



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

Florian Dussap

► To cite this version:

Florian Dussap. Nonparametric Estimation of the Expected Discounted Penalty Function in the Compound Poisson Model. *Electronic Journal of Statistics*, 2022, 16 (1), pp.2124-2174. 10.1214/22-EJS2003 . hal-03212783v2

HAL Id: hal-03212783

<https://hal.science/hal-03212783v2>

Submitted on 17 Jan 2022

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

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

Florian Dussap*

Université de Paris, CNRS, MAP5 UMR 8145, F-75006 Paris, France

Abstract

We propose a nonparametric estimator of the expected discounted penalty function in the compound Poisson risk model. We use a projection estimator on the Laguerre basis and we compute the coefficients using Plancherel theorem. We provide an upper bound on the MISE of our estimator, and we show it achieves parametric rates of convergence on Sobolev–Laguerre spaces without needing a bias-variance compromise. Moreover, we compare our estimator with the Laguerre deconvolution method. We compute an upper bound of the MISE of the Laguerre deconvolution estimator and we compare it on Sobolev–Laguerre spaces with our estimator. Finally, we compare these estimators on simulated data.

MSC-2020: 62G05, 62P05, 91G70.

Keywords: nonparametric estimator, rate of estimation, classical risk model, Gerber-Shiu function, Laguerre series expansion, Fourier inversion.

1 Introduction

1.1 The statistical problem

We consider the classical risk model (compound Poisson model) for the risk reserve process $(U_t)_{t \geq 0}$ of an insurance company:

$$U_t = u + ct - \sum_{i=1}^{N_t} X_i, \quad t \geq 0 \quad (1)$$

where $u \geq 0$ is the initial capital; $c > 0$ is the premium rate; the claim number process (N_t) is a homogeneous Poisson process with intensity λ ; the claim sizes (X_i) are positive and i.i.d. with density f and mean μ , independent of (N_t) . We denote by $\tau(u)$ the ruin time:

$$\tau(u) := \inf \left\{ t \geq 0 \left| \sum_{i=1}^{N_t} X_i - ct > u \right. \right\} \in \mathbb{R}_+ \cup \{+\infty\}$$

and we make the following assumption to ensure that $\tau(u)$ is not almost surely finite.

Assumption 1 (safety loading condition). Let $\theta := \frac{\lambda\mu}{c}$, we assume that $\theta < 1$.

To study simultaneously the ruin time, the deficit at ruin, and the surplus level before the ruin, Gerber and Shiu [1998] introduced the function:

$$\phi(u) := \mathbb{E} \left[e^{-\delta\tau(u)} w(U_{\tau(u)-}, |U_{\tau(u)}|) \mathbf{1}_{\tau(u) < +\infty} \right], \quad (2)$$

where $\delta \geq 0$, and w is a non-negative function of the surplus before the ruin and the deficit at ruin. This function is called the *expected discounted penalty function*, but it will also be referred to as the *Gerber–Shiu function* in the following. For more information concerning the compound Poisson model and the Gerber–Shiu function, see Asmussen and Albrecher [2010].

*florian.dussap@gmail.com

Example 1.1. Several quantities of interest can be put in the form (2),

1. if $\delta = 0$ and $w(x, y) = 1$, then $\phi(u)$ is the probability of ruin;
2. if $\delta > 0$ and $w(x, y) = 1$, then $\phi(u)$ is the Laplace transform of $\tau(u)$;
3. if $\delta = 0$ and $w(x, y) = x + y$, then $\phi(u)$ is the expected jump size causing the ruin.

Observations and goal In all this article, we suppose that the premium rate c is known, but the parameters of the aggregate claims process are not, that is (λ, μ, f) is unknown. We suppose we have observed the process $(U_t)_{t \geq 0}$ during the interval $[0, T]$, with $T > 0$ fixed, so we have access to the number of claims and their size. Our goal is to recover the Gerber–Shiu function from the observations $(N_T; X_1, \dots, X_{N_T})$.

Several authors have considered the problem of estimating the Gerber–Shiu function using nonparametric methods. The first articles had an asymptotic approach: [Frees \[1986\]](#), [Croux and Veraverbeke \[1990\]](#), [Pitts \[1994\]](#), [Politis \[2003\]](#), and [Masiello \[2014\]](#) constructed nonparametric estimators of the ruin probability, and established their consistency and their asymptotic normality.

Concerning non-asymptotic approaches, [Mnatsakanov, Ruymgaart, and Ruymgaart \[2008\]](#) introduced a regularized Laplace inversion method to estimate the ruin probability in the compound Poisson model. [Shimizu \[2011, 2012\]](#) then extended this method to the estimation of the Gerber–Shiu function in more general risk models. However, this method suffers from poor rates of convergence, and numerical difficulties to compute the estimator.

In their paper, [Zhang and Su \[2018\]](#) introduced a projection estimator on the Laguerre basis to overcome these drawbacks. The choice of this basis is motivated by the work of [Comte, Cuenod, Pensky, and Rozenholc \[2017\]](#), where the properties of the Laguerre functions relative to the convolution product are used to solve a Laplace deconvolution problem. The same method was then used in more general risk models: [Zhang and Su \[2019\]](#) estimate the Gerber–Shiu function in a Lévy risk model, where the aggregate claims is a pure-jump Lévy process; [Su, Yong, and Zhang \[2019b\]](#) estimate the Gerber–Shiu function in the compound Poisson model perturbed by a Brownian motion; and [Su, Shi, and Wang \[2019a\]](#) study the model where both the income and the aggregate claims are compound Poisson processes. Recently, [Su and Yu \[2020\]](#) showed the Laguerre projection estimator of the Gerber–Shiu function in the compound Poisson model is pointwise asymptotically normal in the case $\delta = 0$.

In this paper, we construct an estimator of the Gerber–Shiu function (2) in the compound Poisson model (1). As [Zhang and Su \[2018\]](#), our estimator is a projection estimator on the Laguerre basis, but we compute the coefficients using Plancherel theorem instead of using a Laguerre deconvolution method. We emphasize that our estimator achieves parametric rates of convergence on Sobolev–Laguerre spaces regardless of the regularity of the Gerber–Shiu function, and without needing to find a compromise between the bias and the variance.

We also improve the previous results concerning the Laguerre deconvolution method. Previous rates were given in \mathcal{O}_P , and we propose a non-asymptotic bound on the MISE (Mean Integrated Squared Error) of the estimator. To achieve this goal, we introduce two modified versions of the Laguerre deconvolution estimator: the first one depends on a truncation parameter, whereas the second one does not, but it is only defined in the case $\delta = 0$.

To control the MISE of the second version of the Laguerre deconvolution estimator, we had to prove that the primitives of the Laguerre functions were uniformly bounded (see Lemma 3.4). This result is interesting in itself, the proof relies on the study of the properties of the ODE’s satisfied by Laguerre polynomials. The interested reader can find all the details in Appendix B.

Outline of the paper In the remaining part of this section, we introduce the notations and we give preliminary results on the Gerber–Shiu function. In Section 2, we construct our estimator and we study its MISE. In Section 3, we introduce two modified versions of the Laguerre deconvolution estimator and we study their MISE. In Section 4, we compute convergence rates of the different estimators considered on Sobolev–Laguerre spaces and also in the case where the claim sizes are exponentially distributed. In Section 5, we compare numerically the estimators on simulated data. We gathered all the proofs in Section 7.

1.2 Notations and preliminaries on the Gerber–Shiu function

We use the following notations in the paper:

- $\mathbb{N} := \{0, 1, 2, 3, \dots\}$, $\mathbb{N}^* := \mathbb{N} \setminus \{0\}$, $\mathbb{R}_+ := [0, +\infty)$, and $\mathbb{T} := \{z \in \mathbb{C} \mid |z| = 1\}$.
- “ $x \lesssim y$ ” means $x \leq Cy$ for an absolute constant $C > 0$.
- $x \wedge y := \min(x, y)$ and $x \vee y := \max(x, y)$.
- $\mathcal{L}f(s) := \int_0^{+\infty} e^{-sx} f(x) dx$ is the Laplace transform of f .
- $\mathcal{F}\psi(\omega) := \int_{\mathbb{R}} e^{i\omega x} \psi(x) dx$ is the Fourier transform of ψ .
- $\text{Leb}(A)$ is the Lebesgue measure of the set A .
- $\|\mathbf{A}_m\|_{\text{op}} := \sup_{x \in \mathbb{R}^m \setminus \{0\}} \frac{\|\mathbf{A}_m x\|_{\ell^2}}{\|x\|_{\ell^2}}$ is the ℓ^2 -operator norm of the matrix $\mathbf{A}_m \in \mathbb{R}^{m \times m}$.

The key result to estimate the Gerber–Shiu function is the following theorem.

Theorem 1.2 (Gerber and Shiu [1998]). *Under Assumption 1, the Gerber–Shiu function satisfies the equation:*

$$\phi = \phi * g + h. \quad (3)$$

where g and h are given by:

$$g(x) := \frac{\lambda}{c} \int_x^{+\infty} e^{-\rho_\delta(y-x)} f(y) dy, \quad h(u) := \frac{\lambda}{c} \int_u^{+\infty} e^{-\rho_\delta(x-u)} \left(\int_x^{+\infty} w(x, y-x) f(y) dy \right) dx \quad (4)$$

and where ρ_δ is the (unique) non-negative solution of the so-called Lundberg equation:

$$cs - \lambda(1 - \mathcal{L}f(s)) = \delta. \quad (5)$$

Remark 1.3. Since $\mathcal{L}f(s) \in [0, 1]$, it is easy to see that $\rho_\delta \in [\frac{\delta}{c}, \frac{\delta+\lambda}{c}]$. Moreover, we know that $\rho_\delta = 0$ when $\delta = 0$.

We need to ensure that ϕ , g and h belong to $L^2(\mathbb{R}_+)$ in order to use a projection estimator. We see that $\sup_x g(x) \leq \sup_x \frac{\lambda}{c} \mathbb{P}[X > x] \leq \frac{\lambda}{c}$ and $\int_0^\infty g(x) dx \leq \frac{\lambda}{c} \int_0^\infty \mathbb{P}[X > x] dx = \theta$, hence $g \in L^1(\mathbb{R}_+) \cap L^\infty(\mathbb{R}_+)$, therefore $g \in L^2(\mathbb{R}_+)$. To ensure that $h \in L^2(\mathbb{R}_+)$ we make the following assumption.

Assumption 2. We assume that $\int_0^\infty (1+x) \left(\int_x^\infty w(x, y-x) f(y) dy \right) dx$ is finite.

Under this assumption, we have:

$$\begin{aligned} \sup_{u>0} h(u) &\leq \frac{\lambda}{c} \int_0^\infty \left(\int_x^\infty w(x, y-x) f(y) dy \right) dx < +\infty, \\ \int_0^\infty h(u) du &\leq \frac{\lambda}{c} \int_0^\infty x \left(\int_x^\infty w(x, y-x) f(y) dy \right) dx < +\infty. \end{aligned}$$

Hence, h belongs to $L^1(\mathbb{R}_+) \cap L^\infty(\mathbb{R}_+)$, so $h \in L^2(\mathbb{R}_+)$. Integrating Equation (3) yields:

$$\|\phi\|_{L^1} = \int_0^\infty \phi(u) du = \frac{\int_0^\infty h(u) du}{1 - \int_0^\infty g(x) dx},$$

which is finite under Assumption 1 since $\int_0^\infty g(x) dx \leq \theta < 1$. Since ϕ belongs to $L^1(\mathbb{R}_+)$ and g belongs to $L^2(\mathbb{R}_+)$, their convolution product belongs to $L^2(\mathbb{R}_+)$, hence $\phi = \phi * g + h$ belongs to $L^2(\mathbb{R}_+)$ as well.

Remark 1.4. Assumption 2 has already been considered by Shimizu and Zhang [2017], and Zhang and Su [2018]. Actually, the quantity:

$$\omega(x) := \int_x^{+\infty} w(x, y - x) f(y) dy = \mathbb{E}[w(x, X - x) \mathbf{1}_{X \geq x}],$$

can be found on several occasion in the study of the Gerber–Shiu function. The assumption that $\int_0^\infty \omega(x) dx$ is finite ensures that $\phi(u)$ is finite for all u [Asmussen and Albrecher, 2010, Chapter X, Section 1]. The additional requirement that $\int_0^\infty x \omega(x) dx$ is finite serves to prove that ϕ belongs to $L^1(\mathbb{R}_+)$, so that its Fourier transform is well defined. As we have seen, it also ensures that ϕ belongs to $L^2(\mathbb{R}_+)$.

2 The Laguerre–Fourier estimator

We use the Laguerre functions $(\psi_k)_{k \in \mathbb{N}}$ as an orthonormal basis of $L^2(\mathbb{R}_+)$:

$$\forall x \in \mathbb{R}_+, \psi_k(x) := \sqrt{2} L_k(2x) e^{-x}, \quad L_k(x) := \sum_{j=0}^k \binom{k}{j} \frac{(-x)^j}{j!}. \quad (6)$$

We choose this basis for several reasons. First, the support of the Laguerre functions is \mathbb{R}_+ , which is well suited since the functions we want to estimate are defined on \mathbb{R}_+ . Moreover, exponential functions (and more broadly mixtures of gamma functions, see the proof of Lemma 3.9 in Mabon [2017]) have an exponentially small bias in this basis, which is interesting because when the claim sizes distribution is exponential and w is a polynomial, then g and h will be given by products of polynomials with exponentials. Finally, the Fourier transform of the Laguerre function is known explicitly:

$$\forall \omega \in \mathbb{R}, \quad \mathcal{F}\psi_k(\omega) = (-1)^k \sqrt{2} \frac{(1 + i\omega)^k}{(1 - i\omega)^{k+1}}, \quad (7)$$

which is helpful for the computation of the estimated coefficients (8).

We denote by $(a_k)_{k \geq 0}$ the Laguerre coefficients of ϕ . If $m \in \mathbb{N}^*$, we denote by ϕ_m the projection of ϕ on the subspace of $L^2(\mathbb{R}_+)$ spanned by the first m Laguerre functions $\psi_0, \dots, \psi_{m-1}$, that is:

$$\phi_m := \sum_{k=0}^{m-1} a_k \psi_k, \quad \text{with } a_k = \langle \phi, \psi_k \rangle.$$

The Laguerre coefficients of ϕ can be computed using Plancherel theorem:

$$a_k = \langle \phi, \psi_k \rangle = \frac{1}{2\pi} \langle \mathcal{F}\phi, \mathcal{F}\psi_k \rangle.$$

Taking the Fourier transform in equation (3), we see that $\mathcal{F}\phi = \frac{\mathcal{F}h}{1 - \mathcal{F}g}$. Let $\hat{g}, \hat{h} \in L^2(\mathbb{R}_+)$ be some estimators of g and h (we provide these estimators later in equation (14)), we estimate the coefficients of ϕ by:

$$\hat{a}_k := \frac{1}{2\pi} \left\langle \frac{\mathcal{F}\hat{h}}{1 - \widehat{\mathcal{F}g}}, \mathcal{F}\psi_k \right\rangle, \quad (8)$$

where $\widehat{\mathcal{F}g} := (\mathcal{F}\hat{g}) \mathbf{1}_{|\mathcal{F}\hat{g}| \leq \theta_0}$ for some truncation parameter $\theta_0 < 1$. The estimator of ϕ is then:

$$\hat{\phi}_{m_1} := \sum_{k=0}^{m_1-1} \hat{a}_k \psi_k,$$

where m_1 is the dimension of the projection space.

Proposition 2.1. *Under Assumptions 1 and 2, if $\theta < \theta_0$, we have:*

$$\|\phi - \hat{\phi}_{m_1}\|_{L^2}^2 \leq \|\phi - \phi_{m_1}\|_{L^2}^2 + \frac{2}{(1 - \theta_0)^2} \|h - \hat{h}\|_{L^2}^2 + \frac{2 \|h\|_{L^1}^2}{(1 - \theta_0)^2 (1 - \theta)^2} \left(1 + \frac{\|g\|_{L^1}^2}{(\theta_0 - \theta)^2} \right) \|g - \hat{g}\|_{L^2}^2.$$

Remark 2.2. We emphasize the fact that this result is proven using only two properties: the function ϕ satisfies the equation (3) and $\theta_0 > \theta > \|g\|_{L^1}$. Hence, it can be applied to other problems where the target function satisfies an equation of the form (3). For example, it is the case in Zhang and Su [2019], Su et al. [2019a] and Su et al. [2019b].

We now need to provide good estimators of g and h . We choose to estimate them by projection on the Laguerre basis too. Let $(b_k)_{k \geq 0}$ and $(c_k)_{k \geq 0}$ be the coefficients of g and h , that is:

$$g = \sum_{k=0}^{+\infty} b_k \psi_k, \quad \text{with } b_k := \langle g, \psi_k \rangle, \quad (9)$$

$$h = \sum_{k=0}^{+\infty} c_k \psi_k, \quad \text{with } c_k := \langle h, \psi_k \rangle. \quad (10)$$

By Fubini's theorem and using equation (4):

$$\begin{aligned} b_k &= \int_0^{+\infty} g(x) \psi_k(x) dx \\ &= \frac{\lambda}{c} \int_0^{+\infty} \left(\int_x^{+\infty} e^{-\rho_\delta(y-x)} f(y) dy \right) \psi_k(x) dx \\ &= \frac{\lambda}{c} \int_0^{+\infty} \left(\int_0^y e^{-\rho_\delta(y-x)} \psi_k(x) dx \right) f(y) dy \\ &= \frac{\lambda}{c} \mathbb{E} \left[\int_0^X e^{-\rho_\delta(X-x)} \psi_k(x) dx \right]. \end{aligned}$$

The same calculation for c_k yields:

$$c_k = \frac{\lambda}{c} \mathbb{E} \left[\int_0^X \left(\int_u^X e^{-\rho_\delta(x-u)} w(x, X-x) dx \right) \psi_k(u) du \right].$$

We estimate these coefficients by empirical means. However, we first need to estimate ρ_δ . Since ρ_δ is the non-negative solution of the Lundberg equation (5), we estimate it by $\hat{\rho}_\delta$ the non-negative solution of the empirical Lundberg equation:

$$cs - \hat{\lambda}(1 - \widehat{\mathcal{L}}f(s)) = \delta, \quad (11)$$

where $\hat{\lambda} := \frac{N_T}{T}$ and $\widehat{\mathcal{L}}f(s) := \frac{1}{N_T} \sum_{i=1}^{N_T} e^{-sX_i}$. When $\delta = 0$ we know that $\rho_\delta = 0$ so we do not need to estimate it, thus we set $\hat{\rho}_0 = 0$. The estimated coefficients of g and h are:

$$\hat{b}_k = \frac{1}{cT} \sum_{i=1}^{N_T} \int_0^{X_i} e^{-\hat{\rho}_\delta(X_i-x)} \psi_k(x) dx, \quad (12)$$

$$\hat{c}_k = \frac{1}{cT} \sum_{i=1}^{N_T} \int_0^{X_i} \left(\int_u^{X_i} e^{-\hat{\rho}_\delta(x-u)} w(x, X_i-x) dx \right) \psi_k(u) du, \quad (13)$$

and the estimators of g and h are:

$$\hat{g}_{m_2} := \sum_{k=0}^{m_2-1} \hat{b}_k \psi_k, \quad \hat{h}_{m_3} := \sum_{k=0}^{m_3-1} \hat{c}_k \psi_k, \quad (14)$$

where m_2 and m_3 are the dimensions of the projection spaces. As we did for ϕ , we denote by g_{m_2} and h_{m_3} the projections of g and h on the subspaces $\text{Span}(\psi_0, \dots, \psi_{m_2-1})$ and $\text{Span}(\psi_0, \dots, \psi_{m_3-1})$.

Remark 2.3. The dimensions m_1, m_2, m_3 do not have to be the same for the estimation of ϕ, g and h . In practice, we will choose different dimensions.

In order to give a bound on the mean integrated squared error of our estimators \hat{g}_{m_2} and \hat{h}_{m_3} , we need to make an additional assumption.

Assumption 3. Let $W(X) := \int_0^X \left(\int_u^X w(x, X-x) dx \right)^2 du$. If $\delta = 0$, we assume that $\mathbb{E}[W(X)]$ is finite, and if $\delta > 0$, we assume that $\mathbb{E}[W(X)^2]$ is finite.

Remark 2.4 (Applicability of Assumptions 2 and 3). Assumptions 2 and 3 can be thought as moment conditions on the claim sizes distribution, with respect to w . In the special case where w is given by $w(x, y) = x^k(x+y)^\ell$ for $k, \ell \geq 0$, we have:

$$\int_0^{+\infty} (1+x) \left(\int_x^{+\infty} w(x, y-x) f(y) dy \right) dx = \mathbb{E} \left[\frac{X^{k+\ell+1}}{k+1} \right] + \mathbb{E} \left[\frac{X^{k+\ell+2}}{k+2} \right],$$

$$W(X) = \frac{X^{2k+2\ell+3}}{(k+1)^2(2k+3)},$$

so Assumptions 2 and 3 reduce to the moment condition $\mathbb{E}[X^{2k+2\ell+3}] < +\infty$ (if $\delta = 0$). Notice that the functions of Example 1.1 correspond to the cases $(k, \ell) = (0, 0)$ or $(0, 1)$, so that the corresponding moment condition is $\mathbb{E}[X^3] < \infty$ or $\mathbb{E}[X^5] < \infty$. Hence, heavy-tailed distributions can fit into these assumptions, provided they admit sufficiently large moments. On the other hand, if w grows with an exponential rate, for example if $w(x, y-x) := \exp(\gamma(x+y))$, then we also need an exponential moment for X , so that we are restricted to light tailed distributions.

Theorem 2.5. Under Assumptions 1, 2 and 3, if $\delta = 0$ then it holds:

$$\mathbb{E}\|g - \hat{g}_{m_2}\|_{L^2}^2 \leq \|g - g_{m_2}\|_{L^2}^2 + \frac{\lambda}{c^2 T} \mathbb{E}[X],$$

$$\mathbb{E}\|h - \hat{h}_{m_3}\|_{L^2}^2 \leq \|h - h_{m_3}\|_{L^2}^2 + \frac{\lambda}{c^2 T} \mathbb{E}[W(X)],$$

and if $\delta > 0$ then it holds:

$$\mathbb{E}\|g - \hat{g}_{m_2}\|_{L^2}^2 \leq \|g - g_{m_2}\|_{L^2}^2 + \frac{C(\lambda)}{c^2 T} \left(\mathbb{E}[X] + \frac{\mathbb{E}[X^2]^{\frac{1}{2}}}{(1-\theta)^2 \delta^2} \right),$$

$$\mathbb{E}\|h - \hat{h}_{m_3}\|_{L^2}^2 \leq \|h - h_{m_3}\|_{L^2}^2 + \frac{C(\lambda)}{c^2 T} \left(\mathbb{E}[W(X)] + \frac{\mathbb{E}[W(X)^2]^{\frac{1}{2}}}{(1-\theta)^2 \delta^2} \right),$$

where $C(\lambda)$ is a $\mathcal{O}(\lambda^2)$.

Remark 2.6. The variance terms do not depend on m_2 nor m_3 , so no compromise between the bias and the variance is needed: we just have to take m_2 and m_3 as large as possible such that the bias is smaller than $1/T$. See Section 4 for a discussion concerning the choice of m_2 and m_3 when the functions g and h belong to a Sobolev–Laguerre space.

