ADAPTIVE DENSITY ESTIMATION IN THE PILE-UP MODEL INVOLVING MEASUREMENT ERRORS

FABIENNE COMTE[1], TABEA REBAFKA[2]

ABSTRACT. Motivated by fluorescence lifetime measurements, this paper considers the problem of nonparametric density estimation in the pile-up model, where observations suffer also from measurement errors. In the pile-up model, an observation is defined as the minimum of a random number of i.i.d. variables following the target distribution. Adaptive nonparametric estimators are proposed for this pile-up model with measurement errors. Furthermore, oracle type risk bounds for the mean integrated squared error (MISE) are provided. Finally, the estimation method is assessed by a simulation study and the application to real fluorescence lifetime data.

Keywords. Adaptive nonparametric estimation. Biased data. Deconvolution. Fluorescence lifetimes. Projection estimator.
AMS Subject Classification. 62G07-62N01

1. INTRODUCTION

This paper is concerned with nonparametric density estimation in a specific inverse problem. Observations are not directly available from the target distribution, but suffer from both measurement errors and the so-called pile-up effect. The pile-up effect refers to some right-censoring, since an observation is defined as the minimum of a random number of i.i.d. variables from the target distribution. The pile-up distribution is thus the result of a nonlinear distortion of the target distribution. In our setting we also take into account measurement errors, that is the pile-up effect applies to the convolution of the target density and a known error distribution. The aim is to estimate the target density in spite of the pile-up effect and additive noise.

The pile-up model is encountered in time-resolved fluorescence when lifetime measurements are obtained by the technique called Time-Correlated Single-Photon Counting (TC-SPC) (O'Connor and Phillips, 1984). The fluorescence lifetime is the duration that a molecule stays in the excited state before emitting a fluorescence photon (Lakowicz, 1999; Valeur, 2002). The distribution of the fluorescence lifetimes associated with a sample of molecules provides precious information on the underlying molecular processes. Lifetimes are used in various applications as e.g. to determine the speed of rotating molecules or to measure molecular distances. This means that the knowledge of the lifetime distribution is required to obtain information on physical and chemical processes.

In the TCSPC technique, a short laser pulse excites a random number of molecules, but for technical reasons, only the arrival time of the very first fluorescence photon striking the detector can be measured, while the arrival times of the other photons are unobservable. The arrival time of a photon is the sum of the fluorescence lifetime and some noise, which is

some random time due to the measuring instrument as e.g. the time of flight of the photon in the photon-multiplier tube. Hence, TCSPC observations can be described by a pile-up model with measurement errors. The goal is to recover the distribution of the lifetimes of *all* fluorescence photons from the piled-up observations.

Until recently TCSPC was operated in a mode where the pile-up effect is negligible. However, a shortcoming of this mode is that the acquisition time is very long. Recent studies have made clear that from an information viewpoint it is a better strategy to operate TCSPC in a mode with considerable pile-up effect (Rebafka et al., 2010, 2011). Consequently, an estimation procedure is required that takes the pile-up effect into account. The concern of this paper is to provide such a nonparametric estimator of the target density and furthermore to include measurement errors in the model in order to deal with real fluorescence data. Therefore, we develop adequate deconvolution strategies for the correction in the pile-up model and test those methods on simulated data as well as on real fluorescence data.

It is noteworthy that the pile-up model is connected to survival analysis, since it can be considered as a special case of the nonlinear transformation model (Tsodikov, 2003). Indeed, it is straightforward to extend the methods proposed in this paper to this more general case. Moreover, the model can also be viewed as a biased data problem with known bias, see Brunel et al. (2009). Nonetheless, the consideration of additional measurement errors is new and fruitful. Since there are now two sources of mis-measurement, we have to face real technical difficulties to preserve standard deconvolution rates: either a loss is admitted, or the sample has to be split in two independent parts. Finally, we show that deconvolution methods can be used to complete the study in the spirit of Fan (1991), Pensky and Vidakovic (1999), Diggle and Hall (1993) or Comte et al. (2006). These techniques are of unusual use in both survival analysis and pile-up model studies. Numerical results confirm the adequacy of these methods in practice.

In Section 2, the model is described, together with the main assumptions. Then the nonparametric estimation strategy to recover the target density in the pile-up model with measurement errors is presented . In Section 3, the properties of the estimator are described, which are mainly risk bounds for the estimator. The rates obtained in this framework depend on the smoothness of the error density and on the choice of a cut-off parameter. Furthermore, a cut-off selection strategy is proposed in Section 4 to achieve an adequate bias-variance trade-off. In Section 5 the performance of the methods is assessed via simulations and by an application on a dataset of fluorescence lifetime measurements. All proofs are relegated to Section 6.

2. Pile-up Model with measurement errors and assumptions

2.1. **Notations.** In the following, for u and v two functions, we denote by $u \circ v$ the function $x \mapsto u \circ v(x) := u(v(x))$. If u is bijective, we denote by u^{-1} the inverse of the function u, that is the function such that $(u^{-1} \circ u)(x) = (u \circ u^{-1})(x) = x$ for all x. We also denote by $\dot{u}$ the derivative of u and by $\ddot{u}$ the second-order derivative of u, when they exist.

If u and v are real valued and square-integrable, we define the convolution product $u \star v$ of u and v by $(u \star v)(x) = \int u(x-t)v(t) \mathrm{d}t$ and the scalar product $\langle u, v \rangle$ by $\langle u, v \rangle = \int u(t)v(t) \mathrm{d}t$. If the functions are complex valued, the conjugate of v is used instead of simply v. If u is integrable, we define the Fourier transform of u by $u^*(t) = \int \mathrm{e}^{-itx} u(x) \mathrm{d}x$. We recall that for u and v integrable and square-integrable functions, $(u \star v)^* = u^* v^*$. Moreover, Parseval's

formula gives the useful relations

$$(u^*)^*(x) = (2\pi)u(-x) \quad \text{and} \quad \int u^*(t)\bar{v}^*(t)\mathrm{d}t = (2\pi)\int u(x)\bar{v}(x)\mathrm{d}x \ ,$$

where $\bar{z}$ denotes the conjugate of the complex number z.

