New Insights on Inverse Problems:

Multidimensional Strategies for Deconvolution or Regression, and Ruin Probability Estimation

Florian Dussap

Soutenance de thèse de doctorat
24 juin 2022

Overview

Multivariate Laguerre basis

The univariate Laguerre functions $(\varphi_k)_{k\in\mathbb{N}}$ are defined as:

$$\forall x \in \mathbb{R}_+, \quad \varphi_k(x) \coloneqq \sqrt{2}\, L_k(2x)\, \mathrm{e}^{-x}, \quad L_k(x) \coloneqq \sum_{j=0}^{k} \binom{k}{j} \frac{(-x)^j}{j!}.$$

For $\boldsymbol{k} \coloneqq (k_1, \ldots, k_d) \in \mathbb{N}^d$, we define the $\boldsymbol{k}$-th multivariate Laguerre function as:

$$\varphi_{\boldsymbol{k}}(\boldsymbol{x}) \coloneqq (\varphi_{k_1} \otimes \cdots \otimes \varphi_{k_d})(\boldsymbol{x}) \coloneqq \varphi_{k_1}(x_1) \times \cdots \times \varphi_{k_d}(x_d).$$

The functions $(\varphi_{\boldsymbol{k}})_{\boldsymbol{k}\in\mathbb{N}^d}$ form a basis of $\mathrm{L}^2(\mathbb{R}_+^d)$. Hence, a function $f \in \mathrm{L}^2(\mathbb{R}_+^d)$ can be decomposed as:

$$f = \sum_{\boldsymbol{k}\in\mathbb{N}^d} a_{\boldsymbol{k}}\, \varphi_{\boldsymbol{k}}, \quad a_{\boldsymbol{k}} = \langle f, \varphi_{\boldsymbol{k}} \rangle_{\mathrm{L}^2}.$$

Projection estimator

A projection estimator of a function $f \in \mathsf{L}^2(\mathbb{R}_+^d)$ is an estimator of the form:

$$\hat{f}_{\boldsymbol{m}} := \sum_{\boldsymbol{k} \leqslant \boldsymbol{m}-\mathbf{1}} \hat{a}_{\boldsymbol{k}}\, \varphi_{\boldsymbol{k}}, \quad \hat{a}_{\boldsymbol{k}} \text{ is an estimator of } a_{\boldsymbol{k}}, \quad \boldsymbol{m} \in \mathbb{N}_+^d.$$

We quantify its performance by its Mean Integrated Squared Error (MISE):

$$\mathbb{E}\|f - \hat{f}_{\boldsymbol{m}}\|_{\mathsf{L}^2}^2.$$

Let $f_{\boldsymbol{m}}$ be the projection of f on the space:

$$S_{\boldsymbol{m}} := \operatorname{Span}\left(\varphi_{\boldsymbol{k}} : \boldsymbol{k} \leqslant \boldsymbol{m} - \mathbf{1}\right), \quad D_{\boldsymbol{m}} := \dim(S_{\boldsymbol{m}}) = m_1 \cdots m_d.$$

The MISE can be decomposed as the sum of a bias term and a variance term:

$$\begin{aligned}\mathbb{E}\|f - \hat{f}_{\boldsymbol{m}}\|_{\mathsf{L}^2}^2 &= \|f - f_{\boldsymbol{m}}\|_{\mathsf{L}^2}^2 + \mathbb{E}\|\hat{f}_{\boldsymbol{m}} - f_{\boldsymbol{m}}\|_{\mathsf{L}^2}^2 \\ &= \operatorname{dist}_{\mathsf{L}^2}^2(f, S_{\boldsymbol{m}}) + \sum_{\boldsymbol{k} \leqslant \boldsymbol{m}-\mathbf{1}} \mathbb{E}\Big[(\hat{a}_{\boldsymbol{k}} - a_{\boldsymbol{k}})^2\Big].\end{aligned}$$

Sobolev–Laguerre spaces

Definition

For $\boldsymbol{s} \in (0, +\infty)^d$ and $L > 0$, we define the Sobolev–Laguerre ball of regularity $\boldsymbol{s}$ and radius L as:

$$\mathrm{W}^{\boldsymbol{s}}(\mathbb{R}_+, L) := \left\{ f \in \mathrm{L}^2(\mathbb{R}_+^d) \,\middle|\, \sum_{\boldsymbol{k} \in \mathbb{N}^d} \langle f, \varphi_{\boldsymbol{k}} \rangle^2 \, \boldsymbol{k}^{\boldsymbol{s}} \leqslant L \right\}.$$

- When $d = 1$, these spaces were introduced by [Bongioanni and Torrea, 2009] to study the Laguerre operator.
- When $d = 1$, [Comte and Genon-Catalot, 2015] show that s is the regularity of the function f.
- If $f \in \mathrm{W}^{\boldsymbol{s}}(\mathbb{R}_+^d, L)$, then the bias term decreases as $m_1^{-s_1} + \cdots + m_d^{-s_d}$.

Hypermatrices

- For $\boldsymbol{m} \in \mathbb{N}_+^d$, let $\mathbb{R}^{\boldsymbol{m}}$ be the space of $m_1 \times \cdots \times m_d$ hypermatrices.
- The spaces $S_{\boldsymbol{m}}$ and $\mathbb{R}^{\boldsymbol{m}}$ are isometric (function $\leftrightarrow$ coefficients).
- We define the r-contracted product between hypermatrices with compatible shapes as:

$$\left[\mathbf{A} \times_r \mathbf{B}\right]_{\boldsymbol{j},\ell} := \sum_{\boldsymbol{k}=(k_1,\dots,k_r)} \mathbf{A}_{\boldsymbol{j},\boldsymbol{k}} \, \mathbf{B}_{\boldsymbol{k},\ell}.$$

- If $\mathbf{G} \in \mathbb{R}^{\boldsymbol{m}\times\boldsymbol{m}}$, then $\mathbf{a} \mapsto \mathbf{G} \times_d \mathbf{a}$ is an endomorphism of $\mathbb{R}^{\boldsymbol{m}}$.
- As an endomorphism, $\mathbf{G} \in \mathbb{R}^{\boldsymbol{m}\times\boldsymbol{m}}$ has eigenvalues, a trace, an operator norm, a Frobenius norm, ...

Nonparametric Estimation of the Expected Discounted Penalty Function in the Compound Poisson Model

Electronic Journal of Statistics, 16(1), 2022.