Let $m_1, m_2, m_3 \in \mathbb{N}^*$, we estimate g by \hat{g}_{m_2} and h by \hat{h}_{m_3} . We plug these estimators in (8) and we estimate ϕ by:

$$\hat{\phi}_{m_1, m_2, m_3} := \sum_{k=0}^{m_1-1} \left\langle \frac{\mathcal{F}\hat{h}_{m_3}}{1 - \widehat{\mathcal{F}g_{m_2}}}, \mathcal{F}\psi_k \right\rangle \psi_k,$$

with $\widehat{\mathcal{F}g_{m_2}} := \mathcal{F}\hat{g}_{m_2} \mathbf{1}_{|\mathcal{F}\hat{g}_{m_2}| \leq \theta_0}$. Combining Proposition 2.1 with Theorem 2.5, we obtain:

Corollary 2.7. Under Assumptions 1, 2 and 3, if $\theta < \theta_0$ then it holds:

$$\mathbb{E}\|\phi - \hat{\phi}_{m_1, m_2, m_3}\|_{L^2}^2 \leq \|\phi - \phi_{m_1}\|_{L^2}^2 + \frac{C}{(1-\theta_0)^2} \left(\|g - g_{m_2}\|_{L^2}^2 + \|h - h_{m_3}\|_{L^2}^2 + \frac{1}{cT} \right),$$

where C is a constant depending on $\lambda, c, \theta, \|g\|_{L^1}, \|h\|_{L^1}, \mathbb{E}[X], \mathbb{E}[W(X)]$ and $\theta_0 - \theta$; and also $\delta, \mathbb{E}[X^2], \mathbb{E}[W(X)^2]$ if $\delta > 0$.

We want to compare our estimator with the Laguerre deconvolution method. However, there is no result on the MISE of this method for estimating the Gerber–Shiu function, so we study it in the next section.

3 The Laguerre deconvolution estimator

For the Laguerre deconvolution method, we need an additional assumption on the coefficients of g .

Assumption 4. The coefficients $(b_k)_{k \geq 0}$, defined by (9), satisfy $(b_{k+1} - b_k)_{k \geq 0} \in \ell^1(\mathbb{N})$.

Remark 3.1. If g belongs to a Sobolev–Laguerre space $W^s(\mathbb{R}_+)$ with regularity $s > 1$, then Assumption 4 holds automatically. The spaces $W^s(\mathbb{R}_+)$ are regularity spaces associated with the Laguerre basis, see Definition 4.1 below. Indeed, by the Cauchy–Schwarz inequality, we have:

$$\sum_{k=0}^{+\infty} |b_k| = \sum_{k=0}^{+\infty} |b_k| (1+k)^{\frac{s}{2}} (1+k)^{-\frac{s}{2}} \leq \left(\sum_{k=0}^{+\infty} |b_k|^2 (1+k)^s \right)^{\frac{1}{2}} \left(\sum_{k=0}^{+\infty} (1+k)^{-s} \right)^{\frac{1}{2}},$$

which is finite if $g \in W^s(\mathbb{R}_+)$ and $s > 1$. Hence, $(b_k)_{k \geq 0}$ is in $\ell^1(\mathbb{N})$ and so is $(b_{k+1} - b_k)_{k \geq 0}$.

The reason why the Laguerre basis is well suited for deconvolution on \mathbb{R}_+ is the following relation satisfied by the Laguerre functions:

$$\forall k, j \in \mathbb{N}, \quad \psi_k * \psi_j = \frac{1}{\sqrt{2}} (\psi_{k+j} - \psi_{k+j+1}),$$

see formula 22.13.14 in Abramowitz and Stegun [1972]. The reader interested in the use of the Laguerre basis for deconvolution problems is referred to Mabon [2017]. Expanding the renewal equation (3) on the Laguerre basis, one easily obtains the following relation between the coefficients of ϕ , g and h :

$$\forall k \in \mathbb{N}, \quad a_k = (\beta * a)_k + c_k,$$

where the sequence $(\beta_k)_{k \geq 0}$ is defined by $\beta_0 := \frac{b_0}{\sqrt{2}}$ and $\beta_k := \frac{b_k - b_{k-1}}{\sqrt{2}}$ for $k \geq 1$. This relation can be written in a matrix form: if $\mathbf{a}_m := (a_0, \dots, a_{m-1})^T$ and $\mathbf{c}_m := (c_0, \dots, c_{m-1})^T$ are the vectors of the m first coefficients of ϕ and h , then it holds:

$$\mathbf{A}_m \times \mathbf{a}_m = \mathbf{c}_m \iff \mathbf{a}_m = \mathbf{A}_m^{-1} \times \mathbf{c}_m, \quad (15)$$

where \mathbf{A}_m is the lower triangular Toeplitz matrix:

$$\forall i, j \in \{0, \dots, m-1\}, \quad (\mathbf{A}_m)_{i,j} := \begin{cases} 1 - \frac{1}{\sqrt{2}} b_0 & \text{if } i = j, \\ \frac{1}{\sqrt{2}} (b_{i-j-1} - b_{i-j}) & \text{if } i > j, \\ 0 & \text{else.} \end{cases} \quad (16)$$

This matrix is invertible if and only if $1 - \frac{b_0}{\sqrt{2}} \neq 1$, which is the case because $\frac{b_0}{\sqrt{2}} \leq \theta < 1$ under Assumption 1.

Lemma 3.2. Under Assumption 4, we have $\|\mathbf{A}_m^{-1}\|_{\text{op}} \leq \frac{2}{1-\theta}$ for all $m \in \mathbb{N}^*$.

This lemma is borrowed from Zhang and Su [2018] (Lemma 4.3 in their article). There were missing elements in their proof, so we give a new proof of this lemma, for the sake of completeness.

The naive Laguerre deconvolution estimator consists in estimating the matrix \mathbf{A}_m and the coefficients \mathbf{c}_m in (15), to obtain an estimation of the coefficients of ϕ . More precisely, the matrix \mathbf{A}_m is estimated by plugging \hat{b}_k , defined by (12), in (16):

$$\forall i, j \in \{0, \dots, m-1\}, \quad (\hat{\mathbf{A}}_m)_{i,j} := \begin{cases} 1 - \frac{1}{\sqrt{2}} \hat{b}_0 & \text{if } i = j, \\ \frac{1}{\sqrt{2}} (\hat{b}_{i-j-1} - \hat{b}_{i-j}) & \text{if } i > j, \\ 0 & \text{else.} \end{cases} \quad (17)$$

This matrix is invertible if and only if $\frac{\hat{b}_0}{\sqrt{2}} \neq 1$, which is almost surely the case since $\frac{\hat{b}_0}{\sqrt{2}} = \frac{1}{cT} \sum_{i=1}^{N_T} (1 - e^{-X_i})$ is a continuous random variable. The coefficients of ϕ are estimated by:

$$\hat{\mathbf{a}}_m^{\text{Lag}_0} := \hat{\mathbf{A}}_m^{-1} \times \hat{\mathbf{c}}_m, \quad (18)$$

where $\hat{\mathbf{c}}_m := (\hat{c}_0, \dots, \hat{c}_{m-1})^T$. Under Assumptions 1, 2, 3, and 4, Zhang and Su [2018] show that if $\mathbb{E}[X^2]$ is finite and if $m = o(T)$, then $\|\phi - \hat{\phi}_m^{\text{Lag}_0}\|_{L^2}^2 \leq \|\phi - \phi_m\|_{L^2}^2 + \mathcal{O}_P(\frac{m}{T})$.

In the following, we propose two ways inspired by Comte and Mabon [2017] to estimate the Gerber–Shiu function, using the Laguerre deconvolution method. To obtain a non asymptotic result on the MISE of the estimator, a cutoff is required when inverting the matrix $\hat{\mathbf{A}}_m$. Let $\theta_0 < 1$ be a truncation parameter, we estimate \mathbf{A}_m^{-1} by:

$$\tilde{\mathbf{A}}_{m,1}^{-1} := \hat{\mathbf{A}}_m^{-1} \mathbf{1}_{\Delta_m^1} \quad \text{where} \quad \Delta_m^1 := \left\{ \|\hat{\mathbf{A}}_m^{-1}\|_{\text{op}} \leq \frac{2}{1-\theta_0} \right\},$$

and we estimate the coefficients \mathbf{a}_m by $\hat{\mathbf{a}}_m^{\text{Lag}_1} := \tilde{\mathbf{A}}_{m,1}^{-1} \times \hat{\mathbf{c}}_m$.

Theorem 3.3. *Under Assumptions 1, 2, 3, and 4, if $\theta < \theta_0$ then it holds:*

$$\forall m \in \mathbb{N}^*, \quad \mathbb{E}\|\phi - \hat{\phi}_m^{\text{Lag}_1}\|_{L^2}^2 \leq \|\phi - \phi_m\|_{L^2}^2 + C \frac{m}{T},$$

where C is a constant depending on λ , c , θ , $\mathbb{E}[X]$, $\mathbb{E}[W(X)]$ and $\theta_0 - \theta$; and also δ , $\mathbb{E}[X^2]$, $\mathbb{E}[W(X)^2]$ if $\delta > 0$.

We propose a second way to estimate ϕ using the Laguerre deconvolution method, in the case $\delta = 0$. It avoids the use of a truncation parameter θ_0 , but at the expense of an extra $\log(m)$ factor in the upper bound of the MISE, and it uses an additional independence assumption. We estimate the Laguerre coefficients of g by (12), that is in this case:

$$\hat{b}_k := \frac{1}{cT} \sum_{i=1}^{N_T} \Psi_k(X_i),$$

where $\Psi_k(x) := \int_0^x \psi_k(t) dt$. The matrix \mathbf{A}_m is still estimated by (17).

Lemma 3.4. *The functions $\Psi_k(x) := \int_0^x \psi_k(t) dt$ are uniformly bounded.*

This lemma is a technical result interesting in itself and we prove it in Appendix B. Using this lemma, we can control the risk of $\hat{\mathbf{A}}_m$ in operator norm.

Proposition 3.5. *If $\delta = 0$, $p \geq 1$ and $\log m \geq p$, then it holds:*

$$\mathbb{E}\left[\|\hat{\mathbf{A}}_m - \mathbf{A}_m\|_{\text{op}}^{2p}\right] \leq C(p, \lambda) \mu^p \left(\frac{m \log m}{cT}\right)^p + C(p) \left(\frac{m \log m}{cT}\right)^{2p},$$

where $C(p, \lambda)$ is a $\mathcal{O}(\lambda^p)$, and $C(p)$ is a constant depending on p .

This time, we estimate the inverse of the matrix \mathbf{A}_m by:

$$\tilde{\mathbf{A}}_{m,2}^{-1} := \hat{\mathbf{A}}_m^{-1} \mathbf{1}_{\Delta_m^2}, \quad \text{where} \quad \Delta_m^2 := \left\{ \|\hat{\mathbf{A}}_m^{-1}\|_{\text{op}}^2 \leq \frac{cT}{m \log m} \right\},$$

we estimate the coefficients of ϕ by $\hat{\mathbf{a}}_m^{\text{Lag}_2} := \tilde{\mathbf{A}}_{m,2}^{-1} \times \hat{\mathbf{c}}_m$, and we estimate ϕ by:

$$\hat{\phi}_m^{\text{Lag}_2} := \sum_{k=0}^{m-1} \hat{a}_k^{\text{Lag}_2} \psi_k.$$

To provide an upper bound on the MISE of $\hat{\phi}_m^{\text{Lag}_2}$, we need $\tilde{\mathbf{A}}_{m,2}^{-1}$ and $\hat{\mathbf{c}}_m$ to be independent. For this reason, we assume that we have a second observation set $\{N'_T; X'_1, \dots, X'_{N'_T}\}$ identical in law but independent from the main one¹. We use this second set to estimate $\tilde{\mathbf{A}}_{m,2}^{-1}$.

¹alternatively, we could split the data $\{X_1, \dots, X_{N_T}\}$ in two parts: we use half of the data to estimate $\tilde{\mathbf{A}}_{m,2}^{-1}$, and the other half to estimate $\hat{\mathbf{c}}_m$.

Theorem 3.6. *We assume that $\delta = 0$. Under Assumptions 1, 2, 3 and 4, if $m \log m \leq cT$ then it holds:*

$$\mathbb{E} \|\phi - \hat{\phi}_m^{\text{Lag}_2}\|_{L^2}^2 \leq \|\phi - \phi_m\|_{L^2}^2 + \frac{C(\lambda)}{cT(1-\theta)^2} \left(\frac{\mathbb{E}[W(X)]}{c} + \|\phi\|_{L^2}^2 (\mu + \mu^2) m \log m \right) + \mathcal{O}\left(\frac{1}{T^2}\right),$$

with $C(\lambda) = \mathcal{O}(\lambda \vee \lambda^2)$.

Remark 3.7. Contrary to the Laguerre–Fourier method, there is only one bias term with the Laguerre deconvolution method. However, the variance term is more complicated and a bias-variance compromise is needed. It leads to non-parametric rates of convergence, which are slower than the parametric rate $\frac{1}{T}$.

4 Convergence rates of the Laguerre estimators

4.1 Sobolev–Laguerre spaces

To study the bias of a function in the Laguerre basis, we consider the Sobolev–Laguerre spaces. These functional spaces have been introduced by Bongioanni and Torrea [2009] to study the Laguerre operator. The connection with the Laguerre coefficients was established later by Comte and Genon-Catalot [2015].

Definition 4.1. For $s > 0$, we define the Sobolev–Laguerre ball of radius $L > 0$ and regularity s as:

$$W^s(\mathbb{R}_+, L) := \left\{ v \in L^2(\mathbb{R}_+) \left| \sum_{k=0}^{+\infty} \langle v, \psi_k \rangle^2 k^s \leq L \right. \right\},$$

and we define the Sobolev–Laguerre space of regularity s as $W^s(\mathbb{R}_+) := \bigcup_{L>0} W^s(\mathbb{R}_+, L)$.

By Proposition 7.2 in Comte and Genon-Catalot [2015], when s is a natural number, $v \in W^s(\mathbb{R}_+)$ if and only if v is $(s-1)$ times differentiable, $v^{(s-1)}$ is absolutely continuous, and for all $0 \leq k \leq s-1$ we have $x^{\frac{k+1}{2}} \sum_{j=0}^{k+1} \binom{k+1}{j} v^{(j)} \in L^2(\mathbb{R}_+)$. In particular, a function v belongs to $W^1(\mathbb{R}_+)$ if and only if it is absolutely continuous and $\sqrt{x}(v + v') \in L^2(\mathbb{R}_+)$.

We are interested in the Sobolev–Laguerre spaces because of the following observation. If v belongs to a Sobolev–Laguerre ball $W^s(\mathbb{R}_+, L)$, then its bias is controlled by:

$$\|v - v_m\|_{L^2}^2 = \sum_{k=m}^{+\infty} \langle v, \psi_k \rangle^2 = \sum_{k=m}^{+\infty} \langle v, \psi_k \rangle^2 k^s k^{-s} \leq L m^{-s}.$$

Combining this upper bound on the bias term with Corollary 2.7, and Theorems 3.3 and 3.6, we obtain convergence rates for the Laguerre–Fourier estimator and the Laguerre deconvolution estimators, on Sobolev–Laguerre spaces.

Theorem 4.2. *Under Assumptions 1, 2 and 3, if $\theta < \theta_0$ and if $\phi \in W^{s_1}(\mathbb{R}_+)$, $g \in W^{s_2}(\mathbb{R}_+)$ and $h \in W^{s_3}(\mathbb{R}_+)$, then choosing $m_i > (cT)^{\frac{1}{s_i}}$ for all $i \in \{1, 2, 3\}$ yields:*

$$\mathbb{E} \|\phi - \hat{\phi}_{m_1, m_2, m_3}\|_{L^2}^2 = \mathcal{O}\left(\frac{1}{cT}\right).$$

Remark 4.3. If ϕ , g and h belong to some Sobolev–Laguerre spaces with a regularity index greater than 1, we can just choose $m_1 = m_2 = m_3 = \lceil cT \rceil$ and obtain the parametric rate $\mathcal{O}(\frac{1}{cT})$ for the Laguerre–Fourier estimator.

Theorem 4.4. *We make Assumptions 1, 2, 3 and 4, and we assume that $\phi \in W^s(\mathbb{R}_+)$.*

1. *If $\theta < \theta_0$, then choosing $m_{\text{opt}} \propto (cT)^{\frac{1}{1+s}}$ yields:*

$$\mathbb{E} \|\phi - \hat{\phi}_{m_{\text{opt}}}^{\text{Lag}_1}\|_{L^2}^2 = \mathcal{O}\left((cT)^{-\frac{s}{1+s}}\right).$$

2. If $\delta = 0$, then choosing $m_{\text{opt}} \propto (cT)^{\frac{1}{1+s}}$ yields:

$$\mathbb{E} \|\phi - \hat{\phi}_{m_{\text{opt}}}^{\text{Lag}_2}\|_{L^2}^2 = \mathcal{O}\left((cT)^{-\frac{s}{1+s}} \log(cT)\right).$$

Remark 4.5. The Fourier–Laguerre estimator and the Laguerre deconvolution estimator $\hat{\phi}_m^{\text{Lag}_1}$ both depend on a truncation parameter θ_0 that needs to be chosen such that $\theta < \theta_0$. We see two ways to ensure that.

1. We can assume that we know some $\theta_0 < 1$ such that $\theta < \theta_0$. Then our convergence rates are those of Theorems 4.2 and 4.4.
2. We can choose $\theta_0 = 1 - (\log T)^{1/2}$. Then for T large enough (more precisely $T > e^{(1-\theta)^2}$), the convergence rates of the Laguerre–Fourier estimator and $\hat{\phi}_m^{\text{Lag}_1}$ are those of Theorems 4.2 and 4.4 multiplied by $\log(T)$.

In our simulations, we chose the first way.

4.2 The exponential case

In this section, we want to compute the convergence rate of the estimators, in the exponential case: $X \sim \text{Exp}(1/\mu)$. This distribution is often considered in risk theory and closed forms of the Gerber–Shiu function are available in this case. Indeed, the Gerber–Shiu functions of Example 1.1 are given by:

$$\phi(u) = \begin{cases} \theta \exp\left(-\frac{1-\theta}{\mu}u\right) & \text{(ruin probability),} \\ \frac{\theta}{1+\mu\rho_\delta} \exp\left(-\left[\frac{1-\theta}{\mu} + \rho_\delta - \frac{\delta}{c}\right]u\right) & \text{(Laplace transform of the ruin time),} \\ \mu(1+2\theta) \exp\left(-\frac{1-\theta}{\mu}u\right) - \mu \exp\left(-\frac{u}{\mu}\right) & \text{(jump size causing the ruin).} \end{cases} \quad (19)$$

These formulas are obtained by Laplace inversion, see Asmussen and Albrecher [2010], chapter XII. We use the following lemma to compute the bias terms of the functions ϕ , g and h .

Lemma 4.6. Let C, γ be positive numbers and let $F(x) = C \exp(-\gamma x) \mathbf{1}_{\mathbb{R}_+}(x)$. The Laguerre coefficients of F are given by:

$$\langle F, \psi_k \rangle = \frac{C\sqrt{2}}{\gamma+1} \left(\frac{\gamma-1}{\gamma+1}\right)^k.$$

Hence if $m \geq 0$ we have:

$$\sum_{k=m}^{+\infty} \langle F, \psi_k \rangle^2 = \frac{C^2}{2\gamma} \left(\frac{\gamma-1}{\gamma+1}\right)^{2m}.$$

Proposition 4.7. If the density of X is $f(x) = \frac{1}{\mu} e^{-\frac{x}{\mu}}$, then the bias term of ϕ is given by:

$$\|\phi - \phi_m\|_{L^2}^2 \leq L e^{-rm}$$

with L and r given by:

1. Ruin probability: $L = \frac{\theta^2}{2\gamma}$, $r = 2 \log \left| \frac{\gamma+1}{\gamma-1} \right|$ and $\gamma = \frac{1-\theta}{\mu}$.
2. Laplace transform of the ruin time: $L = \frac{\theta^2}{2\gamma(1+\mu\rho_\delta)^2}$, $r = 2 \log \left| \frac{\gamma+1}{\gamma-1} \right|$ and $\gamma = \frac{1-\theta}{\mu} + \rho_\delta - \frac{\delta}{c}$.
3. Jump size causing the ruin: $L = \mu^3 \frac{(1+2\theta)^2}{1-\theta}$, $r = 2 \log \left(\left| \frac{1-\theta+\mu}{1-\theta-\mu} \right| \wedge \left| \frac{1+\mu}{1-\mu} \right| \right)$.

Combining this Proposition with Theorems 3.3 and 3.6, we easily obtain convergence rates for the Gerber–Shiu functions we are interested in.

Theorem 4.8. We assume that the density of X is $f(x) = \frac{1}{\mu}e^{-\frac{x}{\mu}}$, we make Assumptions 1, 2, 3 and 4, and we assume that the bias term of ϕ decreases as:

$$\|\phi - \phi_m\|_{L^2}^2 \leq Le^{-rm}.$$

1. If $\theta < \theta_0$, then choosing $m_{\text{opt}} = \lceil \frac{1}{r} \log(cT) \rceil$ yields:

$$\mathbb{E}\|\phi - \hat{\phi}_{m_{\text{opt}}}^{\text{Lag}_1}\|^2 = \mathcal{O}\left(\frac{\log(cT)}{cT}\right).$$

2. If $\delta = 0$, then choosing $m_{\text{opt}} = \lceil \frac{1}{r} \log(cT) \rceil$ yields:

$$\mathbb{E}\|\phi - \hat{\phi}_{m_{\text{opt}}}^{\text{Lag}_2}\|_{L^2}^2 = \mathcal{O}\left(\frac{\log(cT) \log \log(cT)}{cT}\right).$$

For the Laguerre–Fourier estimator, we also need to know the decreasing rate of the bias term of g and h . For the ruin probability, the Laplace transform of τ , and the jump size causing the ruin, direct calculations show that g and h are given by a positive multiple of $e^{-x/\mu}$. Thus, Lemma 4.6 yields that their bias term is less than $\exp(-r'm)$, with $r' := 2 \log|\frac{1+\mu}{1-\mu}|$. Together with Corollary 2.7, we obtain the convergence rates of the Laguerre–Fourier estimator.

Theorem 4.9. If the density of X is $f(x) = \frac{1}{\mu}e^{-\frac{x}{\mu}}$, under Assumptions 1, 2 and 3, if $\theta < \theta_0$ and if the bias term of ϕ decreases as:

$$\|\phi - \phi_m\|_{L^2}^2 \leq Le^{-rm},$$

then choosing $m_1 > \lceil \frac{1}{r} \log(cT) \rceil$ and $m_2, m_3 > \lceil \frac{1}{r'} \log(cT) \rceil$ with $r' := 2 \log|\frac{1+\mu}{1-\mu}|$ yields:

$$\mathbb{E}\|\phi - \hat{\phi}_{m_1, m_2, m_3}\|_{L^2}^2 = \mathcal{O}\left(\frac{1}{cT}\right).$$

5 Numerical comparison

In this section, we compare the performance on simulated data of the Laguerre–Fourier estimator (see Section 2) and the Laguerre deconvolution estimators (see Section 3). We consider the three Gerber–Shiu functions of Example 1.1:

1. $\phi^{(1)}(u) = \mathbb{P}[\tau(u) < \infty]$ the ruin probability;
2. $\phi^{(2)}(u) = \mathbb{E}[(U_{\tau(u)-} + |U_{\tau(u)}|)\mathbf{1}_{\tau(u) < \infty}]$ the expected claim size causing the ruin;
3. $\phi^{(3)}(u) = \mathbb{E}[e^{-\delta\tau(u)}\mathbf{1}_{\tau(u) < \infty}]$ the Laplace transform of the ruin time, for $\delta = 0.1$.

