mathbb{1}_{\Omega_j} \qquad \text{and} \qquad \widehat{\delta}_k := (2k+1)\widehat{\Delta}_k \frac{\log(\widehat{\Delta}_k \vee (k+4))}{\log(k+4)}.$$

Defining *contrast function* and *penalty* for some $\eta \in (0,1)$ as

$$\Upsilon_{n,m}(t) := \|t\|^2 - 2\Re\langle \widehat{\lambda}_{n\wedge m}, t\rangle \qquad \text{and} \qquad \widehat{\text{pen}}_k := 1375\eta^{-1}\widehat{[\ell]}_0 \cdot \frac{\widehat{\delta}_k}{n},$$

we define $\widehat{k}$ as the minimizer of the *penalized contrast*, that is

$$\widehat{k} := \operatorname*{argmin}_{0\le k\le \widehat{K}} \left\{\Upsilon_{n,m}(\widehat{\lambda}_k) + \widehat{\text{pen}}_k\right\},$$

where the random quantity $\widehat{K}$ does neither depend on γ nor on α. To state our result, we further introduce the quantity

$$\delta_k^\alpha := (2k+1)\alpha_k^{-1} \frac{\log(\alpha_k^{-1} \vee (k+4))}{\log(k+4)}.$$

Theorem 3. *Under some mild technical assumptions it holds that*

$$\sup_{\lambda\in\Lambda_\gamma^r} \sup_{f\in\mathcal{F}_\alpha^d} \mathbb{E}[\|\widehat{\lambda}_{\widehat{k}} - \lambda\|^2] \lesssim \min_{0\le k\le K_\alpha} \left\{\frac{1}{\gamma_k} + \frac{\delta_k^\alpha}{n}\right\} + \Phi_m + \frac{1}{m} + \frac{1}{n},$$

where K_α depends on n, m and α only.

- ▷ Important ingredient of the proof: Talagrand type concentration inequalites for PPPs
- ▷ The adaptive estimator $\widehat{\lambda}_{\widehat{k}}$ attains the optimal rate of convergence in a wide range of scenarios.

5. Bibliography

[Big+13] J. Bigot, S. Gadat, T. Klein, C. Marteau (2013) *Intensity estimation of non-homogeneous Poisson processes from shifted trajectories*, Electronic Journal of Statistics 7, 881–931.

[JS13] J. Johannes, M. Schwarz (2013) *Adaptive circular deconvolution by model selection under unknown error distribution*, Bernoulli 19(5A), 1576–1611.

[Kro16a] M. Kroll (2016) *Concentration inequalities for Poisson point processes with application to intensity estimation*, in preparation.

[Kro16b] M. Kroll (2016) *Nonparametric intensity estimation from indirect point process observations with unknown error distribution*, in preparation.

LaTeX TikZposter