The compound Poisson risk model [Asmussen and Albrecher, 2010]

Let $(U_t)_{t \geqslant 0}$ be the reserve process of an insurance company. In the compound Poisson risk model, this process is given by:

$$U_t = u + ct - \sum_{i=1}^{N_t} X_i,$$

where:

- $u \geqslant 0$ is the initial reserve,
- $c > 0$ is the premium rate,
- the claim number process $(N_t)_{t \geqslant 0}$ is a homogeneous Poisson process with intensity λ,
- the individual claim sizes $(X_i)_{i \geqslant 1}$ are positive, i.i.d. with density f and mean μ, independent of $(N_t)_{t \geqslant 0}$.

The Expected Discounted Penalty Function (EDPF)

We denote the time of ruin by $\tau := \inf\{t \geqslant 0 \,|\, U_t < 0\} \in \mathbb{R}_+ \cup \{\infty\}$.

Assumption (Safety Loading Condition)

A1 We assume that $c > \lambda\mu$. Introducing the parameter $\theta := \frac{\lambda\mu}{c}$, the previous condition is equivalent to $\theta < 1$.

Under the SLC, we have $\mathbb{P}[\tau < \infty] < 1$.

The Expected Discounted Penalty Function ([Gerber and Shiu, 1998]), is defined as:

$$\phi(u) := \mathbb{E}\Big[\mathrm{e}^{-\delta\tau} w(U_{\tau^-}, |U_\tau|)\, \mathbf{1}_{\{\tau<\infty\}} \,\Big|\, U_0 = u\Big],$$

where $\delta \geqslant 0$ is a discounting force of interest, and $w\colon \mathbb{R}_+^2 \to \mathbb{R}_+$ is a penalty function.

In the following, we consider the case of the ruin probability ($\delta = 0$ and $w(x, y) = 1$).

Observations and goal

We assume that c is known but the parameters (λ, μ, f) of the compound Poisson process are not. We suppose we have access to a trajectory of the reserve process $(U_t)_{t\in[0,T]}$ on a time interval $[0, T]$, on which we observe:

$$N_T \text{ and } X_1, \ldots, X_{N_T}.$$

Goal

We want to estimate the Gerber–Shiu function from the observations $(N_T, X_1, \ldots, X_{N_T})$ with c known but (λ, μ, f) unknown.

Renewal equation

Theorem

Under Assumption A1 (SLC), the ruin probability satisfies the equation:

$$\phi = \phi * g + h,$$

with:

$$g(x) := \frac{\lambda}{c} S(x), \quad h(u) := \frac{\lambda}{c} \int_u^{+\infty} S(x)\, \mathrm{d}x,$$

where $S(x) := \mathbb{P}[X_1 > x]$ *is the survival function of the* $(X_i)_{i \geqslant 1}$.

Following the work of [Comte et al., 2017] and [Mabon, 2017], [Zhang and Su, 2018] estimate these functions by projection on the Laguerre basis.

$$\phi = \sum_{k=0}^{+\infty} a_k \varphi_k, \qquad g = \sum_{k=0}^{+\infty} b_k \varphi_k, \qquad h = \sum_{k=0}^{+\infty} c_k \varphi_k.$$

Estimation of g and h

Let $\Phi_k(x) := \int_0^x \varphi_k(t)\,\mathrm{d}t$. The Laguerre coefficients of g and h are given by:

$$b_k = \frac{\lambda}{c}\,\mathbb{E}\big[\Phi_k(X)\big], \quad c_k = \frac{\lambda}{c}\,\mathbb{E}\left[\int_0^X \Phi_k(x)\,\mathrm{d}x\right],$$

so we estimate them with empirical means:

$$\hat{b}_k = \frac{1}{cT}\sum_{i=1}^{N_T} \Phi_k(X_i), \quad \hat{c}_k = \frac{1}{cT}\sum_{i=1}^{N_T} \int_0^{X_i} \Phi_k(x)\,\mathrm{d}x.$$

For $m \in \mathbb{N}_+$, the projection estimators of g and h are:

$$\hat{g}_m := \sum_{k=0}^{m-1} \hat{b}_k\,\varphi_k, \qquad \hat{h}_m := \sum_{k=0}^{m-1} \hat{c}_k\,\varphi_k.$$

MISE of $\hat{g}_m$ and $\hat{h}_m$

Assumption

A2 $\mathbb{E}[X^3]$ is finite.

Proposition

Under Assumptions A1 and A2, we have:

$$\mathbb{E}\|g - \hat{g}_m\|_{\mathrm{L}^2}^2 \leqslant \mathrm{dist}_{\mathrm{L}^2}^2(g, S_m) + \frac{\lambda}{c^2 T}\mathbb{E}[X],$$

$$\mathbb{E}\|h - \hat{h}_m\|_{\mathrm{L}^2}^2 \leqslant \mathrm{dist}_{\mathrm{L}^2}^2(h, S_m) + \frac{\lambda}{3c^2 T}\mathbb{E}[X^3].$$

- The variance term does not depend on m.
- For m large enough, the convergence rate is T^{-1}.

Interlude: Laguerre deconvolution [Comte et al., 2017] [Mabon, 2017]

The Laguerre functions satisfy the relation:

$$\forall j, k \in \mathbb{N}, \quad \varphi_j * \varphi_k = 2^{-\frac{1}{2}}(\varphi_{j+k} - \varphi_{j+k+1}).$$

Using this relation, one can show that if f and g are two functions on $\mathbb{R}_+$ then their Laguerre coefficients satisfy:

$$c(f * g) = c(f) * \Delta(g), \quad \Delta_k(g) := \begin{cases} 2^{-\frac{1}{2}} \left(c_k(g) - c_{k-1}(g)\right) & : k \geqslant 1, \\ 2^{-\frac{1}{2}} \, c_0(g) & : k = 0. \end{cases}$$

If $\mathbf{c}_m(f)$ denotes the vector of the first m coefficients of f, we have:

$$\mathbf{c}_m(f * g) = \mathbf{G}_m \times \mathbf{c}_m(f), \quad \mathbf{G}_m := \begin{bmatrix} \Delta_0 & 0 & 0 & 0 & 0 \\ \Delta_1 & \Delta_0 & 0 & 0 & 0 \\ \Delta_2 & \Delta_1 & \Delta_0 & 0 & 0 \\ \cdots & \cdots & \cdots & \Delta_0 & 0 \\ \Delta_{m-1} & \Delta_{m-2} & \cdots & \cdots & \Delta_0 \end{bmatrix}.$$

Laguerre deconvolution estimator

If we use the convolution property of the Laguerre functions in the equation $\phi = \phi * g + h$, we obtain the following relation between the coefficients of ϕ, g and h:

$$\mathbf{c}_m = \mathbf{A}_m \times \mathbf{a}_m \iff \mathbf{a}_m = \mathbf{A}_m^{-1} \times \mathbf{c}_m,$$

with $\mathbf{A}_m \coloneqq \mathbf{Id}_m - \mathbf{G}_m$.