We also consider three sets of parameters:

1. X follows an exponential distribution, $\lambda = 1$, $\mu = 1$, $c = 1.5$. In this setting, $\theta \approx 0.67$.
2. X follows an exponential distribution, $\lambda = 1.25$, $\mu = 2$, $c = 3$. In this setting, $\theta \approx 0.83$.
3. X follows a $\Gamma(2, \frac{\mu}{2})$ distribution, $\lambda = 1.25$, $\mu = 2$, $c = 3$. In this setting, $\theta \approx 0.83$.

Using Laplace inversion techniques [Asmussen and Albrecher, 2010, Chapter XII], we have access to explicit formulas for these Gerber–Shiu functions. In all cases, they are given by a sum of products of polynomials and exponentials, hence they belong to $W^s(\mathbb{R}_+)$ for all $s > 0$.

Computation of the estimators Let us start on how we compute the Laguerre functions. The Laguerre polynomials, defined by (6), satisfy the relations:

$$\begin{aligned}(k+1)L_{k+1}(x) &= (2k+1-x)L_k(x) - kL_{k-1}(x), \\ xL'_k(x) &= k(L_k(x) - L_{k-1}(x)),\end{aligned}\tag{20}$$

see formulas 22.7.12 and 22.8.6 in Abramowitz and Stegun [1972]. From this formulas, one can prove:

$$(x\psi_k)' = \frac{k+1}{2}\psi_{k+1}(x) + \frac{1}{2}\psi_k(x) - \frac{k}{2}\psi_{k-1}(x).\tag{21}$$

Let $\Psi_k(x) := \int_0^x \psi_k(t) dt$ be the primitive of the Laguerre function ψ_k ; these functions are used to compute the coefficients \hat{b}_k and \hat{c}_k below. From (20), and by integrating (21), we see that the Laguerre functions and their primitives can be computed recursively:

$$\begin{aligned}(k+1)\psi_{k+1}(x) &= (2k+1-x)\psi_k(x) - k\psi_{k-1}(x), \\ (k+1)\Psi_k(x) &= 2x\psi_k(x) - \Psi_k(x) + k\Psi_{k-1}(x).\end{aligned}$$

The expression of \hat{b}_k and \hat{c}_k depends on the value of δ and the form of w :

1. *Ruin probability.* The estimators of the coefficients b_k and c_k are in this case:

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

We compute the integrals in \hat{c}_k using Romberg's method with $2^{10} + 1$ points.

2. *Expected claim size causing the ruin.* The estimators of the coefficients b_k and c_k are in this case:

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

We compute the integrals in \hat{c}_k using Romberg's method with $2^{10} + 1$ points.

3. *Laplace transform.* The estimators of the coefficients b_k and c_k are in this case:

$$\hat{b}_k = \frac{1}{cT} \sum_{i=1}^{N_T} \int_0^{X_i} e^{-\hat{\rho}_\delta(X_i-x)} \psi_k(x) dx, \quad \hat{c}_k = \frac{1}{\hat{\rho}_\delta} \left(\frac{1}{cT} \sum_{i=1}^{N_T} \Psi_k(X_i) - \hat{b}_k \right),$$

where we used integration by parts to obtain this expression of \hat{c}_k . We compute the integrals in \hat{b}_k using Romberg's method with $2^{10} + 1$ points. We compute $\hat{\rho}_\delta$, the solution of Equation (11), with Newton's method using the initial condition $\frac{\delta + \lambda/2}{c}$.

For the Laguerre–Fourier estimator, once we have computed $(\hat{b}_k)_{0 \leq k < m_2}$ and $(\hat{c}_k)_{0 \leq k < m_3}$, we can compute the coefficients \hat{a}_k defined by (8):

$$\begin{aligned}\hat{a}_k &= \frac{1}{2\pi} \int_{\mathbb{R}} \frac{\mathcal{F}\hat{h}(\omega)}{1 - \widehat{\mathcal{F}g}(\omega)} \overline{\mathcal{F}\psi_k(\omega)} d\omega, \\ \widehat{\mathcal{F}g}(\omega) &= \begin{cases} \mathcal{F}\hat{g}(\omega) & \text{if } |\mathcal{F}\hat{g}(\omega)| \leq \theta_0, \\ 0 & \text{if } |\mathcal{F}\hat{g}(\omega)| > \theta_0, \end{cases} \quad \mathcal{F}\hat{g} = \sum_{k=0}^{m_2-1} \hat{b}_k \mathcal{F}\psi_k, \quad \mathcal{F}\hat{h} = \sum_{k=0}^{m_3-1} \hat{c}_k \mathcal{F}\psi_k,\end{aligned}$$

where $\mathcal{F}\psi_k$ is given by (7), and where the integral in \hat{a}_k is computed with Romberg's method on a discretization of $[-10^3, 10^3]$ with $2^{15} + 1$ points.

For the Laguerre deconvolution estimators, once we have computed $(\hat{b}_k)_{0 \leq k < m}$ and $(\hat{c}_k)_{0 \leq k < m}$, we can compute the matrix $\hat{\mathbf{A}}_m$ defined by (17) and then compute the coefficients $\hat{a}_m^{\text{Lag}_i}$ as described in Section 3.

Remark 5.1. While the Gerber–Shiu function is always positive, this is not necessarily the case of the estimators. However, we can always take their positive part, since it does not increase their risk:

$$\mathbb{E}\|\phi - (\hat{\phi})_+\|^2 \leq \mathbb{E}\|\phi - \hat{\phi}\|^2.$$

In Figures 1 and 2, we observe that the estimators stay positive where ϕ is positive, and that they can take small negative values when ϕ becomes small (as u tends to $+\infty$). Hence, it is reasonable to use the estimators without taking their positive part. We choose to do so, in the simulations.

Model selection Each estimator we consider depends on one or several parameters that need to be chosen. The Laguerre–Fourier estimator and the Laguerre deconvolution estimator $\hat{\phi}_m^{\text{Lag}_1}$ depend on a truncation parameter θ_0 , which needs to be chosen such that $\theta < \theta_0$. We choose $\theta_0 = 0.95$ in our simulations.

- The Laguerre–Fourier estimator depends on four parameters: m_1, m_2 and m_3 , the dimensions of the projection spaces for the functions ϕ, g and h , and θ_0 the truncation parameter in the estimation of $\widehat{\mathcal{F}}g$. As said in Remark 4.3, we can choose $m_1 = m_2 = m_3 = \lceil cT \rceil$, no selection procedure is required. Still, we propose a *model reduction procedure* for the choice of m_2 and m_3 , that we describe in Appendix A.
- The naive Laguerre deconvolution estimator $\hat{\phi}_m^{\text{Lag}_0}$, defined by (18), depends on one parameter: m , the dimension of the projection space for ϕ . However, there is no model selection procedure for m . In their numerical section, Zhang and Su [2018] only consider (as we do) Gerber–Shiu functions with exponential decay; hence the bias term also decays with exponential rate. Using this fact, they chose $m = \lfloor 5T^{1/10} \rfloor$. We make the same choice in our simulations and we write $\hat{\phi}^{\text{ZS}}$ this estimator.
- The Laguerre deconvolution estimators $\hat{\phi}_m^{\text{Lag}_1}$ and $\hat{\phi}_m^{\text{Lag}_2}$ also depend on m . For $i \in \{1, 2\}$, we choose \hat{m}^{Lag_i} as the minimizer of a penalized criterion:

$$\hat{m}^{\text{Lag}_i} \in \arg \min_{m \in \mathcal{M}_i} \left\{ -\|\hat{\phi}_m^{\text{Lag}_i}\|_{L^2}^2 + \kappa_i \text{pen}_i(m) \right\} \quad (22)$$

where the model collections are:

$$\begin{aligned} \mathcal{M}_1 &:= \left\{ 1 \leq m \leq M \mid \|\hat{\mathbf{A}}_m^{-1}\|_{\text{op}} \leq \frac{1}{1 - \theta_0} \right\} \\ \mathcal{M}_2 &:= \left\{ 1 \leq m \leq M \mid \|\hat{\mathbf{A}}_m^{-1}\|_{\text{op}}^2 \leq \frac{cT}{m \log(m)} \right\} \end{aligned}$$

with $M = \lceil cT \rceil \wedge 500$ (we do not compute more than 500 coefficients, because of computation time).

In the following, if $F(X)$ is a function of X , we write $\overline{F(X)} := \frac{1}{N_T} \sum_{i=1}^{N_T} F(X_i)$ its empirical mean from the sample $\{X_1, \dots, X_{N_T}\}$. For the penalty terms, we choose empirical versions of the variance terms in Theorems 3.3 and 3.6:

$$\begin{aligned} \text{pen}_1(m) &:= \frac{1}{(1 - \theta_0)^2} \left(\|\hat{\phi}_m^{\text{Lag}_1}\|_{L^2}^2 m \hat{V}_g + \hat{V}_h \right) \\ \text{pen}_2(m) &:= (\hat{\lambda} \vee \hat{\lambda}^2) \left(\frac{\overline{W(X)}}{c} + \|\hat{\phi}_m^{\text{Lag}_1}\|_{L^2}^2 (\overline{X} \vee \overline{X}^2) m \log(m) \right), \end{aligned}$$

with:

$$\hat{V}_g := \begin{cases} \frac{\hat{\lambda}}{c^2 T} \overline{X} & \text{if } \delta = 0, \\ \frac{\hat{\lambda}^2}{c^2 T} \left(\overline{X} + \frac{(\overline{X^2})^{1/2}}{\delta(1 - \theta_0)^2} \right) & \text{if } \delta > 0, \end{cases} \quad \hat{V}_h := \begin{cases} \frac{\hat{\lambda}}{c^2 T} \overline{W(X)} & \text{if } \delta = 0, \\ \frac{\hat{\lambda}^2}{c^2 T} \left(\overline{W(X)} + \frac{(\overline{W(X)^2})^{1/2}}{\delta(1 - \theta_0)^2} \right) & \text{if } \delta > 0. \end{cases}$$

The constants κ_1 and κ_2 are calibrated following the “minimum penalty heuristic” [Arlot and Massart, 2009]. On several preliminary simulations, we compute the selected dimension \hat{m} as a function of κ , and we find κ_{\min} such that for $\kappa < \kappa_{\min}$ the dimension is too high and for $\kappa > \kappa_{\min}$ it is acceptable. Then, the selected constant is $2\kappa_{\min}$. In our cases, we choose:

- $\kappa_1 = 0.01$, $\kappa_2 = 0.01$ for the ruin probability;
- $\kappa_1 = 0.1$, $\kappa_2 = 1$ for the expected claim size causing the ruin;
- $\kappa_1 = 10^{-8}$ for the Laplace transform of the ruin time, $\delta = 0.1$.

There is no constant κ_2 in the last case because the Laguerre deconvolution estimator $\hat{\phi}_m^{\text{Lag}_2}$ is defined only if $\delta = 0$.

We write $\hat{\phi}^{\text{Lag}_1} := \hat{\phi}_{\hat{m}^{\text{Lag}_1}}^{\text{Lag}_1}$ and $\hat{\phi}^{\text{Lag}_2} := \hat{\phi}_{\hat{m}^{\text{Lag}_2}}^{\text{Lag}_2}$ in the following.

MISE calculation We compare the estimators by looking at their MISE: $\mathbb{E}\|\phi - \hat{\phi}\|_{L^2}^2$. We compute the norm $\|\cdot\|_{L^2}$ with Romberg’s method using a discretization of $[0, u_{\max}]$ with $2^{11} + 1$ points. The value of u_{\max} varies from 12 to 50, depending on the parameters set. We compute the expectation by an empirical mean over $n = 200$ paths of the process $(U_t)_{t \in [0, T]}$. We also compute a 95% confidence interval for the MISE, using the asymptotic confidence interval for a mean (CLT approximation):

$$\text{CI} = \left[\overline{\text{ISE}}_n \pm q_{1-\frac{\alpha}{2}} \frac{S_n}{\sqrt{n}} \right], \quad \alpha = 5\%,$$

where $\overline{\text{ISE}}_n$ is the empirical mean of the ISEs, $q_{1-\frac{\alpha}{2}}$ is the $(1 - \frac{\alpha}{2})$ -quantile of the normal distribution, and S_n^2 is the empirical variance of the ISEs. We have two goals in this section:

1. To compare the performance of our Laguerre–Fourier estimator with the Laguerre deconvolution estimators.
2. To see if the model selection procedures (22) for the Laguerre deconvolution estimators lead to the same performance than the naive choice $m = \lfloor 5T^{1/10} \rfloor$.

The code that performed the simulations can be obtained on request.

Results We display our results in Tables 1, 2 and 3. Concerning the estimation of the ruin probability (Table 1), we see that all the estimators perform well with the first set of parameter (exponential distribution, $\theta = 0.67$). However, with the two other sets of parameters (exponential distribution and Gamma(2) distribution, $\theta = 0.83$), the difference is clear: the Laguerre–Fourier estimator has the smallest risk, followed by the estimator of Zhang and Su [2018], and the Laguerre deconvolution estimators come last. We notice that $\hat{\phi}^{\text{Lag}_2}$ seems to be better than $\hat{\phi}^{\text{Lag}_1}$ in this case.

Concerning the estimation of the expected jump size causing the ruin (Table 2), the difference is even clearer. With the first set of parameters, we see that the Laguerre–Fourier is better for small sample size ($\mathbb{E}[N_T] = 100$), but equivalent to the other estimators for larger sample sizes. We also notice that the estimator $\hat{\phi}^{\text{ZS}}$ and $\hat{\phi}^{\text{Lag}_2}$ have the same risk. With the two other sets of parameters, we find again that the Laguerre–Fourier estimator is better than the estimator $\hat{\phi}^{\text{ZS}}$, which is better than the Laguerre deconvolution estimators. This time, we see that $\hat{\phi}^{\text{Lag}_1}$ has better performances than $\hat{\phi}^{\text{Lag}_2}$.

Concerning the estimation of Laplace transform of the ruin time (Table 3), we see no difference between the MISE of the Laguerre–Fourier estimator and the Laguerre deconvolution estimators.

For illustration purposes, on Figures 1 and 2, we show the estimations of the ruin probability and the expected claim size causing the ruin, on 50 independent samples, with the second set of parameters (exponential distribution, $\theta = 0.83$). Qualitatively, we see that the Laguerre–Fourier estimator is better than the others. In contrast, the non data-driven choice of m for estimator of Zhang and Su [2018] seems not appropriate in this setting.

To conclude, we can say that our Laguerre–Fourier estimator has better performances than the Laguerre deconvolution estimators on simulated data, even in the exponential case where they

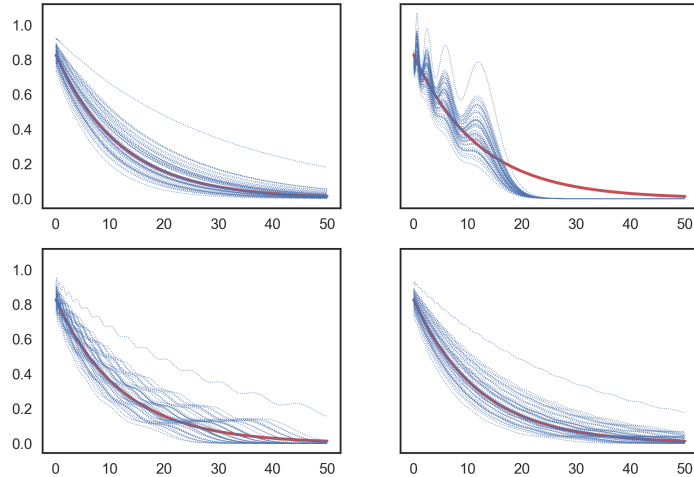


Figure 1: Estimation of the ruin probability when the parameters of the model are $\lambda = 1.25$, $\mu = 2$, $c = 3$, $X \sim \text{Exp}(1/\mu)$ and $T = 800$ (so that $\mathbb{E}[N_T] = 1000$). For each estimation procedure, we plot the estimation of ϕ from 50 independent samples. The true function is in bold red and the estimated functions are in dotted blue. **Top left:** Laguerre–Fourier. **Top right:** estimator of Zhang and Su [2018]. **Bottom left:** Laguerre deconvolution 1. **Bottom right:** Laguerre deconvolution 2.

have theoretically the same MISE (up to a log factor). Furthermore, the Laguerre deconvolution estimators with the model selection procedure (22) fail to match the performance of the estimator of Zhang and Su [2018], for which we choose the parameter m knowing the bias decay rate of ϕ , in most cases.

Remark 5.2. In Tables 1, 2 and 3, the MISEs of the estimators are not normalized by $\|\phi\|_{L^2}^2$, the size of the estimated function. Hence, it is normal that the order of magnitude of the results varies from one function to another. For example, in Table 2, $\|\phi\|_{L^2}^2$ equals respectively 5, 100 and 50, for each set of parameters.

6 Conclusion

Using a projection estimator on the Laguerre basis, and computing the coefficients with Fourier transforms, we constructed an estimator of the Gerber–Shiu function that achieves parametric rates of convergence, without needing a model selection procedure. It is worth noticing that our results are non-asymptotic and concern the MISE of the estimator. In comparison, the Laguerre deconvolution estimators have slower rates of convergence and necessitate a model selection procedure in practice. The better performances of our procedure are confirmed by a numerical study, on simulated data.

Knowing that the Laguerre deconvolution method does not achieve the best rate of convergence in the compound Poisson model is important. Indeed, this method is used to estimate the Gerber–Shiu function in more general models, see Zhang and Su [2019], Su et al. [2019a] and Su et al. [2019b]. These papers have one thing in common: they all want to estimate a function ϕ that satisfies an equation of the form $\phi = \phi * g + h$, with g and h functions that depend on the specificity of each problem. If we applied the procedure described in the beginning of Section 2, we could obtain an estimator that would achieve the same rate of convergence as the estimators of g and h (see Remark 2.2). Hence the Laguerre deconvolution method used in these papers is not optimal since a factor m appears in the variance term in the construction step of $\hat{\phi}_m$ from \hat{g}_m and \hat{h}_m .

Parameters	Estimator	$\mathbb{E}[N_T] = 100$	$\mathbb{E}[N_T] = 200$	$\mathbb{E}[N_T] = 400$
$X \sim \text{Exp}(1)$ $\lambda = 1$ $c = 1.5$ $\theta = 0.67$	LagFou	0.14 [0.07, 0.21] $m_1 = 150$	0.053 [0.039, 0.067] $m_1 = 300$	0.022 [0.017, 0.027] $m_1 = 500$
	ZS	0.23 [0.02, 0.44] $m = 8$	0.053 [0.039, 0.068] $m = 9$	0.022 [0.017, 0.028] $m = 10$
	LagDec1	0.25 [0.01, 0.48] $\hat{m} = 3.3$	0.055 [0.042, 0.069] $\hat{m} = 3.8$	0.024 [0.019, 0.029] $\hat{m} = 4.2$
	LagDec2	0.23 [0.02, 0.45] $\hat{m} = 6.0$	0.053 [0.039, 0.068] $\hat{m} = 6.5$	0.023 [0.017, 0.028] $\hat{m} = 7.0$
$X \sim \text{Exp}(1/2)$ $\lambda = 1.25$ $c = 3$ $\theta = 0.83$	LagFou	0.95 [0.80, 1.09] $m_1 = 240$	0.67 [0.53, 0.80] $m_1 = 480$	0.43 [0.31, 0.55] $m_1 = 500$
	ZS	1.57 [1.26, 1.89] $m = 8$	1.02 [0.74, 1.30] $m = 9$	0.54 [0.46, 0.61] $m = 9$
	LagDec1	8.51 [5.07, 11.95] $\hat{m} = 11.0$	3.82 [1.57, 6.07] $\hat{m} = 12.8$	0.72 [0.36, 1.07] $\hat{m} = 14.4$
	LagDec2	2.96 [2.11, 3.82] $\hat{m} = 14.2$	1.94 [1.08, 2.80] $\hat{m} = 18.1$	0.64 [0.36, 0.92] $\hat{m} = 21.8$
$X \sim \Gamma(2, 1/4)$ $\lambda = 1.25$ $c = 3$ $\theta = 0.83$	LagFou	0.64 [0.52, 0.77] $m_1 = 240$	0.46 [0.37, 0.56] $m_1 = 480$	0.30 [0.22, 0.38] $m_1 = 500$
	ZS	1.77 [1.07, 2.47] $m = 8$	0.62 [0.45, 0.78] $m = 9$	0.30 [0.23, 0.36] $m = 9$
	LagDec1	7.22 [4.25, 10.19] $\hat{m} = 9.3$	1.71 [0.87, 2.56] $\hat{m} = 10.6$	0.45 [0.21, 0.70] $\hat{m} = 11.6$
	LagDec2	2.71 [1.80, 3.62] $\hat{m} = 12.1$	1.07 [0.66, 1.47] $\hat{m} = 15.7$	0.41 [0.22, 0.60] $\hat{m} = 18.0$

Table 1: **Ruin Probability.** For two sets of parameters, we compare the three estimators of the ruin probability: the Laguerre–Fourier estimator (LagFou), the estimator of [Zhang and Su \[2018\]](#) (ZS), and the Laguerre deconvolution estimators (LagDec1 and LagDec2). In each case, we display the estimation of the MISE over 200 samples with a 95% confidence interval and the model used (\hat{m} is the mean selected model in the case of the Laguerre deconvolution estimators).

Parameters	Estimator	$\mathbb{E}[N_T] = 100$	$\mathbb{E}[N_T] = 200$	$\mathbb{E}[N_T] = 400$
$X \sim \text{Exp}(1)$ $\lambda = 1$ $c = 1.5$ $\theta = 0.67$	LagFou	1.71 [1.09, 2.32] $m_1 = 150$	0.60 [0.46, 0.73] $m_1 = 300$	0.34 [0.27, 0.40] $m_1 = 500$
	ZS	1.80 [1.07, 2.53] $m = 8$	0.62 [0.47, 0.77] $m = 9$	0.34 [0.27, 0.41] $m = 10$
	LagDec1	1.41 [1.19, 1.64] $\hat{m} = 1.9$	0.84 [0.74, 0.93] $\hat{m} = 2.2$	0.44 [0.39, 0.50] $\hat{m} = 2.8$
	LagDec2	1.86 [1.11, 2.61] $\hat{m} = 3.7$	0.64 [0.49, 0.78] $\hat{m} = 4.1$	0.35 [0.28, 0.42] $\hat{m} = 4.6$
$X \sim \text{Exp}(1/2)$ $\lambda = 1.25$ $c = 3$ $\theta = 0.83$	LagFou	46.2 [30.0, 62.3] $m_1 = 240$	28.1 [21.7, 34.5] $m_1 = 480$	20.5 [15.1, 25.8] $m_1 = 500$
	ZS	96.3 [62.4, 130.2] $m = 8$	48.0 [31.3, 64.6] $m = 9$	27.9 [23.0, 32.7] $m = 9$
	LagDec1	77.5 [71.6, 83.5] $\hat{m} = 3.0$	56.3 [45.1, 67.4] $\hat{m} = 4.9$	38.9 [29.7, 48.1] $\hat{m} = 7.1$
	LagDec2	197.7 [115.1, 280.3] $\hat{m} = 9.7$	96.7 [47.4, 146.0] $\hat{m} = 12.5$	48.5 [27.5, 69.6] $\hat{m} = 14.6$
$X \sim \Gamma(2, 1/4)$ $\lambda = 1.25$ $c = 3$ $\theta = 0.83$	LagFou	11.7 [9.4, 14.0] $m_1 = 240$	9.2 [7.5, 10.9] $m_1 = 480$	6.2 [4.1, 8.3] $m_1 = 500$
	ZS	18.5 [12.0, 25.0] $m = 8$	13.6 [10.1, 17.1] $m = 9$	5.9 [4.6, 7.2] $m = 9$
	LagDec1	19.2 [18.2, 20.2] $\hat{m} = 2.8$	15.0 [13.1, 17.0] $\hat{m} = 4.3$	8.4 [7.1, 9.7] $\hat{m} = 5.9$
	LagDec2	28.2 [19.4, 37.1] $\hat{m} = 7.8$	24.6 [16.5, 32.7] $\hat{m} = 10.0$	8.3 [5.6, 11.1] $\hat{m} = 11.3$

Table 2: **Expected claim size causing the ruin.** For two sets of parameters, we compare the three estimators of the expected jump size causing the ruin: the Laguerre–Fourier estimator (LagFou), the estimator of [Zhang and Su \[2018\]](#) (ZS), and the Laguerre deconvolution estimators (LagDec1 and LagDec2). In each case, we display the estimation of the MISE over 200 samples with a 95% confidence interval and the model used (\hat{m} is the mean selected model in the case of the Laguerre deconvolution estimators).