Assumption

A3 $(b_{k+1} - b_k)_{k \in \mathbb{N}} \in \ell^1(\mathbb{N})$.

Lemma

Under Assumption A1 and A3, we have $\|\mathbf{A}_m^{-1}\|_{\mathrm{op}} \leqslant \frac{2}{1 - \|g\|_{\mathrm{L}^1}} \leqslant \frac{2}{1-\theta}$.

For $\theta_0 < 1$ a truncation parameter, we estimate ϕ by:

$$\hat{\phi}_m := \sum_{k=0}^{m-1} \hat{a}_k \, \varphi_k, \quad \hat{\mathbf{a}}_m := \widetilde{\mathbf{A}}_m^{-1} \times \hat{\mathbf{c}}_m, \quad \widetilde{\mathbf{A}}_m^{-1} := \hat{\mathbf{A}}_m^{-1} \mathbf{1}_{\left\{ \|\hat{\mathbf{A}}_m^{-1}\|_{\mathrm{op}} \leqslant \frac{2}{1-\theta_0} \right\}}.$$

Proposition

Under Assumptions A1, A2, and A3, if $\theta < \theta_0$ then it holds:

$$\mathbb{E}\|\phi - \hat{\phi}_m\|_{\mathrm{L}^2}^2 \leqslant \mathrm{dist}_{\mathrm{L}^2}^2(\phi, S_m) + C\,\frac{m}{T}.$$

- This method does not recover the rate T^{-1} for the ruin probability ([Pitts, 1994] and [Politis, 2003]).
- The functions g and h are estimated with the rate T^{-1}, but the deconvolution step loses a factor m in the variance term.

Laguerre–Fourier estimator [Dussap, 2022]

Since $\phi = \phi * g + h$, we have $\mathcal{F}\phi = \frac{\mathcal{F}h}{1-\mathcal{F}g}$. We compute the coefficients of ϕ using Plancherel theorem:

$$a_k = \langle \phi, \varphi_k \rangle = \frac{1}{2\pi} \langle \mathcal{F}\phi, \mathcal{F}\varphi_k \rangle = \frac{1}{2\pi} \left\langle \frac{\mathcal{F}h}{1 - \mathcal{F}g}, \mathcal{F}\varphi_k \right\rangle.$$

Definition

For $\hat{g}$ and $\hat{h}$ two estimators of g and h, and for θ_0 a truncation parameter, we estimate ϕ by:

$$\hat{\phi}_{m_1,\hat{g},\hat{h}} \coloneqq \sum_{k=0}^{m_1-1} \hat{a}_{k,\hat{g},\hat{h}} \varphi_k, \quad \hat{a}_{k,\hat{g},\hat{h}} \coloneqq \frac{1}{2\pi} \left\langle \frac{\mathcal{F}\hat{h}}{1 - \widetilde{\mathcal{F}g}}, \mathcal{F}\varphi_k \right\rangle,$$

$$\widetilde{\mathcal{F}g} \coloneqq (\mathcal{F}\hat{g}) \mathbf{1}_{\{|\mathcal{F}\hat{g}| < \theta_0\}}.$$

Proposition

Under Assumption A1 and A2, if $\theta < \theta_0$ then it holds:

$$\|\phi - \hat{\phi}_{m_1,\hat{g},\hat{h}}\|_{\mathrm{L}^2}^2 \leqslant \mathrm{dist}_{\mathrm{L}^2}^2(\phi, S_{m_1}) + \frac{2}{(1-\theta_0)^2}\|h - \hat{h}\|_{\mathrm{L}^2}^2 + \frac{2\,\|h\|_{\mathrm{L}^1}^2}{(1-\theta_0)^2(1-\theta)^2}\left(1 + \frac{\|g\|_{\mathrm{L}^1}^2}{(\theta_0-\theta)^2}\right)\|g - \hat{g}\|_{\mathrm{L}^2}^2.$$

If we use the Laguerre projection estimators $\hat{g}_{m_2}$ and $\hat{h}_{m_3}$, we obtain the following result.

Corollary

Under Assumptions A1 and A2, if $\theta < \theta_0$ then it holds:

$$\mathbb{E}\|\phi - \hat{\phi}_{m_1,m_2,m_3}\|_{\mathrm{L}^2}^2 \leqslant \mathrm{dist}_{\mathrm{L}^2}^2(\phi, S_{m_1}) + C\left(\mathrm{dist}_{\mathrm{L}^2}^2(g, S_{m_2}) + \mathrm{dist}_{\mathrm{L}^2}^2(h, S_{m_3}) + \frac{1}{T}\right).$$

Conclusion and perspectives

- If ϕ belongs to a Sobolev–Laguerre space of regularity greater than 1, it is possible to estimate the EDPF with rate T^{-1}.
- The Laguerre deconvolution method fails to recover the parametric rate.
- The Laguerre–Fourier method could be extended to more general risk models.
- The absence of a bias-variance compromise raises questions about how to perform a model selection procedure in practice.

Anisotropic Multivariate Deconvolution Using Projection on the Laguerre Basis

Journal of Statistical Planning and Inference, 215:23–46, 2021.

Density estimation from indirect observations

We observe random vectors $\boldsymbol{Z}_1, \ldots, \boldsymbol{Z}_n$ in $\mathbb{R}_+^d$ such that:

$$\boldsymbol{Z}_i = \boldsymbol{X}_i + \boldsymbol{Y}_i,$$

where:

- $\boldsymbol{X}_i \in \mathbb{R}_+^d$ are i.i.d. with unknown density f that we want to estimate;
- $\boldsymbol{Y}_i \in \mathbb{R}_+^d$ are i.i.d. with known density g, and are independent from the $\boldsymbol{X}_i$.

Under these assumptions, $\boldsymbol{Z}_1, \ldots, \boldsymbol{Z}_n$ are i.i.d. with density h given by:

$$\forall \boldsymbol{x} \in \mathbb{R}_+^d, \quad h(\boldsymbol{x}) = (f * g)(\boldsymbol{x}) \coloneqq \int_{\mathbb{R}_+^d} f(\boldsymbol{t}) g(\boldsymbol{x} - \boldsymbol{t}) \, \mathrm{d}\boldsymbol{t}.$$

For $d = 1$, this problem is studied by [Mabon, 2017] .

Multivariate Laguerre basis

We assume that f, g and h belong to $\mathrm{L}^2(\mathbb{R}_+^d)$, and we decompose them in the multivariate Laguerre basis:

$$f = \sum_{\boldsymbol{k} \in \mathbb{N}^d} a_{\boldsymbol{k}}\, \varphi_{\boldsymbol{k}}, \qquad g = \sum_{\boldsymbol{k} \in \mathbb{N}^d} b_{\boldsymbol{k}}\, \varphi_{\boldsymbol{k}}, \qquad h = \sum_{\boldsymbol{k} \in \mathbb{N}^d} c_{\boldsymbol{k}}\, \varphi_{\boldsymbol{k}}.$$

Using the relation $\varphi_j * \varphi_k = 2^{-1/2}(\varphi_{j+k} - \varphi_{j+k+1})$, the convolution equation $h = f * g$ is equivalent to:

$$c = \beta * a, \qquad \beta_{\boldsymbol{k}} := 2^{-d/2} \sum_{\boldsymbol{\varepsilon} \in \{0,1\}^d} (-1)^{|\boldsymbol{\varepsilon}|} b_{\boldsymbol{k}-\boldsymbol{\varepsilon}},$$

with $|\boldsymbol{\varepsilon}| := \varepsilon_1 + \cdots + \varepsilon_{\boldsymbol{d}}$.

Hypermatrices and estimation

Let $\mathbf{a}_{\boldsymbol{m}}, \mathbf{c}_{\boldsymbol{m}} \in \mathbb{R}^{\boldsymbol{m}}$ be the hypermatrices of the coefficients $a_{\boldsymbol{k}}$ and $c_{\boldsymbol{k}}$ for $\boldsymbol{k} \leqslant \boldsymbol{m} - \mathbf{1}$, and let $\mathbf{G}_{\boldsymbol{m}} \in \mathbb{R}^{\boldsymbol{m} \times \boldsymbol{m}}$ be the hypermatrix:

$$[\mathbf{G}_{\boldsymbol{m}}]_{\boldsymbol{j},\boldsymbol{k}} \coloneqq \beta_{\boldsymbol{j}-\boldsymbol{k}} \, \mathbf{1}_{\boldsymbol{k} \leqslant \boldsymbol{j}}.$$

Then, we have:

$$\mathbf{c}_{\boldsymbol{m}} = \mathbf{G}_{\boldsymbol{m}} \times_d \mathbf{a}_{\boldsymbol{m}} \iff \mathbf{a}_{\boldsymbol{m}} = \mathbf{G}_{\boldsymbol{m}}^{-1} \times_d \mathbf{c}_{\boldsymbol{m}}.$$

Since the coefficient of h are given by $c_{\boldsymbol{k}} = \mathbb{E}[\varphi_{\boldsymbol{k}}(\boldsymbol{Z})]$, we estimate them with empirical means, and we estimate f with a plug-in estimator:

$$\hat{f}_{\boldsymbol{m}} \coloneqq \sum_{\boldsymbol{k} \leqslant \boldsymbol{m}-\mathbf{1}} \hat{a}_{\boldsymbol{k}} \, \varphi_{\boldsymbol{k}}, \quad \hat{\mathbf{a}}_{\boldsymbol{m}} \coloneqq \mathbf{G}_{\boldsymbol{m}}^{-1} \times_d \hat{\mathbf{c}}_{\boldsymbol{m}}, \quad \hat{c}_{\boldsymbol{k}} \coloneqq \frac{1}{n} \sum_{i=1}^{n} \varphi_{\boldsymbol{k}}(\boldsymbol{Z}_i).$$

Upper bound on the variance term

Assumption

A1 g is bounded.

A2 For all $J \subset \{1, \dots, d\}$, the following moments are finite:

$$M_J(g) := \int_{\mathbb{R}_+^d} \left(\prod_{i \in J} y_i^{-1/2} \right) g(\boldsymbol{y}) \, \mathrm{d}\boldsymbol{y}.$$

A3 $\beta \in \ell^1(\mathbb{N}^d)$.

Proposition

Under Assumptions A1 and A2, we have:

$$\mathbb{E}\|\hat{f}_{\boldsymbol{m}} - f_{\boldsymbol{m}}\|_{\mathsf{L}^2}^2 \leqslant \frac{c_d(g)\sqrt{D_{\boldsymbol{m}}}\,\|\mathbf{G}_{\boldsymbol{m}}^{-1}\|_{\mathrm{op}}^2}{n} \wedge \frac{\|g\|_\infty \|\mathbf{G}_{\boldsymbol{m}}^{-1}\|_{\mathsf{F}}^2}{n},$$

where $c_d(g)$ is a constant depending on $\{M_J(g) : J \subset \{1, \dots, d\}\}$.

Upper bound on the Frobenius norm

We consider the case $d = 1$.

Proposition ([Comte et al., 2017])

We assume A3 and we make the following assumptions:

1. *The Laplace transform $\mathcal{L}g$ of g does not vanish on the half plane $\mathcal{P}_+ := \{s \in \mathbb{C} \mid \mathfrak{Re}\, s \geqslant 0\}$.*
2. *The Fourier transform $\mathcal{F}g$ of g has an asymptotic expansion:*

$$\mathcal{F}g(\omega) = \omega^{-\alpha}\big(K_\alpha + o(1)\big), \ |\omega| \to +\infty$$

with $\alpha \in \mathbb{N}_+$ and $K_\alpha \neq 0$.

Then there exists $C(g) > 0$ depending on g such that for $m \geqslant 4$, we have:

$$\left\|\mathbf{G}_m^{-1}\right\|_{\mathsf{F}}^2 \leqslant C(g)\, m^{2\alpha}.$$

We consider the case $d \geqslant 2$.