Parameters	Estimator	$\mathbb{E}[N_T] = 100$	$\mathbb{E}[N_T] = 200$	$\mathbb{E}[N_T] = 400$
$X \sim \text{Exp}(1)$ $\lambda = 1$ $c = 1.5$ $\theta = 0.67$	LagFou	2.50 [1.91, 3.09] $m_1 = 150$	1.09 [0.87, 1.31] $m_1 = 300$	0.64 [0.52, 0.77] $m_1 = 500$
	ZS	2.50 [1.93, 3.07] $m = 8$	1.10 [0.88, 1.33] $m = 9$	0.66 [0.53, 0.79] $m = 10$
	LagDec1	2.52 [1.95, 3.08] $\hat{m} = 4.2$	1.11 [0.89, 1.34] $\hat{m} = 4.6$	0.67 [0.54, 0.80] $\hat{m} = 4.9$
$X \sim \text{Exp}(1/2)$ $\lambda = 1.25$ $c = 3$ $\theta = 0.83$	LagFou	11.81 [9.26, 14.36] $m_1 = 240$	5.60 [4.49, 6.72] $m_1 = 480$	2.51 [2.04, 2.98] $m_1 = 500$
	ZS	12.47 [10.53, 14.41] $m = 8$	6.13 [5.07, 7.19] $m = 9$	3.30 [2.86, 3.75] $m = 9$
	LagDec1	13.22 [10.61, 15.84] $\hat{m} = 10.2$	5.82 [4.57, 7.06] $\hat{m} = 11.2$	2.65 [2.15, 3.14] $\hat{m} = 12.4$
$X \sim \Gamma(2, 1/4)$ $\lambda = 1.25$ $c = 3$ $\theta = 0.83$	LagFou	10.26 [8.08, 12.45] $m_1 = 240$	4.09 [3.28, 4.91] $m_1 = 480$	2.01 [1.57, 2.46] $m_1 = 500$
	ZS	9.76 [8.07, 11.45] $m = 8$	4.16 [3.36, 4.96] $m = 9$	2.20 [1.77, 2.63] $m = 9$
	LagDec1	10.39 [8.39, 12.38] $\hat{m} = 8.8$	4.15 [3.29, 5.00] $\hat{m} = 9.5$	2.05 [1.59, 2.52] $\hat{m} = 10.3$

Table 3: **Laplace transform**, $\delta = 0.1$. For two sets of parameters, we compare three estimators of the Laplace transform of the ruin time: the Laguerre–Fourier estimator (LagFou), the estimator of [Zhang and Su \[2018\]](#) (ZS) and the Laguerre deconvolution estimator (LagDec1). In each case, we display the estimation of the MISE (multiplied by 10^2 in this table) over 200 samples with a 95% confidence interval and the model used (\hat{m} is the mean selected model in the case of the Laguerre deconvolution estimator).

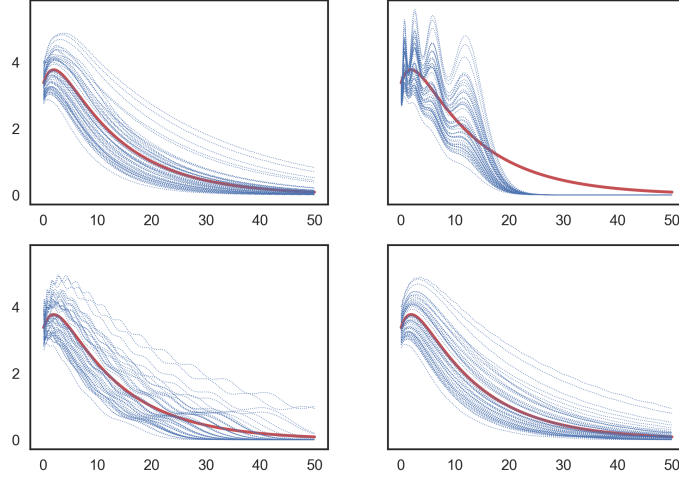


Figure 2: Estimation of the expected claim size causing the ruin when the parameters of the model are $\lambda = 1.25$, $\mu = 2$, $c = 3$, $X \sim \text{Exp}(1/\mu)$ and $T = 800$ (so that $\mathbb{E}[N_T] = 1000$). For each estimation procedure, we plot the estimation of ϕ from 50 independent samples. The true function is in bold red and the estimated functions are in dotted blue. **Top left:** Laguerre–Fourier. **Top right:** estimator of Zhang and Su [2018]. **Bottom left:** Laguerre deconvolution 1. **Bottom right:** Laguerre deconvolution 2.

7 Proofs

Proof of Proposition 2.1. By Pythagoras theorem, $\|\phi - \hat{\phi}_m\|_{L^2}^2 = \|\phi - \phi_m\|_{L^2}^2 + \|\phi_m - \hat{\phi}_m\|_{L^2}^2$. Let Π_m be the projector on $\text{Span}(\mathcal{F}\psi_0, \dots, \mathcal{F}\psi_{m-1})$. Since $\|\mathcal{F}\psi_k\|^2 = 2\pi$, we get:

$$\begin{aligned}
\|\phi_m - \hat{\phi}_m\|_{L^2}^2 &= \sum_{k=0}^{m-1} (\hat{a}_k - a_k)^2 \\
&= \frac{1}{(2\pi)^2} \sum_{k=0}^{m-1} \left\langle \frac{\mathcal{F}\hat{h}}{1 - \widehat{\mathcal{F}g}} - \frac{\mathcal{F}h}{1 - \mathcal{F}g}, \mathcal{F}\psi_k \right\rangle^2 \\
&= \frac{1}{2\pi} \left\| \Pi_m \left(\frac{\mathcal{F}\hat{h}}{1 - \widehat{\mathcal{F}g}} - \frac{\mathcal{F}h}{1 - \mathcal{F}g} \right) \right\|_{L^2}^2 \\
&\leq \frac{1}{2\pi} \left\| \frac{\mathcal{F}\hat{h}}{1 - \widehat{\mathcal{F}g}} - \frac{\mathcal{F}h}{1 - \mathcal{F}g} \right\|_{L^2}^2.
\end{aligned} \tag{23}$$

Then since $|\widehat{\mathcal{F}g}| \leq \theta_0$ by definition, and $|\mathcal{F}g| \leq \|g\|_{L^1} \leq \theta$, we obtain:

$$\begin{aligned}
\left\| \frac{\mathcal{F}\hat{h}}{1 - \widehat{\mathcal{F}g}} - \frac{\mathcal{F}h}{1 - \mathcal{F}g} \right\|_{L^2}^2 &\leq 2 \left\| \frac{\mathcal{F}\hat{h}}{1 - \widehat{\mathcal{F}g}} - \frac{\mathcal{F}h}{1 - \widehat{\mathcal{F}g}} \right\|_{L^2}^2 + 2 \left\| \frac{\mathcal{F}h}{1 - \widehat{\mathcal{F}g}} - \frac{\mathcal{F}h}{1 - \mathcal{F}g} \right\|_{L^2}^2 \\
&\leq \frac{2}{(1 - \theta_0)^2} \|\mathcal{F}\hat{h} - \mathcal{F}h\|_{L^2}^2 + \frac{2\|h\|_{L^1}^2}{(1 - \theta_0)^2(1 - \theta)^2} \|\widehat{\mathcal{F}g} - \mathcal{F}g\|_{L^2}^2.
\end{aligned} \tag{24}$$

To control the last term, we decompose according to the set $\{|\mathcal{F}\hat{g}| \leq \theta_0\}$ and its complement:

$$\|\widehat{\mathcal{F}g} - \mathcal{F}g\|_{L^2}^2 \leq \|\mathcal{F}\hat{g} - \mathcal{F}g\|_{L^2}^2 + \|g\|_{L^1}^2 \text{Leb}(\{|\mathcal{F}\hat{g}| > \theta_0\}).$$

Thus if $\theta < \theta_0$, then $\{|\mathcal{F}\hat{g}| > \theta_0\} \subseteq \{|\mathcal{F}\hat{g} - \mathcal{F}g| \geq \theta_0 - \theta\}$, therefore Markov inequality yields:

$$\|\widehat{\mathcal{F}g} - \mathcal{F}g\|_{L^2}^2 \leq \|\mathcal{F}\hat{g} - \mathcal{F}g\|_{L^2}^2 + \frac{\|g\|_{L^1}^2}{(\theta_0 - \theta)^2} \|\mathcal{F}\hat{g} - \mathcal{F}g\|_{L^2}^2. \tag{25}$$

Finally, gathering (23), (24) and (25), and using Plancherel theorem yield the desired result. \square

7.1 Proof of Theorem 2.5

We start with some preliminary lemmas.

Lemma 7.1. *Let Y_1, \dots, Y_n be i.i.d non-negative random variables. We denote by $\mathcal{L}(s) := \mathbb{E}[e^{-sY_1}]$ their Laplace transform and we denote by $\widehat{\mathcal{L}}(s) := \frac{1}{n} \sum_{i=1}^n e^{-sY_i}$ the empirical Laplace transform. Then for $p \geq 1$, we have:*

$$\mathbb{E} \left[\sup_{s>0} \left| \widehat{\mathcal{L}}(s) - \mathcal{L}(s) \right|^{2p} \right] \leq \frac{p!}{2^{p-1} n^p}.$$

Proof. Let $\widehat{F}(x) := \frac{1}{n} \sum_{i=1}^n \mathbf{1}_{\{Y_i \leq x\}}$ be the empirical c.d.f. of the Y_i 's, and let $F(x) := \mathbb{P}[Y \leq x]$ be their c.d.f. We notice that for $s > 0$:

$$\int_0^{+\infty} s e^{-sx} \widehat{F}(x) dx = \frac{1}{n} \sum_{i=1}^n \int_0^{+\infty} s e^{-sx} \mathbf{1}_{\{Y_i \leq x\}} dx = \frac{1}{n} \sum_{i=1}^n e^{-sY_i} =: \widehat{\mathcal{L}}(s),$$

and by the same argument, $\mathcal{L}(s) = \int_0^{+\infty} s e^{-sx} F(x) dx$. Thus:

$$\sup_{s>0} \left| \widehat{\mathcal{L}}(s) - \mathcal{L}(s) \right| \leq \sup_{s>0} \int_0^{+\infty} s e^{-sx} \left| \widehat{F}(x) - F(x) \right| dx \leq \|\widehat{F} - F\|_\infty.$$

We take the expectation and we get:

$$\mathbb{E} \left[\sup_{s>0} \left| \widehat{\mathcal{L}}(s) - \mathcal{L}(s) \right|^{2p} \right] \leq \mathbb{E} \left[\|\widehat{F} - F\|_\infty^{2p} \right] = 2p \int_0^{+\infty} t^{2p-1} \mathbb{P} \left[\|\widehat{F} - F\|_\infty \geq t \right] dt. \quad (26)$$

By Massart [1990], $\mathbb{P} \left[\sqrt{n} \|\widehat{F} - F\|_\infty \geq x \right] \leq 2e^{-2x^2}$, so by setting $t = x/\sqrt{n}$ in (26), we obtain:

$$\mathbb{E} \left[\sup_{s>0} \left| \widehat{\mathcal{L}}(s) - \mathcal{L}(s) \right|^{2p} \right] \leq \frac{2p}{n^p} \int_0^{+\infty} x^{2p-1} 2e^{-2x^2} dx = \frac{2p}{n^p} \int_0^{+\infty} u^{p-1} e^{-2u} du = \frac{p!}{n^p 2^{p-1}}. \quad \square$$

Lemma 7.2. *Let $Z \sim \mathcal{P}(\lambda)$ and $m_j(\lambda) := \mathbb{E}[(Z - \lambda)^j]$ be the j -th central moment of Z . Then, for all $r \geq 2$ we have:*

$$m_r(\lambda) = \lambda \sum_{j=0}^{r-2} \binom{r-1}{j} m_j(\lambda).$$

Proof. Let $\mathcal{L}(\lambda, t) := e^{\lambda(e^t - t - 1)} = \mathbb{E}[e^{t(Z - \lambda)}]$ and $\varphi(t) := e^t - t - 1$. Then $m_r(\lambda) = \frac{\partial^r \mathcal{L}}{\partial t^r}(\lambda, 0)$. By Leibniz's rule:

$$\frac{\partial^r \mathcal{L}}{\partial t^r}(\lambda, t) = \frac{\partial^{r-1}}{\partial t^{r-1}} \left(\lambda \varphi'(t) \mathcal{L}(\lambda, t) \right) = \lambda \sum_{j=0}^{r-1} \binom{r-1}{j} \frac{\partial^j \mathcal{L}}{\partial t^j}(\lambda, t) \varphi^{(r-j)}(t).$$

Taking $t = 0$ gives the result since $\varphi'(0) = 0$ and $\varphi^{(k)}(0) = 1$ if $k \geq 2$. \square

Corollary 7.3. *The central moments $m_{2r}(\lambda)$ and $m_{2r+1}(\lambda)$ are polynomials of degree r in λ .*

The next proposition provides an upper bound on the L^p -risk of $\hat{\rho}_\delta$.

Proposition 7.4. *Under Assumption 1, for $p \geq 1$, we have:*

$$\mathbb{E}[(\hat{\rho}_\delta - \rho_\delta)^{2p}] \leq \frac{C(p, \lambda)}{c^{2p} (1 - \theta)^{2p} T^p},$$

where $C(p, \lambda)$ is a $\mathcal{O}(\lambda^p)$.

Proof. By definition, ρ_δ is a solution of the Lundberg equation, so it is a zero of the function:

$$\ell_\delta(s) := cs - (\lambda + \delta) + \lambda \mathcal{L}f(s).$$

The estimator $\hat{\rho}_\delta$ is then a zero of the function:

$$\hat{\ell}_\delta(s) := cs - (\hat{\lambda} + \delta) + \hat{\lambda} \widehat{\mathcal{L}f}(s).$$

We use a Taylor–Lagrange expansion:

$$\hat{\ell}_\delta(\hat{\rho}_\delta) = 0 = \ell_\delta(\rho_\delta) = \ell_\delta(\hat{\rho}_\delta) + \ell'_\delta(z)(\rho_\delta - \hat{\rho}_\delta),$$

where z is between ρ_δ and $\hat{\rho}_\delta$.

$$|\ell'_\delta(z)| = \left| c - \lambda \int_0^{+\infty} x e^{-zx} f(x) dx \right| \geq c - \lambda \int_0^{+\infty} x f(x) dx = c - \lambda \mu > 0,$$

under the safety loading condition. Thus:

$$\begin{aligned} |\rho_\delta - \hat{\rho}_\delta| &\leq \frac{1}{c - \lambda \mu} |\hat{\ell}_\delta(\hat{\rho}_\delta) - \ell_\delta(\hat{\rho}_\delta)| \\ &= \frac{1}{c(1 - \theta)} \left| \hat{\lambda} (\widehat{\mathcal{L}f}(\hat{\rho}_\delta) - \mathcal{L}f(\hat{\rho}_\delta)) + (\hat{\lambda} - \lambda) (1 - \mathcal{L}f(\hat{\rho}_\delta)) \right| \\ &\leq \frac{1}{c(1 - \theta)} \left(|\hat{\lambda}| \|\widehat{\mathcal{L}f} - \mathcal{L}f\|_\infty + 2|\hat{\lambda} - \lambda| \right) \\ \mathbb{E}[(\hat{\rho}_\delta - \rho_\delta)^{2p}] &\leq \frac{1}{c^{2p}(1 - \theta)^{2p}} \left(2^{2p-1} \mathbb{E}[\hat{\lambda}^{2p} \|\widehat{\mathcal{L}f} - \mathcal{L}f\|_\infty^{2p}] + 2^{4p-1} \mathbb{E}[|\hat{\lambda} - \lambda|^{2p}] \right). \end{aligned}$$

For the second term, we use Corollary 7.3: $\mathbb{E}|\hat{\lambda} - \lambda|^{2p} = \frac{\mathbb{E}|N_T - \lambda T|^{2p}}{T^{2p}} = \frac{\mathcal{O}(\lambda^p)}{T^p}$. For the first term, we apply Lemma 7.1 conditional to N_T :

$$\mathbb{E}[\hat{\lambda}^{2p} \|\widehat{\mathcal{L}f} - \mathcal{L}f\|_\infty^{2p}] \leq \mathbb{E} \left[C(p) \frac{\hat{\lambda}^{2p}}{N_T^p} \right] = \frac{C(p)}{T^p} \mathbb{E}[\hat{\lambda}^p] = \frac{\mathcal{O}(\lambda^p)}{T^p}.$$

Finally:

$$\mathbb{E}[(\hat{\rho}_\delta - \rho_\delta)^{2p}] \leq \frac{C(p, \lambda)}{c^{2p}(1 - \theta)^{2p} T^p},$$

with $C(p, \lambda) = \mathcal{O}(\lambda^p)$. □

Now, we can prove Theorem 2.5.

Proof of Theorem 2.5. By Pythagoras theorem:

$$\begin{aligned} \mathbb{E}\|g - \hat{g}_{m_2}\|_{L^2}^2 &= \|g - g_{m_2}\|_{L^2}^2 + \mathbb{E}\|g_{m_2} - \hat{g}_{m_2}\|_{L^2}^2 = \|g - g_{m_2}\|_{L^2}^2 + \sum_{k=0}^{m_2-1} \mathbb{E}[(\hat{b}_k - b_k)^2], \\ \mathbb{E}\|h - \hat{h}_{m_3}\|_{L^2}^2 &= \|h - h_{m_3}\|_{L^2}^2 + \mathbb{E}\|h_{m_3} - \hat{h}_{m_3}\|_{L^2}^2 = \|h - h_{m_3}\|_{L^2}^2 + \sum_{k=0}^{m_3-1} \mathbb{E}[(\hat{c}_k - c_k)^2], \end{aligned}$$

hence we need to control the variance terms $\sum_{k=0}^{m_2-1} \mathbb{E}[(\hat{b}_k - b_k)^2]$ and $\sum_{k=0}^{m_3-1} \mathbb{E}[(\hat{c}_k - c_k)^2]$.

Using equations (4.17) to (4.21) and (4.10) to (4.14) in Zhang and Su [2018], we can obtain equations (32) and (33) below. Still, we give the proofs of these equations for the sake of completeness.

We notice that \hat{b}_k and \hat{c}_k (defined by (12) and (13)) can be written as:

$$\frac{1}{cT} \sum_{i=1}^{N_T} \int_0^{+\infty} F(u, X_i, \hat{\rho}_\delta) \psi_k(u) du,$$

and that the coefficients b_k and c_k (defined by (9) and (10)) can be written as:

$$\mathbb{E} \left[\frac{1}{cT} \sum_{i=1}^{N_T} \int_0^{+\infty} F(u, X_i, \rho_\delta) \psi_k(u) du \right],$$

where F is given by:

$$F(u, X, \rho) := \begin{cases} e^{-\rho(X-u)} \mathbf{1}_{X>u} & \text{for the coefficients of } g, \\ \int_u^X e^{-\rho(X-x)} w(x, X-x) dx \mathbf{1}_{X>u} & \text{for the coefficients of } h. \end{cases} \quad (27)$$

Thus, we need to give an upper bound on quantities of the form:

$$V_m := \sum_{k=0}^{m-1} \mathbb{E} \left[\left(\frac{1}{cT} \sum_{i=1}^{N_T} \int_0^{+\infty} F(u, X_i, \hat{\rho}_\delta) \psi_k(u) du - \mathbb{E} \left[\frac{1}{cT} \sum_{i=1}^{N_T} \int_0^{+\infty} F(u, X_i, \rho_\delta) \psi_k(u) du \right] \right)^2 \right]. \quad (28)$$

The bound on V_m is based on the following decomposition:

$$\frac{1}{cT} \sum_{i=1}^{N_T} \int_0^{+\infty} F(u, X_i, \hat{\rho}_\delta) \psi_k(u) du - \mathbb{E} \left[\frac{1}{cT} \sum_{i=1}^{N_T} \int_0^{+\infty} F(u, X_i, \rho_\delta) \psi_k(u) du \right] = Z_k + \Delta_k \quad (29)$$

where:

$$Z_k := \frac{1}{cT} \left(\sum_{i=1}^{N_T} \int_0^{+\infty} F(u, X_i, \hat{\rho}_\delta) \psi_k(u) du - \mathbb{E} \left[\sum_{i=1}^{N_T} \int_0^{+\infty} F(u, X_i, \hat{\rho}_\delta) \psi_k(u) du \right] \right)$$

$$\Delta_k := \frac{1}{cT} \sum_{i=1}^{N_T} \int_0^{+\infty} [F(u, X_i, \hat{\rho}_\delta) - F(u, X_i, \rho_\delta)] \psi_k(u) du.$$

Let us notice that if $\delta = 0$, then $\hat{\rho}_\delta = \rho_\delta = 0$, so $\Delta_k = 0$ and the decomposition reduces to Z_k .

- Bound on $\sum_{k=0}^{m-1} \mathbb{E}[Z_k^2]$. This bound is obtained by a projection argument:

$$\begin{aligned} \sum_{k=0}^{m-1} \mathbb{E}[Z_k^2] &= \sum_{k=0}^{m-1} \text{Var} \left(\frac{1}{cT} \sum_{i=1}^{N_T} \int_0^{+\infty} F(u, X_i, \rho_\delta) \psi_k(u) du \right) \\ &= \sum_{k=0}^{m-1} \frac{\lambda}{c^2 T} \mathbb{E} \left[\left(\int_0^{+\infty} F(u, X, \rho_\delta) \psi_k(u) du \right)^2 \right] \\ &\leq \frac{\lambda}{c^2 T} \mathbb{E} \left[\int_0^{+\infty} F(u, X, \rho_\delta)^2 du \right], \end{aligned}$$

where the last inequality comes from the fact that $(\psi_k)_{k \geq 0}$ is an orthonormal basis of $L^2(\mathbb{R}_+)$. From (27), we see that:

$$\frac{\lambda}{c^2 T} \mathbb{E} \left[\int_0^{+\infty} F(u, X, \rho_\delta)^2 du \right] \leq \begin{cases} \frac{\lambda}{c^2 T} \mathbb{E}[X] & \text{for the coefficients of } g, \\ \frac{\lambda}{c^2 T} \mathbb{E}[W(X)] & \text{for the coefficients of } h. \end{cases} \quad (30)$$

where $W(X)$ is defined in Assumption 3. In the $\delta = 0$ case, this gives the desired results.