Proposition

We assume A3. We assume that $\mathcal{L}g$ does not vanish on $\mathcal{P}_+^d$ and we assume there exists $\boldsymbol{\alpha} \in \mathbb{N}_+^d$ such that the function:

$$K_{\boldsymbol{\alpha}}(\boldsymbol{s}) := (\mathbf{1} + \boldsymbol{s})^{\boldsymbol{\alpha}} \, \mathcal{L}g(\boldsymbol{s}), \quad \boldsymbol{s} \in \mathcal{P}_+^d,$$

can be extended to a nonzero function on $(\mathcal{P}_+ \cup \{\infty\})^d$ such that its restriction to $(\mathrm{i}\mathbb{R} \cup \{\infty\})^d$ is continuous. Then for $\boldsymbol{m} \in \mathbb{N}_+^d$ large enough, there exists $C(g) > 0$ depending on g such that:

$$\|\mathbf{G}_{\boldsymbol{m}}^{-1}\|_{\mathsf{F}}^2 \leqslant C(g)\, \boldsymbol{m}^{2\boldsymbol{\alpha}}.$$

Convergence rates

Theorem

Under Assumptions A1, A2 and A3, if we assume that g satisfies the assumptions of the last proposition with $\boldsymbol{\alpha} \in \mathbb{N}_+^d$, then for $\boldsymbol{m}_{\mathrm{opt}} \in \mathbb{N}_+^d$ given by:

$$m_{\mathrm{opt},j} \propto n^{1/\left(s_j + s_j \sum_{i=1}^d \frac{2\alpha_i}{s_i}\right)}, \quad j = 1, \dots, d,$$

we have:

$$\sup_{f \in \mathrm{W}^{\boldsymbol{s}}(\mathbb{R}_+^d, L)} \mathbb{E} \| f - \hat{f}_{\boldsymbol{m}_{\mathrm{opt}}} \|_{\mathrm{L}^2}^2 \leqslant C \, n^{-1/(1+\sum_{i=1}^d \frac{2\alpha_i}{s_i})}.$$

These rates are similar to those found on Sobolev balls for a kernel estimator by [Comte and Lacour, 2013].

Model selection

We use a procedure similar to the bandwidth selection procedure of [Goldenshluger and Lepski, 2011] that was introduced for model selection by [Chagny, 2013] for the estimation of a conditional density.

We consider the model collection:

$$\mathcal{M}_n := \left\{ \boldsymbol{m} \in \mathbb{N}_+^d \;\middle|\; D_{\boldsymbol{m}} \, \|\mathbf{G}_{\boldsymbol{m}}^{-1}\|_{\mathrm{op}}^2 \leqslant \frac{n}{\log n} \right\}.$$

Let:

$$V(\boldsymbol{m}) := \frac{c_d(g)\sqrt{D_{\boldsymbol{m}}}\,\|\mathbf{G}_{\boldsymbol{m}}^{-1}\|_{\mathrm{op}}^2}{n} \wedge \frac{(\|g\|_\infty \vee 1)\|\mathbf{G}_{\boldsymbol{m}}^{-1}\|_{\mathrm{F}}^2 \log n}{n},$$

$$A(\boldsymbol{m}) := \max_{\boldsymbol{m}' \in \mathcal{M}_n} \Big(\|\hat{f}_{\boldsymbol{m}'} - \hat{f}_{\boldsymbol{m} \wedge \boldsymbol{m}'}\|_{\mathrm{L}^2}^2 - \kappa_1 V(\boldsymbol{m}') \Big)_+ .$$

We choose $\hat{\boldsymbol{m}}$ as:

$$\hat{\boldsymbol{m}} := \operatorname*{arg\,min}_{\boldsymbol{m} \in \mathcal{M}_n} \big\{ A(\boldsymbol{m}) + \kappa_2 V(\boldsymbol{m}) \big\}.$$

Oracle bound

Assumption

A4 $\forall \delta > 0$, $\forall n \in \mathbb{N}_+$, $\sum_{\boldsymbol{m} \in \mathcal{M}_n} \|\mathbf{G}_{\boldsymbol{m}}^{-1}\|_{\mathrm{op}}^2 \, \mathrm{e}^{-\delta\sqrt{D_{\boldsymbol{m}}}} \leqslant C(\delta)$.

Theorem

Under Assumptions A1, A2 and A4, there exists a constant $\kappa_0(d) > 0$ such that for every choice of κ_1, κ_2 satisfying $\kappa_0(d) < \kappa_1 \leqslant \kappa_2$, we have:

$$\mathbb{E}\|f - \hat{f}_{\hat{\boldsymbol{m}}}\|_{\mathrm{L}^2}^2 \leqslant C \inf_{\boldsymbol{m} \in \mathcal{M}_n} \left(\|f - f_{\boldsymbol{m}}\|_{\mathrm{L}^2}^2 + V(\boldsymbol{m}) \right) + \frac{C'}{n}.$$

Illustration

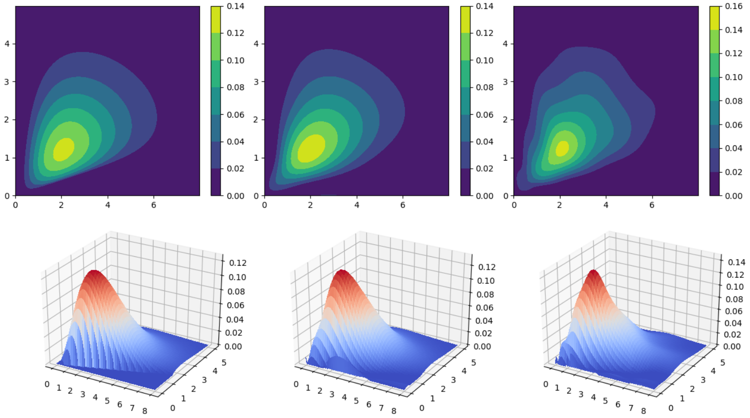

Figure: Density estimation, sample size $n = 5000$. First column: true density, second column: adaptive estimator $\hat{f}_{\hat{m}}$, third column: max model estimator $\hat{f}_{(12,12)}$. The selected model is $\hat{m} = (5, 8)$.

Conclusion and perspective

- We extend the Laguerre deconvolution method to multivariate functions.
- We obtain rates of convergence for the density deconvolution problem on $\mathbb{R}_+^d$ similar to those on $\mathbb{R}^d$.
- Our estimation strategy assumes that the noise distribution is known. A future work would be to construct an estimation procedure where the noise distribution is unknown and has to be estimated too.

Nonparametric Multiple Regression on Non-compact Domains

In revision.

Regression model with random design

Let $A \subset \mathbb{R}^p$, we observe $n \geqslant 1$ r.v. $(\boldsymbol{X}_i, Y_i) \in A \times \mathbb{R}$ given by:

$$Y_i = b(\boldsymbol{X}_i) + \varepsilon_i,$$

where:

- $(\boldsymbol{X}_i)$ are i.i.d. with unknown distribution μ.
- (ε_i) are i.i.d. with zero mean and known variance σ^2.
- $(\boldsymbol{X}_i)$ and (ε_i) are independent.

Our goal is to estimate the regression function $b \colon A \to \mathbb{R}$. To quantify the error of an estimator, we consider two norms:

$$\|t\|_n^2 := \frac{1}{n} \sum_{i=1}^{n} t(\boldsymbol{X}_i)^2, \quad \|t\|_\mu^2 := \int_A t(\boldsymbol{x})^2 \,\mathrm{d}\mu(\boldsymbol{x}).$$

The error relative to the norm $\|\cdot\|_\mu$ can be viewed as a prediction error:

$$\forall \hat{b} \text{ estimator, } \|b - \hat{b}\|_\mu^2 = \mathbb{E}_{\boldsymbol{X}\sim\mu}\Big[\big(b(\boldsymbol{X}) - \hat{b}(\boldsymbol{X})\big)^2 \,\Big|\, \boldsymbol{X}_1, \ldots, \boldsymbol{X}_n\Big].$$

Assumptions

1. Following [Baraud, 2002], we assume that $\mu \ll \nu$ for a fixed measure ν, and that $\frac{\mathrm{d}\mu}{\mathrm{d}\nu}$ is bounded on A. Hence, we have $\mathrm{L}^2(A, \mu) \subset \mathrm{L}^2(A, \nu)$.

2. If A is compact, we assume that:

$$\forall \boldsymbol{x} \in A, \quad \frac{\mathrm{d}\mu}{\mathrm{d}\nu}(\boldsymbol{x}) \geqslant f_0 > 0.$$

   Hence, the norms $\|\cdot\|_\mu$ and $\|\cdot\|_\nu$ are equivalent, and we have $\mathrm{L}^2(A, \mu) = \mathrm{L}^2(A, \nu)$.

3. We assume that $b \in \mathrm{L}^{2r}(A, \mu)$ for some $r \in (1, +\infty]$. We consider $r' \in [1, +\infty)$ such that $\frac{1}{r} + \frac{1}{r'} = 1$.

4. We assume that $A = A_1 \times \cdots \times A_p$ and that $\nu = \nu_1 \otimes \cdots \otimes \nu_p$.

Projection estimator

Let $(\varphi_k^i)_{k\in\mathbb{N}}$ be an orthonormal basis of $\mathrm{L}^2(A_i,\nu_i)$. For $\boldsymbol{k} \in \mathbb{N}^p$, we define:

$$\varphi_{\boldsymbol{k}}(\boldsymbol{x}) \coloneqq (\varphi_{k_1}^1 \otimes \cdots \otimes \varphi_{k_p}^p)(\boldsymbol{x}) \coloneqq \varphi_{k_1}^1(x_1) \times \cdots \times \varphi_{k_p}^p(x_p).$$

We estimate b by a least squares minimization on $S_{\boldsymbol{m}}$:

$$\hat{b}_{\boldsymbol{m}} \coloneqq \operatorname*{arg\,min}_{t\in S_{\boldsymbol{m}}} \frac{1}{n}\sum_{i=1}^{n}\left[Y_i - t(\boldsymbol{X}_i)\right]^2.$$

Example

1. For $A = [-\pi,\pi]$ and $\nu = \mathrm{Leb}$, we choose the trigonometric basis.
2. For $A = \mathbb{R}$ and $\nu = \mathrm{Leb}$, we choose $\varphi_k(x) = c_k H_k(x)\mathrm{e}^{-x^2/2}$ with H_k the k-th Hermite polynomial.

This estimator can be computed using hypermatrix calculus:

$$\hat{b}_{\boldsymbol{m}} = \sum_{\boldsymbol{k} \leqslant \boldsymbol{m}-\mathbf{1}} \hat{a}_{\boldsymbol{k}}^{(\boldsymbol{m})} \varphi_{\boldsymbol{k}}, \qquad \hat{\mathbf{a}}^{(\boldsymbol{m})} := \operatorname*{arg\,min}_{\mathbf{a} \in \mathbb{R}^{\boldsymbol{m}}} \left\| \mathbf{Y} - \hat{\boldsymbol{\Phi}}_{\boldsymbol{m}} \times_p \mathbf{a} \right\|_{\mathbb{R}^n}^2$$

$$= \hat{\mathbf{G}}_{\boldsymbol{m}}^{-1} \times_p \hat{\boldsymbol{\Phi}}_{\boldsymbol{m}}^{*} \times_1 \mathbf{Y},$$

where $\mathbf{Y} := (Y_1, \ldots, Y_n) \in \mathbb{R}^n$, and where:

$$\hat{\mathbf{G}}_{\boldsymbol{m}} := \Big[\langle \varphi_{\boldsymbol{j}}, \varphi_{\boldsymbol{k}} \rangle_n \Big]_{\boldsymbol{j},\boldsymbol{k}} \in \mathbb{R}^{\boldsymbol{m} \times \boldsymbol{m}}, \quad \hat{\boldsymbol{\Phi}}_{\boldsymbol{m}} := \Big[\varphi_{\boldsymbol{j}}(\boldsymbol{X}_i) \Big]_{i,\boldsymbol{j}} \in \mathbb{R}^{n \times \boldsymbol{m}}.$$

In the following, we also consider the expectation of $\hat{\mathbf{G}}_{\boldsymbol{m}}$:

$$\mathbf{G}_{\boldsymbol{m}} := \mathbb{E}\big[\hat{\mathbf{G}}_{\boldsymbol{m}}\big] = \Big[\langle \varphi_{\boldsymbol{j}}, \varphi_{\boldsymbol{k}} \rangle_\mu \Big]_{\boldsymbol{j},\boldsymbol{k}} \in \mathbb{R}^{\boldsymbol{m} \times \boldsymbol{m}}.$$

Basic bound on the empirical risk

We recall the classical bias-variance decomposition of the empirical risk.

Proposition

If $\hat{\mathbf{G}}_{\boldsymbol{m}}$ is invertible, then we have:

$$\mathbb{E}\left[\|b - \hat{b}_{\boldsymbol{m}}\|_n^2 \,\middle|\, \boldsymbol{X}_1, \ldots, \boldsymbol{X}_n\right] = \inf_{t \in S_{\boldsymbol{m}}} \|b - t\|_n^2 + \sigma^2 \frac{D_{\boldsymbol{m}}}{n}.$$

If $\hat{\mathbf{G}}_{\boldsymbol{m}}$ is invertible a.s., then we have:

$$\mathbb{E}\|b - \hat{b}_{\boldsymbol{m}}\|_n^2 \leqslant \inf_{t \in S_{\boldsymbol{m}}} \|b - t\|_\mu^2 + \sigma^2 \frac{D_{\boldsymbol{m}}}{n}.$$

From the empirical norm to the design norm

We introduce the event:

$$\Omega_{\boldsymbol{m}}(\delta) \coloneqq \left\{ \sup_{t \in S_{\boldsymbol{m}} \setminus \{0\}} \frac{\|t\|_\mu^2}{\|t\|_n^2} \leqslant \frac{1}{1-\delta} \right\}, \quad \delta \in (0,1).$$

Using matrix concentration inequalities from [Tropp, 2012], the following bound holds.