- Bound on $\sum_{k=0}^{m-1} \Delta_k^2$. We use a projection argument again:

$$\begin{aligned} \sum_{k=0}^{m-1} \Delta_k^2 &\leq \sum_{k=0}^{m-1} \frac{N_T}{c^2 T^2} \sum_{i=1}^{N_T} \left(\int_0^{+\infty} [F(u, X_i, \hat{\rho}_\delta) - F(u, X_i, \rho_\delta)] \psi_k(u) du \right)^2 \\ &\leq \frac{\hat{\lambda}}{c^2 T} \sum_{i=1}^{N_T} \int_0^{+\infty} |F(u, X_i, \hat{\rho}_\delta) - F(u, X_i, \rho_\delta)|^2 du, \end{aligned}$$

where $\hat{\lambda} := \frac{N_T}{T}$. By Remark 1.3, we know that $\rho_\delta \in [\frac{\delta}{c}, \frac{\delta+\lambda}{c}]$ and $\hat{\rho}_\delta \in [\frac{\delta}{c}, \frac{\delta+\hat{\lambda}}{c}]$, so by the mean value theorem:

$$|F(u, X_i, \hat{\rho}_\delta) - F(u, X_i, \rho_\delta)| \leq |\hat{\rho}_\delta - \rho_\delta| \sup_{\rho \geq \frac{\delta}{c}} \left| \frac{\partial F}{\partial \rho}(u, X_i, \rho) \right|.$$

Since the function $te^{-\rho t}\mathbf{1}_{t>0}$ achieves its maximum at $t = \frac{1}{\rho}$, we see that:

$$\sup_{\rho \geq \frac{\delta}{c}} \left| \frac{\partial F}{\partial \rho}(u, X_i, \rho) \right| \leq \begin{cases} \frac{c}{e\delta} \mathbf{1}_{X_i > u} & \text{for the coefficients of } g, \\ \frac{c}{e\delta} \int_u^{X_i} w(x, X_i - x) dx \mathbf{1}_{X_i > u} & \text{for the coefficients of } h. \end{cases}$$

Thus,

$$\sum_{k=0}^{m-1} \Delta_k^2 \leq \frac{\hat{\lambda}|\hat{\rho}_\delta - \rho_\delta|^2}{e^2\delta^2} \times \begin{cases} \frac{1}{T} \sum_{i=1}^{N_T} X_i & \text{for the coefficients of } g, \\ \frac{1}{T} \sum_{i=1}^{N_T} W(X_i) & \text{for the coefficients of } h. \end{cases} \quad (31)$$

Using the decomposition (29) in (28), we get $V_m \leq 2 \sum_{k=0}^{m-1} \mathbb{E}[Z_k^2] + 2 \sum_{k=0}^{m-1} \mathbb{E}[\Delta_k^2]$. Combining (30) and (31) yields:

$$\mathbb{E}\|\hat{g}_{m_2} - g_{m_2}\|_{L^2}^2 \leq 2 \frac{\lambda}{c^2 T} \mathbb{E}[X] + 2 \mathbb{E} \left[\frac{\hat{\lambda}|\hat{\rho}_\delta - \rho_\delta|^2}{e^2\delta^2} \frac{1}{T} \sum_{i=1}^{N_T} X_i \right], \quad (32)$$

$$\mathbb{E}\|\hat{h}_{m_3} - h_{m_3}\|_{L^2}^2 \leq 2 \frac{\lambda}{c^2 T} \mathbb{E}[W(X)] + 2 \mathbb{E} \left[\frac{\hat{\lambda}|\hat{\rho}_\delta - \rho_\delta|^2}{e^2\delta^2} \frac{1}{T} \sum_{i=1}^{N_T} W(X_i) \right]. \quad (33)$$

We apply Hölder's inequality on the second term in (32) and we use Proposition 7.4:

$$\begin{aligned} \mathbb{E} \left[\hat{\lambda}|\hat{\rho}_\delta - \rho_\delta|^2 \frac{1}{T} \sum_{i=1}^{N_T} X_i \right] &\leq \mathbb{E}[\hat{\lambda}^4]^{1/4} \mathbb{E}[|\hat{\rho}_\delta - \rho_\delta|^8]^{1/4} \mathbb{E} \left[\left(\frac{1}{T} \sum_{i=1}^{N_T} X_i \right)^2 \right]^{1/2} \\ &\leq \frac{C(\lambda)}{c^2(1-\theta)^2 T} \mathbb{E} \left[\left(\frac{1}{T} \sum_{i=1}^{N_T} X_i \right)^2 \right]^{1/2}, \end{aligned}$$

with $C(\lambda) = \mathcal{O}(\lambda^2)$. We need to evaluate this last expectation:

$$\mathbb{E} \left[\left(\frac{1}{T} \sum_{i=1}^{N_T} X_i \right)^2 \right] \leq \mathbb{E} \left[\frac{N_T}{T^2} \sum_{i=1}^{N_T} X_i^2 \right] = \mathbb{E} \left[\frac{N_T^2}{T^2} \right] \mathbb{E}[X^2] = \left(\frac{\lambda}{T} + \lambda^2 \right) \mathbb{E}[X^2].$$

Thus, we obtain:

$$\mathbb{E}\|\hat{g}_m - g_m\|_{L^2}^2 \leq 2 \frac{\lambda}{c^2 T} \mathbb{E}[X] + 2 \frac{C(\lambda)}{c^2 T(1-\theta)^2 \delta^2} \mathbb{E}[X^2]^{1/2}$$

with $C(\lambda) = \mathcal{O}(\lambda^2)$. We make the same reasoning for h , replacing X_i by $W(X_i)$. \square

7.2 Proofs of Section 3

Let us recall some facts about Toeplitz matrices; the interested reader can find more details in the book of Böttcher and Grudsky [2000]. Given $(\alpha_n)_{n \in \mathbb{Z}}$ a sequence of complex numbers, a Toeplitz matrix is an infinite matrix of the form:

$$\begin{pmatrix} \alpha_0 & \alpha_{-1} & \alpha_{-2} & \cdots \\ \alpha_1 & \alpha_0 & \alpha_{-1} & \cdots \\ \alpha_2 & \alpha_1 & \alpha_0 & \cdots \\ \cdots & \cdots & \cdots & \cdots \end{pmatrix}. \quad (34)$$

The classical result from O. Toeplitz says that this matrix induces a bounded operator on $\ell^2(\mathbb{N})$ if and only if $(\alpha_n)_{n \in \mathbb{Z}}$ are the Fourier coefficients of some function $\alpha \in L^\infty(\mathbb{T})$, where \mathbb{T} denotes the complex unit circle. We denote both the matrix (34) and its induced operator on $\ell^2(\mathbb{N})$ by $\mathbf{T}(\alpha)$, the function α being called *the symbol* of the Toeplitz matrix. Finally, if $m \in \mathbb{N}^*$ and if $\mathbf{T}(\alpha)$ is a Toeplitz matrix, we denote by $\mathbf{T}_m(\alpha)$ the $m \times m$ matrix:

$$\mathbf{T}_m(\alpha) := \begin{pmatrix} \alpha_0 & \cdots & \alpha_{-(m-1)} \\ \vdots & \ddots & \vdots \\ \alpha_{m-1} & \cdots & \alpha_0 \end{pmatrix}. \quad (35)$$

The operator norm of $\mathbf{T}(\alpha)$ depends on the properties of its symbol. In the case where $\alpha_k = 0$ for all $k < 0$, we have the following lemma.

Lemma 7.5. *Let $(\alpha_k)_{k \geq 0} \in \ell^1(\mathbb{N})$ be a sequence of complex numbers. Then the Toeplitz matrix $\mathbf{T}(\alpha)$ is lower triangular and we have:*

$$\forall x \in \ell^2(\mathbb{N}), \quad \mathbf{T}(\alpha) \times x = \alpha * x.$$

In particular, $\|\mathbf{T}(\alpha)\|_{\text{op}} \leq \|\alpha\|_{\ell^1}$.

Proof. The fact that $\mathbf{T}(\alpha)$ is lower triangular and that $\mathbf{T}(\alpha) \times x = \alpha * x$ is clear from the definition of a Toeplitz matrix. Then, Young's inequality for convolution yields $\|\alpha * x\|_{\ell^2} \leq \|\alpha\|_{\ell^1} \|x\|_{\ell^2}$. \square

Concerning the inverse of a Toeplitz matrix, its norm depends on the position of zero relatively to the range of the symbol. More precisely, we use the following result.

Lemma 7.6 (Lemma 3.8 in Böttcher and Grudsky [2000]). *Let $\alpha \in L^\infty(\mathbb{T})$ and let $E(\alpha)$ be the convex hull of the essential range of α . If $d := \text{dist}(0, E(\alpha)) > 0$, then $\mathbf{T}_m(\alpha)$ is invertible for all $m \geq 1$, and we have $\|\mathbf{T}_m^{-1}(\alpha)\|_{\text{op}} < \frac{2}{d}$.*

The matrix \mathbf{A}_m defined by (16) is a Toeplitz matrix and its symbol is given by:

$$\alpha(t) := \sum_{k=0}^{+\infty} \alpha_k t^k, \quad \text{with} \quad \alpha_k := \begin{cases} 1 - \frac{b_0}{\sqrt{2}} & \text{if } k = 0, \\ \frac{b_{k-1} - b_k}{\sqrt{2}} & \text{if } k \geq 1. \end{cases}$$

Let us notice that under Assumption 4, we have $(\alpha_k)_{k \geq 0} \in \ell^1(\mathbb{N})$ so the symbol α is continuous on \mathbb{T} , and thus $\alpha \in L^\infty(\mathbb{T})$.

Proof of Lemma 3.2. We apply Lemma C.1² in Comte et al. [2017] to the coefficients of g : the sequence $(\beta_k)_{k \geq 0}$, defined by $\beta_0 := \frac{b_0}{\sqrt{2}}$ and $\beta_k := \frac{b_k - b_{k-1}}{\sqrt{2}}$ for $k \geq 1$, are the Fourier coefficients of the function $t \in \mathbb{T} \mapsto \mathcal{L}g\left(\frac{1+t}{1-t}\right) \in \mathbb{C}$. Thus, we have:

$$\forall t \in \mathbb{T}, \quad \mathcal{L}g\left(\frac{1+t}{1-t}\right) = \sum_{k=0}^{+\infty} \beta_k t^k,$$

with the convention $\mathcal{L}g(\infty) = 0$. Since $\alpha(t) = 1 - \sum_{k \geq 0} \beta_k t^k$, we get:

$$\forall t \in \mathbb{T}, \quad \alpha(t) = 1 - \mathcal{L}g\left(\frac{1+t}{1-t}\right),$$

We notice that if $t \in \mathbb{T} \setminus \{1\}$, then there exists $\omega \in \mathbb{R}$ such that $\frac{1+t}{1-t} = i\omega$. Thus:

$$\begin{aligned} \forall t \in \mathbb{T} \setminus \{1\}, \quad \Re \alpha(t) &= 1 - \Re \left[\mathcal{L}g\left(\frac{1+t}{1-t}\right) \right] \\ &= 1 - \Re [\mathcal{L}g(i\omega)] \\ &= 1 - \Re \left[\int_0^{+\infty} e^{-i\omega x} g(x) dx \right] \\ &= 1 - \int_0^{+\infty} \cos(\omega x) g(x) dx \\ &\geq 1 - \int_0^{+\infty} g(x) dx \geq 1 - \theta. \end{aligned}$$

This inequality holds for $t = 1$ as well, hence $\alpha(\mathbb{T})$ is included in the half-plane $\{z \in \mathbb{C} \mid \Re(z) \geq 1 - \theta\}$, and so is its convex hull. By Lemma 7.6:

$$\|\mathbf{A}_m^{-1}\|_{\text{op}} \leq \frac{2}{1 - \theta}. \quad \square$$

Remark 7.7. In their article, Zhang and Su [2018] show that $\inf_{|z|=1} |\alpha(z)| \geq 1 - \theta > 0$, that is $\text{dist}(0, \alpha(\mathbb{T})) > 0$, which is not sufficient to apply Lemma 7.6.

²this lemma is stated for the generalized Laguerre basis, which depends on a parameter a . This parameter is equal to 1 in our case.

7.2.1 Proof of Theorem 3.3

Proposition 7.8. *Under Assumption 4, if $\theta < \theta_0$ then it holds:*

$$\forall m \in \mathbb{N}^*, \quad \mathbb{E} \|(\tilde{\mathbf{A}}_{m,1}^{-1} - \mathbf{A}_m^{-1})\mathbf{c}_m\|_{\ell^2}^2 \leq \frac{C(\theta, \theta_0)}{(1 - \theta_0)^2} \|\phi_m\|_{L^2}^2 \mathbb{E} \|\hat{\mathbf{A}}_m - \mathbf{A}_m\|_{\text{op}}^2,$$

where $C(\theta, \theta_0)$ is a constant satisfying $C(\theta, \theta_0) \lesssim \frac{(1-\theta_0)^2}{(\theta_0-\theta)^2}$.

Proof. We decompose the expectation according to the event $\Delta_m^1 := \{\|\hat{\mathbf{A}}_m^{-1}\|_{\text{op}} \leq \frac{2}{1-\theta_0}\}$:

$$\begin{aligned} \mathbb{E} \|(\tilde{\mathbf{A}}_{m,1}^{-1} - \mathbf{A}_m^{-1})\mathbf{c}_m\|_{\ell^2}^2 &= \|\mathbf{A}_m^{-1}\mathbf{c}_m\|_{\ell^2}^2 \mathbb{P}[\Delta_m^c] + \mathbb{E} \left[\|\tilde{\mathbf{A}}_{m,1}^{-1}(\mathbf{A}_m - \hat{\mathbf{A}}_m)\mathbf{A}_m^{-1}\mathbf{c}_m\|_{\ell^2}^2 \mathbf{1}_{\Delta_m^c} \right] \\ &\leq \|\mathbf{A}_m^{-1}\mathbf{c}_m\|_{\ell^2}^2 \left(\mathbb{P}[\Delta_m^c] + \frac{4}{(1 - \theta_0)^2} \mathbb{E} \|\hat{\mathbf{A}}_m - \mathbf{A}_m\|_{\text{op}}^2 \right) \\ &= \|\mathbf{a}_m\|_{\ell^2}^2 \left(\mathbb{P}[\Delta_m^c] + \frac{4}{(1 - \theta_0)^2} \mathbb{E} \|\hat{\mathbf{A}}_m - \mathbf{A}_m\|_{\text{op}}^2 \right). \end{aligned}$$

Since $\theta < \theta_0$ and $\|\mathbf{A}_m^{-1}\|_{\text{op}} \leq \frac{2}{1-\theta}$ (see Lemma 3.2), we get:

$$\begin{aligned} \mathbb{P}[\Delta_m^c] &\leq \mathbb{P} \left[\|\hat{\mathbf{A}}_m^{-1} - \mathbf{A}_m^{-1}\|_{\text{op}} > \frac{2}{1 - \theta_0} - \frac{2}{1 - \theta} \right] \\ &= \mathbb{P} \left[\left\{ \|\hat{\mathbf{A}}_m^{-1} - \mathbf{A}_m^{-1}\|_{\text{op}} > \frac{2}{1 - \theta_0} - \frac{2}{1 - \theta} \right\} \cap \left\{ \|\mathbf{A}_m^{-1}(\hat{\mathbf{A}}_m - \mathbf{A}_m)\|_{\text{op}} < \frac{1}{2} \right\} \right] \\ &\quad + \mathbb{P} \left[\left\{ \|\hat{\mathbf{A}}_m^{-1} - \mathbf{A}_m^{-1}\|_{\text{op}} > \frac{2}{1 - \theta_0} - \frac{2}{1 - \theta} \right\} \cap \left\{ \|\mathbf{A}_m^{-1}(\hat{\mathbf{A}}_m - \mathbf{A}_m)\|_{\text{op}} \geq \frac{1}{2} \right\} \right]. \end{aligned}$$

- First term. We apply Theorem C.1 and we conclude using Markov's inequality:

$$\begin{aligned} &\mathbb{P} \left[\left\{ \|\hat{\mathbf{A}}_m^{-1} - \mathbf{A}_m^{-1}\|_{\text{op}} > \frac{2}{1 - \theta_0} - \frac{2}{1 - \theta} \right\} \cap \left\{ \|\mathbf{A}_m^{-1}(\hat{\mathbf{A}}_m - \mathbf{A}_m)\|_{\text{op}} < \frac{1}{2} \right\} \right] \\ &\leq \mathbb{P} \left[\left\{ \frac{\|\mathbf{A}_m^{-1}\|_{\text{op}}^2 \|\hat{\mathbf{A}}_m - \mathbf{A}_m\|_{\text{op}}}{1 - \|\mathbf{A}_m^{-1}(\hat{\mathbf{A}}_m - \mathbf{A}_m)\|_{\text{op}}} > \frac{2}{1 - \theta_0} - \frac{2}{1 - \theta} \right\} \cap \left\{ \|\mathbf{A}_m^{-1}(\hat{\mathbf{A}}_m - \mathbf{A}_m)\|_{\text{op}} < \frac{1}{2} \right\} \right] \\ &\leq \mathbb{P} \left[\|\mathbf{A}_m^{-1}\|_{\text{op}}^2 \|\hat{\mathbf{A}}_m - \mathbf{A}_m\|_{\text{op}} > \frac{1}{1 - \theta_0} - \frac{1}{1 - \theta} \right] \\ &\leq \frac{(1 - \theta)^2 (1 - \theta_0)^2}{(\theta_0 - \theta)^2} \|\mathbf{A}_m^{-1}\|_{\text{op}}^4 \mathbb{E} \|\hat{\mathbf{A}}_m - \mathbf{A}_m\|_{\text{op}}^2 \leq \frac{16(1 - \theta_0)^2}{(\theta_0 - \theta)^2 (1 - \theta)^2} \mathbb{E} \|\hat{\mathbf{A}}_m - \mathbf{A}_m\|_{\text{op}}^2. \end{aligned}$$

- Second term. We use Markov's inequality:

$$\begin{aligned} &\mathbb{P} \left[\left\{ \|\hat{\mathbf{A}}_m^{-1} - \mathbf{A}_m^{-1}\|_{\text{op}} > \frac{2}{1 - \theta_0} - \frac{2}{1 - \theta} \right\} \cap \left\{ \|\mathbf{A}_m^{-1}(\hat{\mathbf{A}}_m - \mathbf{A}_m)\|_{\text{op}} \geq \frac{1}{2} \right\} \right] \\ &\leq \mathbb{P} \left[\|\mathbf{A}_m^{-1}(\hat{\mathbf{A}}_m - \mathbf{A}_m)\|_{\text{op}} \geq \frac{1}{2} \right] \\ &\leq 4 \|\mathbf{A}_m^{-1}\|_{\text{op}}^2 \mathbb{E} \|\hat{\mathbf{A}}_m - \mathbf{A}_m\|_{\text{op}}^2 \leq \frac{16}{(1 - \theta)^2} \mathbb{E} \|\hat{\mathbf{A}}_m - \mathbf{A}_m\|_{\text{op}}^2. \end{aligned}$$

Thus, we obtain:

$$\mathbb{E} \|(\tilde{\mathbf{A}}_{m,1}^{-1} - \mathbf{A}_m^{-1})\mathbf{c}_m\|_{\ell^2}^2 \leq \|\phi_m\|_{L^2}^2 \left(\frac{16}{(1 - \theta)^2} \left(1 + \frac{(1 - \theta_0)^2}{(\theta_0 - \theta)^2} \right) + \frac{4}{(1 - \theta_0)^2} \right) \mathbb{E} \|\hat{\mathbf{A}}_m - \mathbf{A}_m\|_{\text{op}}^2. \quad \square$$

We can now prove Theorem 3.3.

Proof of Theorem 3.3. By Pythagoras Theorem, $\|\phi - \hat{\phi}_m^{\text{Lag}_1}\|_{L^2}^2 = \|\phi - \phi_m\|_{L^2}^2 + \|\phi_m - \hat{\phi}_m^{\text{Lag}}\|_{L^2}^2$.

$$\begin{aligned} \mathbb{E} \|\phi_m - \hat{\phi}_m^{\text{Lag}_1}\|_{L^2}^2 &= \mathbb{E} \|\hat{\mathbf{a}}_m^{\text{Lag}_1} - \mathbf{a}_m\|_{\ell^2}^2 = \mathbb{E} \|\tilde{\mathbf{A}}_{m,1}^{-1} \hat{\mathbf{c}}_m - \mathbf{A}_m^{-1} \mathbf{c}_m\|_{\ell^2}^2 \\ &\leq 3 \mathbb{E} \|(\tilde{\mathbf{A}}_{m,1}^{-1} - \mathbf{A}_m^{-1})\mathbf{c}_m\|_{\ell^2}^2 + 3 \mathbb{E} \|(\mathbf{A}_m^{-1} - \tilde{\mathbf{A}}_{m,1}^{-1})(\mathbf{c}_m - \hat{\mathbf{c}}_m)\|_{\ell^2}^2 + 3 \mathbb{E} \|\mathbf{A}_m^{-1}(\mathbf{c}_m - \hat{\mathbf{c}}_m)\|_{\ell^2}^2. \end{aligned}$$

- First term. We apply Proposition 7.8 with Lemma 7.5:

$$\begin{aligned}
\mathbb{E}\|(\tilde{\mathbf{A}}_{m,1}^{-1} - \mathbf{A}_m^{-1})\mathbf{c}_m\|_{\ell^2}^2 &\leq C(\theta, \theta_0) \|\phi_m\|_{L^2}^2 \mathbb{E}\|\hat{\mathbf{A}}_m - \mathbf{A}_m\|_{L^2}^2 \\
&\leq C(\theta, \theta_0) \|\phi_m\|_{L^2}^2 \mathbb{E}\|\hat{\mathbf{b}}_m - \mathbf{b}_m\|_{\ell^1}^2 \\
&\leq C(\theta, \theta_0) \|\phi_m\|_{L^2}^2 m \mathbb{E}\|\hat{\mathbf{b}}_m - \mathbf{b}_m\|_{\ell^2}^2 \\
&= C(\theta, \theta_0) \|\phi\|_{L^2}^2 m \mathbb{E}\|\hat{g}_m - g_m\|_{L^2}^2.
\end{aligned}$$

- Second term.

$$\begin{aligned}
\mathbb{E}\|(\mathbf{A}_m^{-1} - \tilde{\mathbf{A}}_{m,1}^{-1})(\mathbf{c}_m - \hat{\mathbf{c}}_m)\|_{\ell^2}^2 &\leq \mathbb{E}\left[\|\mathbf{A}_m^{-1} - \tilde{\mathbf{A}}_{m,1}^{-1}\|_{\text{op}}^2 \|\mathbf{c}_m - \hat{\mathbf{c}}_m\|_{\ell^2}^2\right] \\
&\leq \left(\frac{8}{(1-\theta)^2} + \frac{8}{(1-\theta_0)^2}\right) \mathbb{E}\|\hat{h}_m - h_m\|_{L^2}^2.
\end{aligned}$$

- Third term.

$$\begin{aligned}
\mathbb{E}\|\mathbf{A}_m^{-1}(\mathbf{c}_m - \hat{\mathbf{c}}_m)\|_{\ell^2}^2 &\leq \|\mathbf{A}_m^{-1}\|_{\text{op}}^2 \mathbb{E}\|\mathbf{c}_m - \hat{\mathbf{c}}_m\|_{\ell^2}^2 \\
&\leq \frac{4}{(1-\theta)^2} \mathbb{E}\|\hat{h}_m - h_m\|_{L^2}^2.
\end{aligned}$$

$$\mathbb{E}\|\phi_m - \hat{\phi}_m^{\text{Lag}_1}\|_{L^2}^2 \leq 3 \times \frac{C(\theta, \theta_0)}{(1-\theta_0)^2} \|\phi\|_{L^2}^2 m \mathbb{E}\|\hat{g}_m - g_m\|_{L^2}^2 + \frac{60}{(1-\theta_0)^2} \mathbb{E}\|\hat{h}_m - h_m\|_{L^2}^2,$$

with $C(\theta, \theta_0) \lesssim \frac{(1-\theta_0)^2}{(1-\theta)^2}$. To conclude, we use the upper bounds established in the proof of Theorem 2.5. If $\delta = 0$, we have:

$$\mathbb{E}\|\hat{g}_m - g_m\|_{L^2}^2 \leq \frac{\lambda}{c^2 T} \mathbb{E}[X], \quad \mathbb{E}\|\hat{h}_m - h_m\|_{L^2}^2 \leq \frac{\lambda}{c^2 T} \mathbb{E}[W(X)],$$

and if $\delta > 0$, we have:

$$\begin{aligned}
\mathbb{E}\|\hat{g}_m - g_m\|_{L^2}^2 &\leq \frac{C(\lambda)}{c^2 T} \left(\mathbb{E}[X] + \frac{\mathbb{E}[X^2]^{\frac{1}{2}}}{(1-\theta)^2 \delta^2} \right), \\
\mathbb{E}\|\hat{h}_m - h_m\|_{L^2}^2 &\leq \frac{C(\lambda)}{c^2 T} \left(\mathbb{E}[W(X)] + \frac{\mathbb{E}[W(X)^2]^{\frac{1}{2}}}{(1-\theta)^2 \delta^2} \right),
\end{aligned}$$

with $C(\lambda) = \mathcal{O}(\lambda^2)$. □

7.2.2 Proof of Proposition 3.5

Let us introduce the sequence of functions $(D_k)_{k \geq 0}$ as:

$$D_k(x) := \begin{cases} \frac{\Psi_0(x)}{\sqrt{2}} & \text{if } k = 0, \\ \frac{\Psi_k(x) - \Psi_{k-1}(x)}{\sqrt{2}} & \text{if } k \geq 1. \end{cases}$$

so we can rewrite $\mathbf{A}_m = \mathbf{I}_m - \frac{\lambda}{c} \mathbf{T}_m(\mathbb{E}[D(X)])$ and $\hat{\mathbf{A}}_m = \mathbf{I}_m - \frac{1}{cT} \sum_{i=1}^{N_T} \mathbf{T}_m(D(X_i))$, with $\mathbf{T}_m(\bullet)$ defined by (35). Now, the difference between $\hat{\mathbf{A}}_m$ and \mathbf{A}_m can be decomposed as:

$$\hat{\mathbf{A}}_m - \mathbf{A}_m = \frac{1}{cT} \sum_{i=1}^{N_T} \{\mathbf{T}_m(D(X_i)) - \mathbb{E}[\mathbf{T}_m(D(X_i))]\} + \frac{N_T}{cT} \mathbf{T}_m(\mathbb{E}[D(X)]) - \frac{\lambda}{c} \mathbf{T}_m(\mathbb{E}[D(X)]). \quad (36)$$

The next lemma gives a control on the first term in the decomposition (36).

Lemma 7.9. Let $\mathbf{S}_n := \sum_{i=1}^n \mathbf{Z}_i$, with $\mathbf{Z}_i := \mathbf{T}_m(D(X_i)) - \mathbb{E}[\mathbf{T}_m(D(X_i))]$. Then for $p \geq 1$ and $\log m \geq p$, we have:

$$\mathbb{E} \|\mathbf{S}_n\|_{\text{op}}^{2p} \leq C(p) \left[(n\mu m \log m)^p + (m \log m)^{2p} \right],$$

with $C(p)$ a constant depending on p .

Proof. We want to apply Theorem C.2. First, we need upper bounds on $\|\mathbf{Z}_i\|_{\text{op}}$ and $\lambda_{\max}(\mathbb{E}[\mathbf{S}_n^\top \mathbf{S}_n])$.

- Bound on $\|\mathbf{Z}_i\|_{\text{op}}$:

$$\begin{aligned} \|\mathbf{Z}_i\|_{\text{op}} &= \sup_{\|x\|_{\ell^2} \leq 1} \|(\mathbf{T}_m(D(X_i)) - \mathbb{E}[\mathbf{T}_m(D(X_i))])x\|_{\ell^2} \\ &= \sup_{\|x\|_{\ell^2} \leq 1} \|(D(X_i) - \mathbb{E}[D(X_i)]) * x\|_{\ell^2} \\ &\leq \|D(X_i) - \mathbb{E}[D(X_i)]\|_{\ell^1} \\ &\leq \sqrt{2} \sum_{k=0}^{m-1} |\Psi_k(X_i) - \mathbb{E}[\Psi_k(X_i)]| \\ &\leq 2\sqrt{2} \sum_{k=0}^{m-1} \|\Psi_k\|_{\infty} \end{aligned}$$

By Lemma 3.4, there exists an absolute constant $C > 0$ such that $\|\Psi_k\|_{\infty} \leq C$, hence $\|\mathbf{Z}_i\|_{\text{op}} \leq C2\sqrt{2}m$.

- Bound on $\lambda_{\max}(\mathbb{E}[\mathbf{S}_n^\top \mathbf{S}_n])$:

$$\begin{aligned} \lambda_{\max}(\mathbb{E}[\mathbf{S}_n^\top \mathbf{S}_n]) &= \sup_{\|x\|_{\ell^2}=1} x^\top \mathbb{E}[\mathbf{S}_n^\top \mathbf{S}_n] x \\ &= n \sup_{\|x\|_{\ell^2}=1} x^\top \mathbb{E}[\mathbf{Z}_1^\top \mathbf{Z}_1] x \\ &= n \sup_{\|x\|_{\ell^2}=1} \mathbb{E}[\|\mathbf{Z}_1 x\|_{\ell^2}^2] \\ &= n \sup_{\|x\|_{\ell^2}=1} \mathbb{E}[\|(D(X_1) - \mathbb{E}[D(X_1)]) * x\|_{\ell^2}^2]. \end{aligned}$$

If $x \in \mathbb{R}^m$, we have:

$$\begin{aligned} \mathbb{E}[\|(D(X_1) - \mathbb{E}[D(X_1)]) * x\|_{\ell^2}^2] &= \sum_{j=0}^{m-1} \mathbb{E}[(\{D(X_1) - \mathbb{E}[D(X_1)]\} * x)_j^2] \\ &= \sum_{j=0}^{m-1} \text{Var}[(D(X_1) * x)_j] \leq \sum_{j=0}^{m-1} \mathbb{E}[(D(X_1) * x)_j^2], \end{aligned}$$

and by Cauchy–Schwarz inequality:

$$\begin{aligned} (D(X_1) * x)_j^2 &\leq \left(\sum_{k=0}^j D_k(X_1)^2 \right) \left(\sum_{k=0}^j x_k^2 \right) \\ &\leq \|x\|_{\ell^2}^2 \sum_{k=0}^j \Psi_k(X_1)^2 \leq \|x\|_{\ell^2}^2 \|\mathbf{1}_{X_1 > \bullet}\|_{\text{L}^2}^2 = \|x\|_{\ell^2}^2 X_1, \end{aligned}$$

because $\Psi_k(X_1) = \langle \mathbf{1}_{X_1 > \bullet}, \psi_k \rangle$ and (ψ_k) is an orthonormal basis of $\text{L}^2(\mathbb{R}_+)$. Hence, we obtain $\lambda_{\max}(\mathbb{E}[\mathbf{S}_n^\top \mathbf{S}_n]) \leq nm\mu$.

We want apply Theorem C.2 to our matrix \mathbf{S}_n , which is not Hermitian. We use the following trick, called the Paulsen dialtation. For \mathbf{M} a rectangular matrix, we define:

$$\mathbf{M} \mapsto \mathcal{H}(\mathbf{M}) = \begin{pmatrix} 0 & \mathbf{M} \\ \mathbf{M}^\dagger & 0 \end{pmatrix},$$

where \mathbf{M}^\dagger denotes the conjugate transpose of \mathbf{M} . Now, $\mathcal{H}(\mathbf{M})$ is an Hermitian matrix, and:

$$\mathcal{H}(\mathbf{M})^2 = \begin{pmatrix} \mathbf{M}\mathbf{M}^\dagger & 0 \\ 0 & \mathbf{M}^\dagger\mathbf{M} \end{pmatrix},$$

hence $\lambda_{\max}(\mathcal{H}(\mathbf{M})^2) = \|\mathbf{M}\|_{\text{op}}^2$ and $\lambda_{\max}(\mathcal{H}(\mathbf{M})) = \|\mathbf{M}\|_{\text{op}}$. We can now apply Theorem C.2: for $\mathbf{M} = \mathbf{S}_n$, we have:

$$\mathcal{H}(\mathbf{S}_n) = \begin{pmatrix} 0 & \sum_i \mathbf{Z}_i \\ \sum_i \mathbf{Z}_i^\top & 0 \end{pmatrix} = \sum_i \begin{pmatrix} 0 & \mathbf{Z}_i \\ \mathbf{Z}_i^\top & 0 \end{pmatrix} = \sum_i \mathcal{H}(\mathbf{Z}_i),$$

thus for $p \geq 1$ and $r \geq \max(2p, 2 \log m)$, we get that:

$$\begin{aligned} (\mathbb{E}\|\mathbf{S}_n\|_{\text{op}}^{2p})^{1/2p} &= \left(\mathbb{E} \lambda_{\max} \left(\mathcal{H} \left(\sum_i \mathbf{Z}_i \right) \right)^{2p} \right)^{1/2p} \\ &\leq \sqrt{er} \lambda_{\max}^{1/2} \left(\sum_i \mathbb{E} \mathcal{H}(\mathbf{Z}_i)^2 \right) + 2er \left(\mathbb{E} \max_i \lambda_{\max}(\mathcal{H}(\mathbf{Z}_i))^{2p} \right)^{1/2p} \\ &\leq \sqrt{er \lambda_{\max}(\mathbb{E} \mathbf{S}_n^\top \mathbf{S}_n)} + 2er \left(\mathbb{E} \max_i \|\mathbf{Z}_i\|_{\text{op}}^{2p} \right)^{1/2p} \\ &\leq \sqrt{ernm\mu} + C4\sqrt{2erm}. \end{aligned}$$

If $\log m \geq p$, then $r = 2 \log m$ and we get $\mathbb{E}\|\mathbf{S}_n\|_{\text{op}}^{2p} \lesssim 2^{2p-1} (n\mu m \log m)^p + 2^{6p-1} (m \log m)^{2p}$. \square

Now we can prove Proposition 3.5.

Proof of Proposition 3.5. From the decomposition (36), we get:

$$\mathbb{E}\|\hat{\mathbf{A}}_m - \mathbf{A}_m\|_{\text{op}}^{2p} \leq 2^{2p-1} \frac{1}{(cT)^{2p}} \mathbb{E}\|\mathbf{S}_{N_T}\|_{\text{op}}^{2p} + 2^{2p-1} \frac{\mathbb{E}|N_T - \lambda T|^{2p}}{(cT)^{2p}} \|\mathbf{T}_m(\mathbb{E}[D(X)])\|_{\text{op}}^{2p}.$$

For the first term, we apply Lemma 7.9 conditional on N_T :

$$\begin{aligned} \frac{1}{(cT)^{2p}} \mathbb{E}\|\mathbf{S}_{N_T}\|_{\text{op}}^{2p} &\leq C(p) \left[\frac{\mathbb{E}[N_T^p] \mu^p (m \log m)^p}{(cT)^{2p}} + \left(\frac{m \log m}{cT} \right)^{2p} \right] \\ &= C(p) \left[\mu^p \mathbb{E} \left[\left(\frac{N_T}{cT} \right)^p \right] \left(\frac{m \log m}{cT} \right)^p + \left(\frac{m \log m}{cT} \right)^{2p} \right], \end{aligned}$$

with $\mathbb{E} \left[\left(\frac{N_T}{cT} \right)^p \right] = \mathcal{O}(\lambda^p)$. For the second term, we know from Corollary 7.3 that $\mathbb{E}[(N_T - \lambda T)^{2p}] = \mathcal{O}(\lambda^p T^p)$, and:

$$\begin{aligned} \|\mathbf{T}_m(\mathbb{E}[D(X)])\|_{\text{op}} &\leq \sum_{k=0}^{m-1} |\mathbb{E}[D_k(X)]| \leq \sqrt{2} \sum_{k=0}^{m-1} |\mathbb{E}[\Psi_k(X)]| \\ &\leq \sqrt{2m} \left(\sum_{k=0}^{m-1} \mathbb{E}[\Psi_k(X)^2] \right)^{1/2} \\ &= \sqrt{2m} \left(\mathbb{E} \left[\sum_{k=0}^{m-1} \langle \mathbf{1}_{X > \bullet}, \psi_k \rangle^2 \right] \right)^{1/2} \leq \sqrt{2m\mu}, \end{aligned}$$

thus:

$$\frac{\mathbb{E}|N_T - \lambda T|^{2p}}{(cT)^{2p}} \|\mathbf{T}_m(\mathbb{E}[D(X)])\|_{\text{op}}^{2p} \leq \frac{\mathcal{O}(\lambda^p)}{T^p} \mu^p m^p. \quad \square$$

7.2.3 Proof of Theorem 3.6

The following results are based on the proofs of Lemma 3.1 and Corollary 3.2 in [Comte and Mabon \[2017\]](#).

Proposition 7.10. *If $m \log m \leq cT$, then it holds:*

$$\mathbb{E} \|\tilde{\mathbf{A}}_{m,2}^{-1} - \mathbf{A}_m^{-1}\|_{\text{op}}^{2p} \leq C(p, \lambda) (\mu^p \|\mathbf{A}_m^{-1}\|_{\text{op}}^{2p}) \wedge \left((\mu^p + \mu^{2p}) \|\mathbf{A}_m^{-1}\|_{\text{op}}^{4p} \left(\frac{m \log m}{cT} \right)^p \right),$$

with $C(p, \lambda) = \mathcal{O}(\lambda^p \vee \lambda^{2p})$.

Proof. We decompose the expectation according to the event $\Delta_m^2 := \{\|\hat{\mathbf{A}}_m^{-1}\|_{\text{op}}^1 \leq \frac{cT}{m \log m}\}$:

$$\begin{aligned} \mathbb{E} \left[\|\mathbf{A}_m^{-1} - \tilde{\mathbf{A}}_{m,2}^{-1}\|_{\text{op}}^{2p} \right] &= \mathbb{E} \left[\|\mathbf{A}_m^{-1}\|_{\text{op}}^{2p} \mathbf{1}_{\Delta_m^c} + \|\hat{\mathbf{A}}_m^{-1}(\mathbf{A}_m - \hat{\mathbf{A}}_m)\mathbf{A}_m^{-1}\|_{\text{op}}^{2p} \mathbf{1}_{\Delta_m} \right] \\ &= \|\mathbf{A}_m^{-1}\|_{\text{op}}^{2p} \mathbb{P}[\Delta_m^c] + \mathbb{E} \left[\|\hat{\mathbf{A}}_m^{-1}(\mathbf{A}_m - \hat{\mathbf{A}}_m)\mathbf{A}_m^{-1}\|_{\text{op}}^{2p} \mathbf{1}_{\Delta_m} \right]. \end{aligned} \quad (37)$$

We now give two bounds on (37), depending on the value of $\|\mathbf{A}_m^{-1}\|_{\text{op}}$.

- First case: $\|\mathbf{A}_m^{-1}\|_{\text{op}} > \frac{1}{2} \sqrt{\frac{cT}{m \log m}}$.

Starting from Equation (37) and using the set Δ_m^2 , we have that:

$$\begin{aligned} \mathbb{E} \left[\|\mathbf{A}_m^{-1} - \tilde{\mathbf{A}}_{m,2}^{-1}\|_{\text{op}}^{2p} \right] &\leq \|\mathbf{A}_m^{-1}\|_{\text{op}}^{2p} + \|\mathbf{A}_m^{-1}\|_{\text{op}}^{2p} \mathbb{E} \left[\|\hat{\mathbf{A}}_m^{-1}\|_{\text{op}}^{2p} \|\mathbf{A}_m - \hat{\mathbf{A}}_m\|_{\text{op}}^{2p} \mathbf{1}_{\Delta_m} \right] \\ &\leq \|\mathbf{A}_m^{-1}\|_{\text{op}}^{2p} + \|\mathbf{A}_m^{-1}\|_{\text{op}}^{2p} \left(\frac{cT}{m \log m} \right)^p \mathbb{E} \left[\|\mathbf{A}_m - \hat{\mathbf{A}}_m\|_{\text{op}}^{2p} \right]. \end{aligned}$$

We apply Proposition 3.5 and get:

$$\begin{aligned} \mathbb{E} \left[\|\mathbf{A}_m^{-1} - \tilde{\mathbf{A}}_{m,2}^{-1}\|_{\text{op}}^{2p} \right] &\leq \|\mathbf{A}_m^{-1}\|_{\text{op}}^{2p} + \|\mathbf{A}_m^{-1}\|_{\text{op}}^{2p} \left(\frac{cT}{m \log m} \right)^p C(p, \lambda) \mu^p \left(\frac{m \log m}{cT} \right)^p \\ &\leq (1 + C(p, \lambda) \mu^p) \|\mathbf{A}_m^{-1}\|_{\text{op}}^{2p}, \end{aligned}$$

with $C(p, \lambda) = \mathcal{O}(\lambda^p)$.

- Second case: $\|\mathbf{A}_m^{-1}\|_{\text{op}} < \frac{1}{2} \sqrt{\frac{cT}{m \log m}}$.

Starting from (37) again, we get:

$$\mathbb{E} \left[\|\mathbf{A}_m^{-1} - \tilde{\mathbf{A}}_{m,2}^{-1}\|_{\text{op}}^{2p} \right] \leq \|\mathbf{A}_m^{-1}\|_{\text{op}}^{2p} \mathbb{P}[\Delta_m^c] + \|\mathbf{A}_m^{-1}\|_{\text{op}}^{2p} \mathbb{E} \left[\|\mathbf{A}_m - \hat{\mathbf{A}}_m\|_{\text{op}}^{2p} \|\hat{\mathbf{A}}_m^{-1}\|_{\text{op}}^{2p} \mathbf{1}_{\Delta_m} \right].$$

1. Upper bound on $\mathbb{E} \left[\|\mathbf{A}_m - \hat{\mathbf{A}}_m\|_{\text{op}}^{2p} \|\hat{\mathbf{A}}_m^{-1}\|_{\text{op}}^{2p} \mathbf{1}_{\Delta_m} \right]$.

First let us notice that

$$\|\hat{\mathbf{A}}_m^{-1}\|_{\text{op}}^{2p} \leq 2^{2p-1} \|\hat{\mathbf{A}}_m^{-1} - \mathbf{A}_m^{-1}\|_{\text{op}}^{2p} + 2^{2p-1} \|\mathbf{A}_m^{-1}\|_{\text{op}}^{2p}.$$

Applying Proposition 3.5, we get:

$$\begin{aligned} &\mathbb{E} \left[\|\mathbf{A}_m - \hat{\mathbf{A}}_m\|_{\text{op}}^{2p} \|\hat{\mathbf{A}}_m^{-1}\|_{\text{op}}^{2p} \mathbf{1}_{\Delta_m} \right] \\ &\leq 2^{2p-1} \|\mathbf{A}_m^{-1}\|_{\text{op}}^{2p} \mathbb{E} \left[\|\mathbf{A}_m - \hat{\mathbf{A}}_m\|_{\text{op}}^{2p} \mathbf{1}_{\Delta_m} \right] + 2^{2p-1} \mathbb{E} \left[\|\mathbf{A}_m - \hat{\mathbf{A}}_m\|_{\text{op}}^{2p} \|\hat{\mathbf{A}}_m^{-1} - \mathbf{A}_m^{-1}\|_{\text{op}}^{2p} \mathbf{1}_{\Delta_m} \right] \\ &\leq 2^{2p-1} \|\mathbf{A}_m^{-1}\|_{\text{op}}^{2p} \mathbb{E} \left[\|\mathbf{A}_m - \hat{\mathbf{A}}_m\|_{\text{op}}^{2p} \mathbf{1}_{\Delta_m} \right] + 2^{2p-1} \|\mathbf{A}_m^{-1}\|_{\text{op}}^{2p} \mathbb{E} \left[\|\mathbf{A}_m - \hat{\mathbf{A}}_m\|_{\text{op}}^{4p} \|\hat{\mathbf{A}}_m^{-1}\|_{\text{op}}^{2p} \mathbf{1}_{\Delta_m} \right] \\ &\leq C(p, \lambda) \mu^p \|\mathbf{A}_m^{-1}\|_{\text{op}}^{2p} \left(\frac{m \log m}{cT} \right)^p + \|\mathbf{A}_m^{-1}\|_{\text{op}}^{2p} \left(\frac{cT}{m \log m} \right)^p C(2p, \lambda) \mu^{2p} \left(\frac{m \log m}{cT} \right)^{2p} \\ &\leq C'(p, \lambda) (\mu^p + \mu^{2p}) \|\mathbf{A}_m^{-1}\|_{\text{op}}^{2p} \left(\frac{m \log m}{cT} \right)^p, \end{aligned} \quad (38)$$

with $C'(p, \lambda) = \mathcal{O}(\lambda^p \vee \lambda^{2p})$.

2. Upper bound on $\mathbb{P}[\Delta_m^c] = \mathbb{P}\left[\|\hat{\mathbf{A}}_m^{-1}\|_{\text{op}} > \sqrt{\frac{cT}{m \log m}}\right]$.