Lemma

For all $\delta \in (0,1)$ and all $\boldsymbol{m} \in \mathbb{N}_+^p$, we have:

$$\mathbb{P}[\Omega_{\boldsymbol{m}}(\delta)^c] \leqslant D_{\boldsymbol{m}} \exp\left(-h(\delta) \frac{n}{L(\boldsymbol{m}) \|\mathbf{G}_{\boldsymbol{m}}^{-1}\|_{\mathrm{op}}} \right),$$

where $h(\delta) \coloneqq (1-\delta)\log(1-\delta) + \delta$, and where:

$$L(\boldsymbol{m}) \coloneqq \Big\| \sum_{\boldsymbol{k} \leqslant \boldsymbol{m}-\boldsymbol{1}} \varphi_{\boldsymbol{k}}^2 \Big\|_\infty = \sup_{t \in S_{\boldsymbol{m}} \setminus \{0\}} \frac{\|t\|_\infty^2}{\|t\|_\nu^2}.$$

Remarks on the lemma

- For the trigonometric basis, we have $L(m) \leqslant m$.
- For the Hermite basis, we have $L(m) \leqslant C\sqrt{m}$.
- If A is compact, then we have $\|\mathbf{G}_{\boldsymbol{m}}^{-1}\|_{\mathrm{op}} \leqslant 1/f_0$.
- If $A = \mathbb{R}$ and $(\varphi_k)_{k\in\mathbb{N}}$ is the Hermite basis, then we have $\|\mathbf{G}_{\boldsymbol{m}}^{-1}\|_{\mathrm{op}} \geqslant C(\mu)\sqrt{m}$ [Comte and Genon-Catalot, 2020].

Bound on the prediction risk

Let us consider the collection:

$$\mathcal{M}_{n,\alpha} \coloneqq \left\{ \boldsymbol{m} \in \mathbb{N}_+^p \,\middle|\, L(\boldsymbol{m})\left(\|\mathbf{G}_{\boldsymbol{m}}^{-1}\|_{\mathrm{op}} \vee 1\right) \leqslant \alpha \frac{n}{\log n} \right\}.$$

If $\boldsymbol{m} \in \mathcal{M}_{n,\alpha}$, then we have $\mathbb{P}[\Omega_{\boldsymbol{m}}(\delta)^{\mathsf{c}}] \leqslant D_{\boldsymbol{m}}\, n^{-\alpha} \leqslant n^{-\alpha+1}$.

Theorem

For all $\alpha \in (0, \frac{1}{2r'+1})$ and for all $\boldsymbol{m} \in \mathcal{M}_{n,\alpha}$ we have:

$$\mathbb{E}\|b - \hat{b}_{\boldsymbol{m}}\|_\mu^2 \leqslant C_n(\alpha, r') \inf_{t \in S_{\boldsymbol{m}}} \|b - t\|_\mu^2 + C'(\alpha, r')\, \sigma^2 \frac{D_{\boldsymbol{m}}}{n} + \mathrm{o}\left(\frac{1}{n}\right).$$

Model selection and oracle bound for the empirical risk

We choose the model with a penalized criterion:

$$\hat{\boldsymbol{m}} := \underset{\boldsymbol{m}\in\widehat{\mathcal{M}}_{n,\beta}}{\arg\min}\left(-\|\hat{b}_{\boldsymbol{m}}\|_n^2 + \mathrm{pen}(\boldsymbol{m})\right), \quad \mathrm{pen}(\boldsymbol{m}) := (1+\theta)\,\sigma^2\frac{D_{\boldsymbol{m}}}{n},$$

$$\widehat{\mathcal{M}}_{n,\beta} := \left\{\boldsymbol{m}\in\mathbb{N}_+^p \,\middle|\, L(\boldsymbol{m})(\|\hat{\mathbf{G}}_{\boldsymbol{m}}^{-1}\|_{\mathrm{op}}\vee 1) \leqslant \beta\frac{n}{\log n}\right\}.$$

Using a fixed design result of [Baraud, 2000], we obtain the following oracle bound.

Theorem

If $\mathbb{E}|\varepsilon_1|^q$ is finite for some $q>6$, then there exists a constant $\alpha_{\beta,r'}>0$ such that for all $\alpha\in(0,\alpha_{\beta,r'})$, we have:

$$\mathbb{E}\|b-\hat{b}_{\hat{\boldsymbol{m}}}\|_n^2 \leqslant C(\theta)\inf_{\boldsymbol{m}\in\mathcal{M}_{n,\alpha}}\left(\inf_{t\in S_{\boldsymbol{m}}}\|b-t\|_\mu^2 + \sigma^2\frac{D_{\boldsymbol{m}}}{n}\right) + \sigma^2\frac{\Sigma(\theta,q)}{n} + o\left(\frac{1}{n}\right),$$

with $\Sigma(\theta,q) := C'(\theta,q)\frac{\mathbb{E}|\varepsilon_1|^q}{\sigma^q}\sum_{\boldsymbol{m}\in\mathbb{N}_+^p} D_{\boldsymbol{m}}^{-(\frac{q}{2}-2)}$.

Oracle bound for the prediction risk

Theorem

If A is compact:

If $\mathbb{E}|\varepsilon_1|^q$ is finite for some $q > 6$, then there exists $\beta^* > 0$ such that for all $\beta \in (0, \beta^*)$, there exists $\alpha_{\beta, r'} > 0$ such that for all $\alpha \in (0, \alpha_{\beta, r'})$, we have:

$$\mathbb{E}\|b - \hat{b}_{\hat{\boldsymbol{m}}}\|_\mu^2 \leqslant C(\theta, \beta, r) \inf_{\boldsymbol{m} \in \mathcal{M}_{n,\alpha}} \left(\inf_{t \in S_{\boldsymbol{m}}} \|b - t\|_\mu^2 + \sigma^2 \frac{D_{\boldsymbol{m}}}{n} \right) + C'(\beta, r)\, \sigma^2 \frac{\Sigma(\theta, q)}{n} + \mathrm{o}\left(\frac{1}{n}\right),$$

with:

$$\mathcal{M}_{n,\alpha} \coloneqq \left\{ \boldsymbol{m} \in \mathbb{N}_+^p \,\middle|\, L(\boldsymbol{m}) \left(\|\mathbf{G}_{\boldsymbol{m}}^{-1}\|_{\mathrm{op}} \vee 1 \right) \leqslant \alpha \frac{n}{\log n} \right\},$$

$$\widehat{\mathcal{M}}_{n,\beta} \coloneqq \left\{ \boldsymbol{m} \in \mathbb{N}_+^p \,\middle|\, L(\boldsymbol{m}) \left(\|\hat{\mathbf{G}}_{\boldsymbol{m}}^{-1}\|_{\mathrm{op}} \vee 1 \right) \leqslant \beta \frac{n}{\log n} \right\}.$$

Oracle bound for the prediction risk

Theorem

If** A **is not compact:

If $\mathbb{E}|\varepsilon_1|^q$ is finite for some $q > 6$, then there exists $\beta^* > 0$ such that for all $\beta \in (0, \beta^*)$, there exists $\alpha_{\beta,r'} > 0$ such that for all $\alpha \in (0, \alpha_{\beta,r'})$, we have:

$$\mathbb{E}\|b - \hat{b}_{\hat{\boldsymbol{m}}}\|_\mu^2 \leqslant C(\theta, \beta, r) \inf_{\boldsymbol{m} \in \mathcal{M}_{n,\alpha}} \left(\inf_{t \in S_{\boldsymbol{m}}} \|b - t\|_\mu^2 + \sigma^2 \frac{D_{\boldsymbol{m}}}{n} \right) + C'(\beta, r)\, \sigma^2 \frac{\Sigma(\theta, q)}{n} + \mathrm{o}\left(\frac{1}{n}\right),$$

with:

$$\mathcal{M}_{n,\alpha} \coloneqq \left\{ \boldsymbol{m} \in \mathbb{N}_+^p \;\middle|\; L(\boldsymbol{m}) \left(\|\mathbf{G}_{\boldsymbol{m}}^{-1}\|_{\mathrm{op}}^2 \vee 1 \right) \leqslant \alpha \frac{n}{\log n} \right\},$$

$$\widehat{\mathcal{M}}_{n,\beta} \coloneqq \left\{ \boldsymbol{m} \in \mathbb{N}_+^p \;\middle|\; L(\boldsymbol{m}) \left(\|\hat{\mathbf{G}}_{\boldsymbol{m}}^{-1}\|_{\mathrm{op}}^2 \vee 1 \right) \leqslant \beta \frac{n}{\log n} \right\}.$$

Conclusion and perspective

- We obtain a bound on the prediction risk by using concentration inequalities of [Gittens and Tropp, 2011] and [Tropp, 2012] on the eigenvalues of a random matrix.
- We improve the oracle bounds of [Baraud, 2002] and [Comte and Genon-Catalot, 2020].
- I think that these results can be extended to more general approximation spaces $(S_m)_{m\in\mathcal{M}_n}$, that are not constructed from an orthonormal basis.

General conclusion

- I use tensorized bases to construct projection estimators of multavariate functions in deconvolution and regression problems.

- Hypermatrices are a natural extension of matrices that allow me to study the MISE of projection estimators in a way that is similar to the one-dimensional case.

- The Goldenshluger and Lepski's method provides a general framework to construct adaptive estimators in this context.

- These techniques can be used to study more complex inverse problems in a multivariate setting.

General conclusion

- I use tensorized bases to construct projection estimators of multavariate functions in deconvolution and regression problems.
- Hypermatrices are a natural extension of matrices that allow me to study the MISE of projection estimators in a way that is similar to the one-dimensional case.
- The Goldenshluger and Lepski's method provides a general framework to construct adaptive estimators in this context.
- These techniques can be used to study more complex inverse problems in a multivariate setting.

Merci pour votre attention !

Asmussen, S. and Albrecher, H. (2010).
Ruin probabilities, volume 14 of *Advanced series on statistical science and applied probability*.
World Scientific, Singapore ; New Jersey, 2nd edition.

Baraud, Y. (2000).
Model selection for regression on a fixed design.
Probability Theory and Related Fields, 117(4):467–493.

Baraud, Y. (2002).
Model selection for regression on a random design.
ESAIM: Probability and Statistics, 6:127–146.

Bongioanni, B. and Torrea, J. L. (2009).
What is a Sobolev space for the Laguerre function systems?
Studia Mathematica, 192(2):147–172.

Chagny, G. (2013).
Warped bases for conditional density estimation.
Mathematical Methods of Statistics, 22(4):253–282.

Comte, F., Cuenod, C.-A., Pensky, M., and Rozenholc, Y. (2017).
Laplace deconvolution on the basis of time domain data and its application to dynamic contrast-enhanced imaging.
Journal of the Royal Statistical Society: Series B (Statistical Methodology), 79(1):69–94.

Comte, F. and Genon-Catalot, V. (2015).
Adaptive Laguerre density estimation for mixed Poisson models.
Electronic Journal of Statistics, 9(1):1113–1149.

Comte, F. and Genon-Catalot, V. (2020).
Regression function estimation as a partly inverse problem.
Annals of the Institute of Statistical Mathematics, 72(4):1023–1054.

Comte, F. and Lacour, C. (2013).
Anisotropic adaptive kernel deconvolution.
Annales de l'Institut Henri Poincaré, Probabilités et Statistiques, 49(2):569–609.

Dussap, F. (2022).
Nonparametric estimation of the expected discounted penalty function in the compound Poisson model.
Electronic Journal of Statistics, 16(1).

Gerber, H. U. and Shiu, E. S. (1998).
On the Time Value of Ruin.
North American Actuarial Journal, 2(1):48–72.

Gittens, A. and Tropp, J. A. (2011).
Tail bounds for all eigenvalues of a sum of random matrices.
arXiv:1104.4513 [math].

Goldenshluger, A. and Lepski, O. (2011).
Bandwidth selection in kernel density estimation: Oracle inequalities and adaptive minimax optimality.
The Annals of Statistics, 39(3):1608–1632.

Mabon, G. (2017).
Adaptive Deconvolution on the Non-negative Real Line: Adaptive deconvolution on $\mathbb{R}_+$.
Scandinavian Journal of Statistics, 44(3):707–740.

Pitts, S. M. (1994).
Nonparametric estimation of compound distributions with applications in insurance.
Annals of the Institute of Statistical Mathematics, 46(3):537–555.

Politis, K. (2003).
Semiparametric Estimation for Non-Ruin Probabilities.
Scandinavian Actuarial Journal, 2003(1):75–96.

Tropp, J. A. (2012).
User-Friendly Tail Bounds for Sums of Random Matrices.
Foundations of Computational Mathematics, 12(4):389–434.

Zhang, Z. and Su, W. (2018).
A new efficient method for estimating the Gerber–Shiu function in the classical risk model.
Scandinavian Actuarial Journal, 2018(5):426–449.