From the triangular inequality:

$$\|\hat{\mathbf{A}}_m^{-1}\|_{\text{op}} \leq \|\hat{\mathbf{A}}_m^{-1} - \mathbf{A}_m^{-1}\|_{\text{op}} + \|\mathbf{A}_m^{-1}\|_{\text{op}}$$

we obtain:

$$\mathbb{P}\left[\|\hat{\mathbf{A}}_m^{-1}\|_{\text{op}} > \sqrt{\frac{cT}{m \log m}}\right] \leq \mathbb{P}\left[\|\hat{\mathbf{A}}_m^{-1} - \mathbf{A}_m^{-1}\|_{\text{op}} > \sqrt{\frac{cT}{m \log m}} - \|\mathbf{A}_m^{-1}\|_{\text{op}}\right].$$

Moreover we have assumed that $\|\mathbf{A}_m^{-1}\|_{\text{op}} < \frac{1}{2}\sqrt{\frac{cT}{m \log m}}$, so:

$$\mathbb{P}\left[\|\hat{\mathbf{A}}_m^{-1}\|_{\text{op}} > \sqrt{\frac{cT}{m \log m}}\right] \leq \mathbb{P}\left[\|\hat{\mathbf{A}}_m^{-1} - \mathbf{A}_m^{-1}\|_{\text{op}} > \|\mathbf{A}_m^{-1}\|_{\text{op}}\right].$$

Now let us rewrite this probability, as:

$$\begin{aligned} & \mathbb{P}\left[\|\hat{\mathbf{A}}_m^{-1} - \mathbf{A}_m^{-1}\|_{\text{op}} > \|\mathbf{A}_m^{-1}\|_{\text{op}}\right] \\ &= \mathbb{P}\left[\left\{\|\hat{\mathbf{A}}_m^{-1} - \mathbf{A}_m^{-1}\|_{\text{op}} > \|\mathbf{A}_m^{-1}\|_{\text{op}}\right\} \cap \left\{\|\mathbf{A}_m^{-1}(\hat{\mathbf{A}}_m - \mathbf{A}_m)\|_{\text{op}} < \frac{1}{2}\right\}\right] \\ &+ \mathbb{P}\left[\left\{\|\hat{\mathbf{A}}_m^{-1} - \mathbf{A}_m^{-1}\|_{\text{op}} > \|\mathbf{A}_m^{-1}\|_{\text{op}}\right\} \cap \left\{\|\mathbf{A}_m^{-1}(\hat{\mathbf{A}}_m - \mathbf{A}_m)\|_{\text{op}} \geq \frac{1}{2}\right\}\right] \\ &\leq \mathbb{P}\left[\left\{\|\hat{\mathbf{A}}_m^{-1} - \mathbf{A}_m^{-1}\|_{\text{op}} > \|\mathbf{A}_m^{-1}\|_{\text{op}}\right\} \cap \left\{\|\mathbf{A}_m^{-1}(\hat{\mathbf{A}}_m - \mathbf{A}_m)\|_{\text{op}} < \frac{1}{2}\right\}\right] \\ &+ \mathbb{P}\left[\|\mathbf{A}_m^{-1}(\hat{\mathbf{A}}_m - \mathbf{A}_m)\|_{\text{op}} \geq \frac{1}{2}\right]. \end{aligned} \quad (39)$$

To control the second term, we apply Markov inequality and Proposition 3.5:

$$\begin{aligned} \mathbb{P}\left[\|\mathbf{A}_m^{-1}(\hat{\mathbf{A}}_m - \mathbf{A}_m)\|_{\text{op}} \geq \frac{1}{2}\right] &\leq \mathbb{P}\left[\|\mathbf{A}_m^{-1}\|_{\text{op}} \|\hat{\mathbf{A}}_m - \mathbf{A}_m\|_{\text{op}} \geq \frac{1}{2}\right] \\ &\leq C(p, \lambda) \mu^p \left(\frac{m \log m}{cT}\right)^p \|\mathbf{A}_m^{-1}\|_{\text{op}}^{2p}. \end{aligned} \quad (40)$$

Next to control the first term on the right hand side of Equation (39), we apply Theorem C.1:

$$\begin{aligned} & \mathbb{P}\left[\left\{\|\hat{\mathbf{A}}_m^{-1} - \mathbf{A}_m^{-1}\|_{\text{op}} > \|\mathbf{A}_m^{-1}\|_{\text{op}}\right\} \cap \left\{\|\mathbf{A}_m^{-1}(\hat{\mathbf{A}}_m - \mathbf{A}_m)\|_{\text{op}} < \frac{1}{2}\right\}\right] \\ &\leq \mathbb{P}\left[\left\{\frac{\|\hat{\mathbf{A}}_m - \mathbf{A}_m\|_{\text{op}} \|\mathbf{A}_m^{-1}\|_{\text{op}}^2}{1 - \|\mathbf{A}_m^{-1}(\hat{\mathbf{A}}_m - \mathbf{A}_m)\|_{\text{op}}} > \|\mathbf{A}_m^{-1}\|_{\text{op}}\right\} \cap \left\{\|\mathbf{A}_m^{-1}(\hat{\mathbf{A}}_m - \mathbf{A}_m)\|_{\text{op}} < \frac{1}{2}\right\}\right] \\ &\leq \mathbb{P}\left[\|\hat{\mathbf{A}}_m - \mathbf{A}_m\|_{\text{op}} > \frac{1}{2}\|\mathbf{A}_m^{-1}\|_{\text{op}}^{-1}\right]. \end{aligned} \quad (41)$$

We apply Markov inequality again, along with Proposition 3.5:

$$\begin{aligned} & \mathbb{P}\left[\left\{\|\hat{\mathbf{A}}_m^{-1} - \mathbf{A}_m^{-1}\|_{\text{op}} > \|\mathbf{A}_m^{-1}\|_{\text{op}}\right\} \cap \left\{\|\mathbf{A}_m^{-1}(\hat{\mathbf{A}}_m - \mathbf{A}_m)\|_{\text{op}} < \frac{1}{2}\right\}\right] \\ &\leq C(p, \lambda) \mu^p \left(\frac{m \log m}{cT}\right)^p \|\mathbf{A}_m^{-1}\|_{\text{op}}^{2p}. \end{aligned}$$

So starting from Equation (39) and gathering Equations (40) and (41) gives:

$$\mathbb{P}\left[\|\hat{\mathbf{A}}_m^{-1}\|_{\text{op}} > \sqrt{\frac{cT}{m \log m}}\right] \leq C(p, \lambda) \mu^p \left(\frac{m \log m}{cT}\right)^p \|\mathbf{A}_m^{-1}\|_{\text{op}}^{2p}, \quad (42)$$

with $C(p, \lambda) = \mathcal{O}(\lambda^p)$.

Finally gathering Equations (38) with (42), we get that

$$\mathbb{E} \left[\|\mathbf{A}_m^{-1} - \tilde{\mathbf{A}}_{m,2}^{-1}\|_{\text{op}}^{2p} \right] \leq C(p, \lambda) (\mu^p + \mu^{2p}) \left(\|\mathbf{A}_m^{-1}\|_{\text{op}}^4 \frac{m \log m}{cT} \right)^p.$$

with $C(p, \lambda) = \mathcal{O}(\lambda^p \vee \lambda^{2p})$. \square

The next proposition is a variant of the last one. It gives a better bound than applying directly Proposition 7.10 to $\mathbb{E} \left[\|\tilde{\mathbf{A}}_{m,2}^{-1} - \mathbf{A}_m^{-1}\|_{\text{op}}^2 \|\mathbf{c}_m\|_{\ell^2}^2 \right]$.

Proposition 7.11. *If $m \log m \leq cT$, then it holds:*

$$\mathbb{E} \left[\|\tilde{\mathbf{A}}_{m,2}^{-1} - \mathbf{A}_m^{-1}\|_{\ell^2}^2 \|\mathbf{c}_m\|_{\ell^2}^2 \right] \leq C(\lambda) \|\phi_m\|_{L^2}^2 \left(\mu \wedge \left\{ (\mu + \mu^2) \|\mathbf{A}_m^{-1}\|_{\text{op}}^2 \frac{m \log m}{cT} \right\} \right),$$

with $C(\lambda) = \mathcal{O}(\lambda \vee \lambda^2)$.

Proof. The proof follows the lines of the proof of Proposition 7.10, but starting from the following decomposition:

$$\begin{aligned} \mathbb{E} \left[\|\mathbf{A}_m^{-1} - \tilde{\mathbf{A}}_{m,2}^{-1}\|_{\ell^2}^2 \|\mathbf{c}_m\|_{\ell^2}^2 \right] &= \|\mathbf{A}_m^{-1} \mathbf{c}_m\|_{\ell^2}^2 \mathbb{P}[\Delta_m^c] + \mathbb{E} \left[\|\hat{\mathbf{A}}_m^{-1} (\mathbf{A}_m - \hat{\mathbf{A}}_m) \mathbf{A}_m^{-1} \mathbf{c}_m\|_{\ell^2}^2 \mathbf{1}_{\Delta_m} \right] \\ &= \|\mathbf{a}_m\|_{\ell^2}^2 \mathbb{P}[\Delta_m^c] + \mathbb{E} \left[\|\hat{\mathbf{A}}_m^{-1} (\mathbf{A}_m - \hat{\mathbf{A}}_m) \mathbf{A}_m^{-1} \mathbf{c}_m\|_{\ell^2}^2 \mathbf{1}_{\Delta_m} \right], \end{aligned}$$

It yields the following upper bound

$$\mathbb{E} \left[\|\mathbf{A}_m^{-1} - \tilde{\mathbf{A}}_{m,2}^{-1}\|_{\ell^2}^2 \|\mathbf{c}_m\|_{\ell^2}^2 \right] \leq \|\mathbf{a}_m\|_{\ell^2}^2 \mathbb{P}[\Delta_m^c] + \|\mathbf{a}_m\|_{\ell^2}^2 \mathbb{E} \left[\|\hat{\mathbf{A}}_m^{-1}\|_{\text{op}}^2 \|\mathbf{A}_m - \hat{\mathbf{A}}_m\|_{\text{op}}^2 \mathbf{1}_{\Delta_m} \right].$$

Following the proof of Proposition 7.10, we get:

$$\mathbb{E} \left[\|\mathbf{A}_m^{-1} - \tilde{\mathbf{A}}_{m,2}^{-1}\|_{\ell^2}^2 \|\mathbf{c}_m\|_{\ell^2}^2 \right] \leq C(\lambda) \|\mathbf{a}_m\|_{L^2}^2 \left(\mu \wedge \left\{ (\mu + \mu^2) \|\mathbf{A}_m^{-1}\|_{\text{op}}^2 \frac{m \log m}{cT} \right\} \right),$$

with $C(\lambda) = \mathcal{O}(\lambda \vee \lambda^2)$. \square

Now we can prove Theorem 3.6.

Proof of Theorem 3.6. By Pythagoras Theorem, $\|\phi - \hat{\phi}_m^{\text{Lag}_2}\|_{L^2}^2 = \|\phi - \phi_m\|_{L^2}^2 + \|\phi_m - \hat{\phi}_m^{\text{Lag}_2}\|_{L^2}^2$. In the proof of Theorem 2.5, we saw that:

$$\mathbb{E} \|\hat{\mathbf{c}}_m - \mathbf{c}_m\|_{\ell^2}^2 = \mathbb{E} \|\hat{h}_m - h_m\|_{L^2}^2 \leq \frac{\lambda}{c^2 T} \mathbb{E}[W(X)].$$

We decompose the variance term in three terms:

$$\begin{aligned} \mathbb{E} \|\phi_m - \hat{\phi}_m^{\text{Lag}_2}\|_{L^2}^2 &= \mathbb{E} \|\hat{\mathbf{a}}_m^{\text{Lag}_2} - \mathbf{a}_m\|_{\ell^2}^2 = \mathbb{E} \|\tilde{\mathbf{A}}_{m,2}^{-1} \hat{\mathbf{c}}_m - \mathbf{A}_m^{-1} \mathbf{c}_m\|_{\ell^2}^2 \\ &\leq 3 \mathbb{E} \|\tilde{\mathbf{A}}_{m,2}^{-1} - \mathbf{A}_m^{-1}\|_{\ell^2}^2 \|\mathbf{c}_m\|_{\ell^2}^2 + 3 \mathbb{E} \|(\tilde{\mathbf{A}}_{m,2}^{-1} - \mathbf{A}_m^{-1})(\mathbf{c}_m - \hat{\mathbf{c}}_m)\|_{\ell^2}^2 + 3 \mathbb{E} \|\mathbf{A}_m^{-1}(\mathbf{c}_m - \hat{\mathbf{c}}_m)\|_{\ell^2}^2. \end{aligned}$$

For the first term, we apply Proposition 7.10. For the second term, we use the fact that $\tilde{\mathbf{A}}_{m,2}^{-1}$ and $\hat{\mathbf{c}}_m$ are independent, and we apply Proposition 7.11:

$$\mathbb{E} \|(\tilde{\mathbf{A}}_{m,2}^{-1} - \mathbf{A}_m^{-1})(\mathbf{c}_m - \hat{\mathbf{c}}_m)\|_{\ell^2}^2 \leq \mathbb{E} \|\tilde{\mathbf{A}}_{m,2}^{-1} - \mathbf{A}_m^{-1}\|_{\text{op}}^2 \times \mathbb{E} \|\mathbf{c}_m - \hat{\mathbf{c}}_m\|_{\ell^2}^2 = \mathcal{O} \left(\frac{1}{T^2} \right).$$

For the third term:

$$\mathbb{E} \|\mathbf{A}_m^{-1}(\mathbf{c}_m - \hat{\mathbf{c}}_m)\|_{\ell^2}^2 \leq \|\mathbf{A}_m^{-1}\|_{\text{op}}^2 \mathbb{E} \|\mathbf{c}_m - \hat{\mathbf{c}}_m\|_{\ell^2}^2 \leq \|\mathbf{A}_m^{-1}\|_{\text{op}}^2 \frac{\lambda}{c^2 T} \mathbb{E}[W(X)].$$

We apply Lemma 3.2 and we obtain the following bound, with $C(\lambda) = \mathcal{O}(\lambda \vee \lambda^2)$:

$$\begin{aligned} \mathbb{E} \|\phi_m - \hat{\phi}_m^{\text{Lag}_2}\|_{L^2}^2 &\leq 3 \|\mathbf{A}_m^{-1}\|_{\text{op}}^2 \frac{C(\lambda)}{cT} \left(\|\phi_m\|_{L^2}^2 (\mu + \mu^2) m \log(m) + \frac{\mathbb{E}[W(X)]}{c} \right) + \mathcal{O} \left(\frac{1}{T^2} \right) \\ &\leq 12 \frac{C(\lambda)}{cT(1-\theta)^2} \left(\|\phi_m\|_{L^2}^2 (\mu + \mu^2) m \log(m) + \frac{\mathbb{E}[W(X)]}{c} \right) + \mathcal{O} \left(\frac{1}{T^2} \right). \quad \square \end{aligned}$$

7.3 Proofs of Section 4

Proof of Lemma 4.6. It follows from direct calculation using (6):

$$\begin{aligned}
\langle F, \psi_k \rangle &= \int_0^{+\infty} C \exp(-\gamma x) \psi_k(x) dx \\
&= C\sqrt{2} \sum_{j=0}^k \binom{k}{j} \frac{(-2)^j}{j!} \int_0^{+\infty} x^j e^{-(1+\gamma)x} dx \\
&= C\sqrt{2} \sum_{j=0}^k \binom{k}{j} \frac{(-2)^j}{(1+\gamma)^{j+1}} \\
&= \frac{C\sqrt{2}}{\gamma+1} \left(1 - \frac{2}{\gamma+1}\right)^k = \frac{C\sqrt{2}}{\gamma+1} \left(\frac{\gamma-1}{\gamma+1}\right)^k.
\end{aligned}$$

Since γ is positive, we have $\left|\frac{\gamma-1}{\gamma+1}\right| < 1$ and we can compute the geometric series:

$$\sum_{k=m}^{+\infty} \langle F, \psi_k \rangle^2 = \frac{2C}{(\gamma+1)^2} \frac{\left(\frac{\gamma-1}{\gamma+1}\right)^{2m}}{1 - \left(\frac{\gamma-1}{\gamma+1}\right)^2} = \frac{C^2}{2\gamma} \left(\frac{\gamma-1}{\gamma+1}\right)^{2m}. \quad \square$$

Proof of Proposition 4.7. For the ruin probability and the Laplace transform of the ruin time, we start from (19) and we apply Lemma 4.6. For the expected jump size causing the ruin, we also start from (19) and we write $\phi = F_1 + F_2$ with:

$$F_1(u) := \mu(1+2\theta)e^{-\frac{1-\theta}{\mu}u} \mathbf{1}_{u>0}, \quad F_2(u) := \mu e^{-u/\mu} \mathbf{1}_{u>0}.$$

Hence, $\|\phi - \phi_m\|_{L^2}^2 \leq 2 \sum_{k=m}^{+\infty} \langle F_1, \psi_k \rangle^2 + 2 \sum_{k=m}^{+\infty} \langle F_2, \psi_k \rangle^2$. We apply Lemma 4.6:

$$\begin{aligned}
\|\phi - \phi_m\|_{L^2}^2 &\leq \frac{\mu^2(1+2\theta)^2}{\frac{1-\theta}{\mu}} \left(\frac{\frac{1-\theta}{\mu} - 1}{\frac{1-\theta}{\mu} + 1}\right)^{2m} + \mu^3 \left(\frac{\frac{1}{\mu} - 1}{\frac{1}{\mu} + 1}\right)^{2m} \\
&= \frac{\mu^3(1+2\theta)^2}{1-\theta} \left(\frac{1-\theta-\mu}{1-\theta+\mu}\right)^{2m} + \mu^3 \left(\frac{1-\mu}{1+\mu}\right)^{2m} \\
&\leq \frac{\mu^3(1+2\theta)^2}{1-\theta} \left(\left|\frac{1-\theta-\mu}{1-\theta+\mu}\right| \vee \left|\frac{1-\mu}{1+\mu}\right|\right)^{2m}. \quad \square
\end{aligned}$$

Acknowledgments

I want to thank Fabienne Comte and Céline Duval for their helpful advice and their support of my work. I also want to thank fedja for their help with Lemma 3.4 regarding the primitives of the Laguerre functions, and Emmanuel Rio for his hints for Lemma 7.1.

Funding

This work was supported by a grant from Région Île-de-France.

References

- M. Abramowitz and I. A. Stegun. *Handbook of Mathematical Functions: with Formulas, Graphs, and Mathematical Tables*. Number 55 in Applied Mathematics Series. National Bureau of Standards, New York, NY, 10th edition, 1972.
- S. Arlot and P. Massart. Data-driven Calibration of Penalties for Least-Squares Regression. *Journal of Machine Learning Research*, 10(10):245–279, 2009.

- S. Asmussen and H. Albrecher. *Ruin probabilities*, volume 14 of *Advanced series on statistical science and applied probability*. World Scientific, Singapore ; New Jersey, 2nd edition, 2010.
- B. Bongioanni and J. L. Torrea. What is a Sobolev space for the Laguerre function systems? *Studia Mathematica*, 192(2):147–172, 2009.
- A. Böttcher and S. M. Grudsky. *Toeplitz Matrices, Asymptotic Linear Algebra, and Functional Analysis*. Birkhäuser Basel, Basel, 2000.
- R. Y. Chen, A. Gittens, and J. A. Tropp. The masked sample covariance estimator: an analysis using matrix concentration inequalities. *Information and Inference*, 1(1):2–20, Dec. 2012.
- F. Comte and V. Genon-Catalot. Adaptive Laguerre density estimation for mixed Poisson models. *Electronic Journal of Statistics*, 9(1):1113–1149, 2015.
- F. Comte and G. Mabon. Laguerre deconvolution with unknown matrix operator. *Mathematical Methods of Statistics*, 26(4):237–266, Oct. 2017.
- F. Comte, C.-A. Cuenod, M. Pensky, and Y. Rozenholc. 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, Jan. 2017.
- K. Croux and N. Veraverbeke. Nonparametric estimators for the probability of ruin. *Insurance: Mathematics and Economics*, 9(2-3):127–130, Sept. 1990.
- F. Dussap. Anisotropic multivariate deconvolution using projection on the Laguerre basis. *Journal of Statistical Planning and Inference*, 215:23–46, 2021.
- fedja. Proving that the primitives of the laguerre functions are uniformly bounded. MathOverflow, Feb. 2021. URL <https://mathoverflow.net/q/383022>.
- E. W. Frees. Nonparametric Estimation of the Probability of Ruin. *ASTIN Bulletin*, 16(S1):S81–S90, Apr. 1986.
- H. U. Gerber and E. S. Shiu. On the Time Value of Ruin. *North American Actuarial Journal*, 2(1):48–72, Jan. 1998.
- G. Mabon. Adaptive Deconvolution on the Non-negative Real Line: Adaptive deconvolution on \mathbb{R}_+ . *Scandinavian Journal of Statistics*, 44(3):707–740, Sept. 2017.
- E. Masiello. On semiparametric estimation of ruin probabilities in the classical risk model. *Scandinavian Actuarial Journal*, 2014(4):283–308, June 2014.
- P. Massart. The Tight Constant in the Dvoretzky-Kiefer-Wolfowitz Inequality. *The Annals of Probability*, 18(3):1269–1283, July 1990.
- R. Mnatsakanov, L. L. Ruymgaart, and F. H. Ruymgaart. Nonparametric estimation of ruin probabilities given a random sample of claims. *Mathematical Methods of Statistics*, 17(1):35–43, Mar. 2008.
- S. M. Pitts. Nonparametric estimation of compound distributions with applications in insurance. *Annals of the Institute of Statistical Mathematics*, 46(3):537–555, Sept. 1994.
- K. Politis. Semiparametric Estimation for Non-Ruin Probabilities. *Scandinavian Actuarial Journal*, 2003(1):75–96, Jan. 2003.
- Y. Shimizu. Estimation of the expected discounted penalty function for Lévy insurance risks. *Mathematical Methods of Statistics*, 20(2):125–149, June 2011.
- Y. Shimizu. Non-parametric estimation of the Gerber–Shiu function for the Wiener–Poisson risk model. *Scandinavian Actuarial Journal*, 2012(1):56–69, Mar. 2012.
- Y. Shimizu and Z. Zhang. Estimating Gerber–Shiu functions from discretely observed Lévy driven surplus. *Insurance: Mathematics and Economics*, 74:84–98, May 2017.

- W. Su and W. Yu. Asymptotically Normal Estimators of the Gerber-Shiu Function in Classical Insurance Risk Model. *Mathematics*, 8(10):1638, Sept. 2020.
- W. Su, B. Shi, and Y. Wang. Estimating the Gerber-Shiu function under a risk model with stochastic income by Laguerre series expansion. *Communications in Statistics - Theory and Methods*, pages 1–23, May 2019a.
- W. Su, Y. Yong, and Z. Zhang. Estimating the Gerber-Shiu function in the perturbed compound Poisson model by Laguerre series expansion. *Journal of Mathematical Analysis and Applications*, 469(2):705–729, Jan. 2019b.
- Z. Zhang and W. Su. A new efficient method for estimating the Gerber-Shiu function in the classical risk model. *Scandinavian Actuarial Journal*, 2018(5):426–449, May 2018.
- Z. Zhang and W. Su. Estimating the Gerber-Shiu function in a Lévy risk model by Laguerre series expansion. *Journal of Computational and Applied Mathematics*, 346:133–149, Jan. 2019.

A Model reduction procedure

We propose a *model reduction procedure* to choose the dimensions m_2 and m_3 , defined by (14). We explain the method for the choice of m_2 in the case $\delta = 0$.

Let us assume we have estimated the M first coefficients of g , for a large M . By Remark 2.6, we know that the best estimator is \hat{g}_M . Our goal is to choose \hat{m}_2 smaller than M that achieves a similar MISE. This provides a parsimonious version of the estimator without degrading its MISE. By Theorem 2.5, the MISE of \hat{g}_m is given by:

$$\mathbb{E}\|g - \hat{g}_m\|_{L^2}^2 \leq \|g - g_m\|_{L^2}^2 + \frac{\lambda}{c^2 T} \mathbb{E}[X].$$

Ideally, we would like to choose the first m such that the bias term $\|g - g_m\|_{L^2}^2$ is smaller than the variance term $\frac{\lambda}{c^2 T} \mathbb{E}[X]$. Since these terms are unknown, we estimate them by $\sum_{k=m}^{M-1} \hat{b}_k^2$ and $\frac{1}{(cT)^2} \sum_{i=1}^{N_T} X_i$ respectively. We choose \hat{m}_2 as:

$$\hat{m}_2 = \min \left\{ 1 \leq m \leq M-1 \left| \sum_{k=m}^{M-1} \hat{b}_k^2 \leq \frac{\kappa_2}{(cT)^2} \sum_{i=1}^{N_T} X_i \right. \right\}, \quad (43)$$

with κ_2 an adjustment constant. The next proposition shows that the MISE of $\hat{g}_{\hat{m}_2}$ does not exceed the MISE of \hat{g}_M by more than $\kappa_2 \times (\text{variance term})$.

Proposition A.1. *Let $\kappa_2 > 0$, if \hat{m}_2 is chosen as (43) then the MISE of $\hat{g}_{\hat{m}_2}$ is:*

$$\mathbb{E}\|g - \hat{g}_{\hat{m}_2}\|_{L^2}^2 \leq \|g - g_M\|_{L^2}^2 + (1 + \kappa_2) \frac{\lambda}{c^2 T} \mathbb{E}[X].$$

Proof. By Pythagoras Theorem:

$$\begin{aligned} \|g - \hat{g}_{\hat{m}_2}\|_{L^2}^2 &= \|g - \hat{g}_M\|_{L^2}^2 + \|\hat{g}_M - \hat{g}_{\hat{m}_2}\|_{L^2}^2 \\ &= \|g - \hat{g}_M\|_{L^2}^2 + \sum_{k=\hat{m}_2}^{M-1} \hat{b}_k^2 \\ &\leq \|g - \hat{g}_M\|_{L^2}^2 + \frac{\kappa_2}{(cT)^2} \sum_{i=1}^{N_T} X_i. \end{aligned}$$

We take the expectation, and we apply Theorem 2.5:

$$\mathbb{E}\|g - \hat{g}_{\hat{m}_2}\|_{L^2}^2 \leq \|g - g_M\|_{L^2}^2 + (1 + \kappa_2) \frac{\lambda}{c^2 T} \mathbb{E}[X]. \quad \square$$

	With Model Reduction	Without Model Reduction
Ruin Probability	1.06 [0.86, 1.26] $\hat{m}_2 = 4.2, \hat{m}_3 = 4.1$	1.06 [0.86, 1.25] $m_2 = m_3 = N_T$
Jump size causing the ruin	40.1 [33.7, 46.9] $\hat{m}_2 = 4.1, \hat{m}_3 = 4.3$	41.6 [34.7, 48.4] $m_2 = m_3 = N_T$
Laplace transform, $\delta = 0.1$	0.092 [0.075, 0.110] $\hat{m}_2 = 3.9, \hat{m}_3 = 4.0$	0.098 [0.079, 0.117] $m_2 = m_3 = N_T$

Table 4: Comparison between the MISE of the Laguerre–Fourier estimator with and without model reduction. In each case, we chose the following parameters: $X \sim \text{Exp}(1/2)$, $\lambda = 1.25$, $c = 3$, $T = 80$. With this set of parameters, $\mathbb{E}[N_T] = 100$. Each cell displays an estimation of the MISE over 200 samples with a 95% confidence interval, and the mean selected models \hat{m}_2 and \hat{m}_3 . In every case, m_1 is equal to N_T .

The same goes for \hat{m}_3 : we estimate the bias term by $\sum_{k=m}^{M-1} \hat{c}_k^2$ and the variance term by $\frac{1}{(cT)^2} \sum_{i=1}^{N_T} W(X_i)$; we choose \hat{m}_3 as:

$$\hat{m}_3 = \min \left\{ 1 \leq m \leq M-1 \left| \sum_{k=m}^{M-1} \hat{c}_k^2 \leq \frac{\kappa_3}{(cT)^2} \sum_{i=1}^{N_T} W(X_i) \right. \right\}.$$

By the same arguments, the MISE of \hat{h}_{m_3} is given by:

$$\mathbb{E} \|h - \hat{h}_{\hat{m}_3}\|_{L^2}^2 \leq \|h - h_M\|_{L^2}^2 + (1 + \kappa_3) \frac{\lambda}{c^2 T} \mathbb{E}[W(X)].$$

In the case $\delta > 0$, we choose the same \hat{m}_2 and \hat{m}_3 as in the case $\delta = 0$. By the same arguments, we obtain:

$$\begin{aligned} \mathbb{E} \|g - \hat{g}_{\hat{m}_2}\|_{L^2}^2 &\leq \|g - g_M\|_{L^2}^2 + \frac{C(\lambda)}{c^2 T} \left(\mathbb{E}[X] + \frac{\mathbb{E}[X^2]^{\frac{1}{2}}}{(1-\theta)^2 \delta^2} \right) + \kappa_2 \frac{\lambda}{c^2 T} \mathbb{E}[X] \\ &\leq \|g - g_M\|_{L^2}^2 + (1 + \kappa_2) \frac{C(\lambda)}{c^2 T} \left(\mathbb{E}[X] + \frac{\mathbb{E}[X^2]^{\frac{1}{2}}}{(1-\theta)^2 \delta^2} \right). \\ \mathbb{E} \|h - \hat{h}_{\hat{m}_3}\|_{L^2}^2 &\leq \|h - h_M\|_{L^2}^2 + \frac{C(\lambda)}{c^2 T} \left(\mathbb{E}[W(X)] + \frac{\mathbb{E}[W(X)^2]^{\frac{1}{2}}}{(1-\theta)^2 \delta^2} \right) + \kappa_3 \frac{\lambda}{c^2 T} \mathbb{E}[W(X)] \\ &\leq \|h - h_M\|_{L^2}^2 + (1 + \kappa_3) \frac{C(\lambda)}{c^2 T} \left(\mathbb{E}[W(X)] + \frac{\mathbb{E}[W(X)^2]^{\frac{1}{2}}}{(1-\theta)^2 \delta^2} \right). \end{aligned}$$

Numerically, we compared the MISE's of the Laguerre–Fourier estimator with and without the model reduction procedure for \hat{m}_2 and \hat{m}_3 , with the choice $\kappa_2 = \kappa_3 = 0.3$. We show the results in Table 4. We see that the model reduction procedure does not affect the MISE of the estimator and we emphasize that the selected dimensions are far lower than the maximum dimension (\hat{m} 's are less than 10 whereas the maximum dimension is 100).

B Uniform bound on the primitives of the Laguerre functions

In this section, we prove that the primitives of the Laguerre function are uniformly bounded. The sketch of the proof comes from [fedja \[2021\]](#).

Proof of Lemma 3.4. Let $u_k(x) := L_k(x)e^{-x/2}$. We first notice that the complete integral of u_k is uniformly bounded:

$$\begin{aligned} \int_0^{+\infty} u_k(x) dx &= 2 \int_0^{+\infty} L_k(2x)e^{-x} dx \\ &= 2 \sum_{j=0}^k \binom{k}{j} \frac{(-2)^j}{j!} \int_0^{+\infty} x^j e^{-x} dx \\ &= 2 \sum_{j=0}^k \binom{k}{j} (-2)^j = 2(-1)^k. \end{aligned}$$

We will show that $|\int_0^x u_k| \leq C |\int_0^\infty u_k| = 2C$ for an absolute constant $C > 0$. The k -th Laguerre polynomial satisfies the ODE:

$$xL_k'' + (1-x)L_k' + kL_k = 0,$$

thus, the function u_k satisfies:

$$xu_k'' + u_k' + \left(k + \frac{1}{2} - \frac{1}{4}x\right)u_k = 0 \quad (44)$$

To kill the first derivative, we consider $v_k(x) := u_k(x^{3/2})x^{1/2}$. The functions u_k and v_k have the same partial integrals (up to the constant $2/3$):

$$\forall x \geq 0, \quad \int_0^x v_k(t) dt = \frac{2}{3} \int_0^{x^{2/3}} u_k(t) dt.$$

The first two derivatives of v_k are:

$$\begin{aligned} v_k'(x) &= \frac{3}{2}xu_k'(x^{3/2}) + \frac{1}{2}x^{-1/2}u(x^{3/2}) \\ v_k''(x) &= \frac{9}{4}x^{3/2}u''(x^{3/2}) + \frac{9}{4}u'(x^{3/2}) - \frac{1}{4}x^{-3/2}u(x^{3/2}) \end{aligned}$$

so using the ODE (44) for u yields the following ODE for v :

$$v_k'' + \Phi_k v_k = 0$$

where Φ_k is given by:

$$\Phi_k(x) := \frac{9}{4} \left(\frac{k + \frac{1}{2}}{x^{1/2}} - \frac{1}{4}x \right) + \frac{1}{4x^2}.$$

The important properties of this function are that it is convex and decreasing.

Since the Laguerre polynomials have simple zeros, the function v_k has $(k+1)$ simple zeros (the zeros of the k -th Laguerre polynomial, and 0) so the integral of v_k can be decomposed as an alternating sum $\int_0^\infty v_k =: I_k = A_0 - A_1 + A_2 - \dots + (-1)^k A_k$, where the A_i 's are the unsigned areas of the excursions of v_k (see Figure 3). Based on the following lemma (proven later), we claim that $A_0 < A_1 < \dots < A_k$.

Lemma B.1. *Let F_1, F_2 be two \mathcal{C}^1 functions defined on an open interval I containing 0, and let y_1, y_2 be the solutions of the following ODE:*

$$\begin{cases} y_1'' + F_1(x)y_1 = 0 \\ y_2'' + F_2(x)y_2 = 0 \\ y_1(0) = y_2(0) = 0 \\ y_2'(0) \geq y_1'(0) > 0. \end{cases}$$

Let $M > 0$ such that y_1 and y_2 are positive on $J := (0, M) \subseteq I$. If $F_1 > F_2$ on J , then we have $y_1 \leq y_2$ on J .

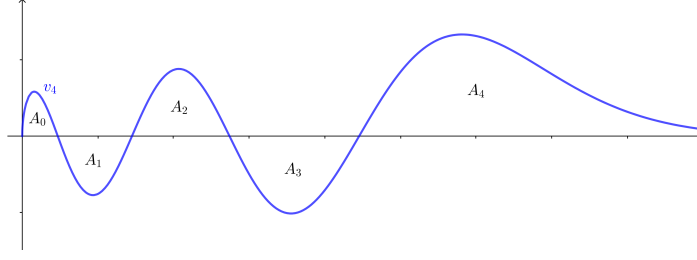


Figure 3: Graph of v_4 . We see that the area of the excursions is increasing.

Indeed, let $z > 0$ be a zero of v_k and assume w.l.o.g. that v_k is positive after z and negative before. Let $y_1(x) := -v_k(z - x)$ and $y_2(x) := v_k(z + x)$ (i.e. y_1 is the central inversion of v_k with respect to z). Thus, y_1 and y_2 satisfy the ODE:

$$\begin{cases} y_1'' + \Phi_k(z - x)y_1 = 0 \\ y_2'' + \Phi_k(z + x)y_2 = 0 \end{cases}$$

with the initial conditions $y_1(0) = y_2(0) = v_k(z) = 0$ and $y_1'(0) = y_2'(0) = v_k'(z) > 0$. Since Φ_k is decreasing, we have $\Phi_k(z - x) > \Phi_k(z + x)$, so Lemma B.1 yields that $y_1(x) \leq y_2(x)$ for $x > 0$ as long as $y_1(x)$ and $y_2(x)$ are positive. Hence, the area of the excursion preceding z is smaller than the area of the excursion following z .

Now let z be the last zero of v_k , and let us assume w.l.o.g. that v_k is positive after z (otherwise we consider $-v_k$, it satisfies the same ODE than v_k). In this case, $I_k = A_0 - A_1 + \dots - A_{k-1} + A_k$ with $A_0 < \dots < A_{k-1} < A_k$. Thus the maximum value of $|\int_0^x v_k|$ is attained either by taking the complete integral (it is the maximum of $\int_0^x v_k$), either by leaving out the last excursion (it is the minimum of $\int_0^x v_k$). We show that the second option is dominated by the first one: there exists an absolute constant $C > 0$ such that $A_k - I_k \leq CI_k$. To do that it suffices to show that $A_{k-1} \leq cA_k$ for absolute constant $c \in (0, 1)$, hence $A_k - I_k \leq \frac{c}{1-c}I_k$.

The strategy is to compare the function v_k to the Airy function of the first kind. This function is solution of the ODE:

$$\begin{cases} y'' - xy = 0 \\ \lim_{x \rightarrow +\infty} y(x) = 0. \end{cases} \quad (45)$$

Let z_ℓ^* and z_p^* be respectively the last and the penultimate zeros of Ai. We recall that z_ℓ^* is negative (all the zeros of Ai are negative), and that Ai is negative on (z_p^*, z_ℓ^*) and positive on $(z_\ell^*, +\infty)$.

Lemma B.2. *Let Ai be the Airy function of the first kind, and let z be the last zero of v_k . There exists a decreasing linear function Υ_k satisfying $\Upsilon_k(z) = \Phi_k(z)$, and there exists $a > 0$ and $b \in \mathbb{R}$, such that the function $\text{Ai}(ax + b)$ satisfies:*

$$\begin{cases} y'' + \Upsilon_k(x)y = 0 \\ y(z) = 0 \end{cases}$$

and such that it stays positive on $(z, +\infty)$ and tends to 0 at $+\infty$.

Let $w_k(x) := \varepsilon \text{Ai}(ax + b)$ where a and b are given by Lemma B.2, and $\varepsilon > 0$ is small enough such that $0 < w_k'(z) < v_k'(z)$. Let z_ℓ and z_p be respectively the last and the penultimate zeros of w_k . By definition of w_k , its last zero is the same than v_k 's, that is $z_\ell = z$. Moreover, the zeros of w_k and Ai are linked by the linear transformation $x \mapsto ax + b$: we have $z_\ell^* = az_\ell + b$ and $z_p^* = az_p + b$.

Let us consider:

$$\mathcal{W} := \det \begin{pmatrix} v_k & w_k \\ v_k' & w_k' \end{pmatrix} = v_k w_k' - v_k' w_k,$$

the Wronskian of v_k and w_k . Since \mathcal{W} vanishes at z and $+\infty$, we have:

$$0 = \int_z^{+\infty} \mathcal{W}'(x) dx = \int_z^{+\infty} (\Phi_k(x) - \Upsilon_k(x)) \underbrace{v_k(x)w_k(x)}_{>0} dx,$$

so the sign of $\Phi_k - \Upsilon_k$ must change on $(z, +\infty)$. Since $\Phi_k - \Upsilon_k$ is convex and $(\Phi_k - \Upsilon_k)(z) = 0$, this is possible only if it is negative and then positive. Hence, \mathcal{W} is first decreasing and then increasing on $(z, +\infty)$. Since it starts from zero at z and tends to zero at $+\infty$, we conclude that \mathcal{W} is negative on $(z, +\infty)$.

It follows that $w_k < v_k$ on $(z, +\infty)$. Indeed, by contradiction if $x_0 := \inf\{x \in (z, +\infty) \mid w_k(x) \geq v_k(x)\}$ existed and was finite, we would have $v_k(x_0) = w_k(x_0) > 0$ and $v'_k(x_0) \leq w'_k(x_0)$. Thus, we would have $\mathcal{W}(x_0) = v_k(x_0)w'_k(x_0) - v'_k(x_0)w_k(x_0) \geq 0$; contradiction. We conclude that the area of the last excursion of w_k is less than v_k 's: $\int_{z_\ell}^\infty w_k \leq A_k$.

We have seen that $\Phi_k - \Upsilon_k$ was negative then positive on $(z, +\infty)$. By convexity, it has to be positive on $(0, z)$, i.e. $\Phi_k > \Upsilon_k$ on the left of z . We apply Lemma B.1 to $-v_k(z-x)$ and $-w_k(z-x)$, and we conclude that the area of the penultimate excursion of v_k is less than w_k 's: $A_{k-1} \leq |\int_{z_p}^{z_\ell} w_k|$.

We conclude that A_{k-1}/A_k is bounded above by the ratio of the areas of the penultimate excursion to the last excursion of w_k . By a linear change of variable, this ratio is equal to the ratio of the areas of the penultimate excursion to the last excursion of the Airy function:

$$\frac{A_{k-1}}{A_k} \leq \frac{\left| \int_{z_p}^{z_\ell} w_k(x) dx \right|}{\int_{z_\ell}^{+\infty} w_k(x) dx} = \frac{\left| \int_{z_p^*}^{z_\ell^*} \text{Ai}(x) dx \right|}{\int_{z_\ell^*}^{+\infty} \text{Ai}(x) dx} =: c.$$

This is an absolute constant, we just need to prove it is smaller than 1 to end the proof. The function Ai satisfies an ODE of the form $y'' + F(x)y = 0$ with $F(x) = -x$ a decreasing function, thus by considering the functions $y_1(x) = -\text{Ai}(z_\ell^* - x)$ and $y_2 := \text{Ai}(z_\ell^* + x)$, we can apply again Lemma B.1 (as we did with v_k), to conclude that $y_1(x) \leq y_2(x)$ for $x > 0$ as long as $y_1(x)$ is positive (y_2 being positive for every $x > 0$). This proves that $|\int_{z_p^*}^{z_\ell^*} \text{Ai}| < \int_{z_\ell^*}^\infty \text{Ai}$, that is $c < 1$. \square

Proof of Lemma B.1. First, let us consider the case $y'_2(0) > y'_1(0)$. Let $\mathcal{W} := y_1 y'_2 - y'_1 y_2$ be the Wronskian of y_1 and y_2 . Then $\mathcal{W}' = (F_1 - F_2)y_1 y_2$ is positive on J and $\mathcal{W}(0) = 0$, so $\mathcal{W} > 0$ on J .

By contradiction, suppose there exists $x \in J$ such that $y_1(x) \geq y_2(x) > 0$ and consider $x_0 := \inf\{x \in J \mid y_1(x) \geq y_2(x) > 0\}$. Since $y'_2(0) > y'_1(0)$, we know that y_1 is below y_2 on a right neighborhood of 0, so $x_0 > 0$. By continuity, we have $y_1(x_0) = y_2(x_0) > 0$, so we must have $y'_1(x_0) \geq y'_2(x_0)$; otherwise, we would have $y_1(x) \geq y_2(x) > 0$ on a left neighborhood of x_0 , which is impossible if $x_0 > 0$. Thus $\mathcal{W}(x_0) = y_1(x_0)[y'_2(x_0) - y'_1(x_0)] \leq 0$ with $x_0 \in J$; contradiction.

Now consider the case $y'_2(0) = y'_1(0)$. For $\varepsilon > 0$ small enough, we consider y_ε the solution of:

$$\begin{cases} y''_\varepsilon + F_1(x)y_\varepsilon = 0 \\ y_\varepsilon(0) = 0 \\ y'_\varepsilon(0) = y'_2(0) - \varepsilon > 0. \end{cases}$$

Applying the first case, we have $y_\varepsilon \leq y_2$ while y_ε and y_2 are positive. Taking $\varepsilon \rightarrow 0$, we obtain $y_1 \leq y_2$. \square

Proof of Lemma B.2. The Airy function is solution of (45), so $\text{Ai}(ax+b)$ (with $a > 0$) satisfies:

$$\begin{cases} y'' - a^2(ax+b)y = 0 \\ \lim_{x \rightarrow +\infty} y(x) = 0. \end{cases}$$

We need to determine (a, b) such that the two following conditions hold:

1. Let $\Upsilon_k(x) := -a^2(ax+b)$, we need to choose a, b such that $\Upsilon_k(z) = \Phi_k(z)$.
2. Let z_ℓ^* be the last zero of Ai . We know that $z_\ell^* < 0$ and that Ai is positive on $(z_\ell^*, +\infty)$, thus we need to choose a, b such that $az + b = z_\ell^*$ so $\text{Ai}(ax+b)$ stays positive on $(z, +\infty)$ and vanishes at z .

Thus, $(a, b) \in \mathbb{R}_+^* \times \mathbb{R}$ must be solution of:

$$\begin{cases} -a^2(az+b) = \Phi_k(z) \\ az+b = z_\ell^* \end{cases} \iff \begin{cases} a^2 = -\frac{\Phi_k(z)}{z_\ell^*} \\ az+b = z_\ell^* \end{cases} \quad (46)$$

Since $z_\ell^* < 0$, this system has a solution iff $\Phi_k(z) > 0$. By contradiction, if we had $\Phi_k(z) \leq 0$, then Φ_k would be negative on $(z, +\infty)$. Since the function v_k is positive after z and satisfies $v_k''(x) = -\Phi_k(x)v_k(x)$, then v_k would be strictly convex on $(z, +\infty)$. But the function v_k starts from zero at z with a positive derivative, stays positive, and tends to 0 at $+\infty$, so it cannot be strictly convex on $(z, +\infty)$; contradiction. Thus, the system (46) has a solution. \square

C Miscellaneous results

Theorem C.1. *Let \mathbf{A}, \mathbf{B} be $(m \times m)$ matrices. If \mathbf{A} is invertible and $\|\mathbf{A}^{-1}\mathbf{B}\|_{\text{op}} < 1$, then $\mathbf{A} + \mathbf{B}$ is invertible and it holds:*

$$\|(\mathbf{A} + \mathbf{B})^{-1} - \mathbf{A}^{-1}\|_{\text{op}} \leq \frac{\|\mathbf{A}^{-1}\|_{\text{op}}^2 \|\mathbf{B}\|_{\text{op}}}{1 - \|\mathbf{A}^{-1}\mathbf{B}\|_{\text{op}}}.$$

Proof. Since $\|\mathbf{A}^{-1}\mathbf{B}\|_{\text{op}} < 1$, its Neumann series is normally convergent and we have:

$$\sum_{k=0}^{+\infty} (-1)^k (\mathbf{A}^{-1}\mathbf{B})^k = (\mathbf{I}_m + \mathbf{A}^{-1}\mathbf{B})^{-1}.$$

Hence:

$$\begin{aligned} \|(\mathbf{A} + \mathbf{B})^{-1} - \mathbf{A}^{-1}\|_{\text{op}} &\leq \|\mathbf{A}^{-1}\|_{\text{op}} \|(\mathbf{I}_m + \mathbf{A}^{-1}\mathbf{B})^{-1} - \mathbf{I}_m\|_{\text{op}} \\ &\leq \|\mathbf{A}^{-1}\|_{\text{op}} \sum_{k=1}^{+\infty} \|\mathbf{A}^{-1}\mathbf{B}\|_{\text{op}}^k \\ &= \|\mathbf{A}^{-1}\|_{\text{op}} \frac{\|\mathbf{A}^{-1}\mathbf{B}\|_{\text{op}}}{1 - \|\mathbf{A}^{-1}\mathbf{B}\|_{\text{op}}} \leq \frac{\|\mathbf{A}^{-1}\|_{\text{op}}^2 \|\mathbf{B}\|_{\text{op}}}{1 - \|\mathbf{A}^{-1}\mathbf{B}\|_{\text{op}}}. \end{aligned} \quad \square$$

Theorem C.2 (Theorem A.1 in [Chen et al. \[2012\]](#)). *Suppose that $q \geq 2$ and fix $r \geq \max(q, 2 \log p)$. Consider a finite sequence $\{\mathbf{Y}_i\}$ of independent, symmetric, random, self-adjoint matrices with dimension $p \times p$. Then:*

$$\left[\mathbb{E} \lambda_{\max} \left(\sum_i \mathbf{Y}_i \right)^q \right]^{1/q} \leq \sqrt{er \lambda_{\max} \left(\sum_i \mathbb{E} \mathbf{Y}_i^2 \right)} + 2er \left[\mathbb{E} \max_i \lambda_{\max}^q(\mathbf{Y}_i) \right]^{1/q}.$$