2.2. The pile-up model with measurement error.

We consider experiments that provide observations

$$(1) \qquad\qquad Z_1, \ldots, Z_n$$

of independent identically distributed (i.i.d.) random variables with common density g and cumulative distribution function (c.d.f.) G. The random variables Z_k follow the model defined by

$$(2) \qquad Z_k = \min\{Y_{1,k} + \eta_{1,k}, \ldots, Y_{N_k,k} + \eta_{N_k,k}\} \ , \quad k = 1, \ldots, n \ ,$$

where

- [M1] the $(Y_{i,k})_{i,k\geq 1}$ are i.i.d. random variables with density f and c.d.f. F,
- [M2] the $(\eta_{i,k})_{i,k\geq 1}$ are i.i.d. with density f_η, which is assumed to be known and such that f_η^* does not vanish (i.e. $f_\eta^*(t) \neq 0$ for all $t \in \mathbb{R}$),
- [M3] the random variables N_k take their values in $\mathbb{N}^* = \{1, 2, \ldots\}$, are i.i.d. with the same distribution as N and such that $\mathbb{E}(N) < +\infty$. We denote by $M_\theta(u) = \mathbb{E}(u^N) = \sum_{k=1}^\infty u^k \mathbb{P}(N = k)$, for $u \in [0, 1]$, the probability generating function associated with N_1, and we assume that the function M_θ is known, up to some possible parameter θ.
- [M4] $(Y_{i,k})_{i,k\geq 1}$, $(\eta_{i,k})_{i,k\geq 1}$ and $(N_k)_{k\geq 1}$ are independent.

Our aim is to estimate the density f of the random variables $Y_{i,k}$ from the observations (1). Note that the random variables N_k are not observed.

Model (2) differs from a compound Poisson process in two aspects: the Z_j's are defined by a minimum instead of a sum $\sum_{i=1}^{N_j} Y_{i,j}$ and we have additional noise measurements η. Thus decompounding as in van Es et al. (2007) or Comte and Genon-Catalot (2010) does not apply. Note also that the assumption $f_\eta^*(t) \neq 0$ rules out uniform type distributions, for which specific methods have recently been developed (see Johnstone et al. (2004), Delaigle and Meister (2011)).

Main example. In the fluorescence application it is assumed that the number N of photons per excitation cycle follows a Poisson distribution with known parameter θ. Note that the events where no photon is detected, i.e. $N = 0$, are discarded from the sample. Hence, we consider a Poisson distribution restricted on $\mathbb{N}^*$ with renormalized probability masses given by

$$\mathbb{P}(N = k) = \frac{1}{\mathrm{e}^\theta - 1}\frac{\theta^k}{k!} \ .$$

Generally, θ is considered as known. Then the functions M_θ and $\dot{M}_\theta$ are known as well and given by $M_\theta(u) = (\mathrm{e}^{\theta u} - 1)/(\mathrm{e}^\theta - 1)$ and $\dot{M}_\theta(u) = \theta \mathrm{e}^{\theta u}/(\mathrm{e}^\theta - 1)$. We will discuss how θ is estimated and the cost of the substitution in Sections 3.3 and 4.2.

We shall also provide a discussion about the assumption that f_η is known in Section 4.3.

2.3. **Definition of the estimator.** It is easy to see that the c.d.f. G verifies

$$
\begin{aligned}
1 - G(z) &= \mathbb{P}(Z_1 > z) = \sum_{k=1}^{\infty} \mathbb{P}(N_1 = k)\mathbb{P}\left(\min_{1 \le i \le k}(Y_{i,1} + \eta_{i,1}) > z\right) \\
&= \sum_{k=1}^{\infty} \mathbb{P}(N_1 = k)[\mathbb{P}(Y_{1,1} + \eta_{1,1} > z)]^k .
\end{aligned}
$$

Therefore, if we denote by $F_{Y+\eta}$ the c.d.f. of $Y_{1,1} + \eta_{1,1}$, we get

$$
(3) \qquad\qquad G(z) = 1 - M_\theta \circ (1 - F_{Y+\eta})(z) .
$$

In Rebafka et al. (2010) and in the absence of noise, G is referred to as the *pile-up distribution function*. Recalling that M_θ is bijective, we deduce from (3) that

$$
(4) \qquad\qquad F_{Y+\eta}(z) = 1 - M_\theta^{-1} \circ (1 - G)(z) .
$$

The assumption $\mathbb{E}(N_1) < +\infty$ on N_1 implies that $\dot{M}_\theta$ exists with $\forall u \in [0,1]$, $\dot{M}_\theta(u) = \mathbb{E}(N_1 u^{N_1-1})$. Consequently, derivation can be applied on both sides of relation (4) and by using that the derivative of M_θ^{-1} is equal to $1/\dot{M}_\theta \circ M_\theta^{-1}$, we get for $f_{Y+\eta}$ denoting the density of $Y_{1,1} + \eta_{1,1}$,

$$
(5) \qquad f_{Y+\eta}(z) = w_\theta \circ G(z)g(z) \quad \text{with} \quad w_\theta(z) = \frac{1}{\dot{M}_\theta \circ M_\theta^{-1}(1-z)} .
$$

Note that since M_θ is assumed to be known, the weight function w_θ is also known. Moreover, equation (5) implies that moments of the target distribution of $Y_{1,1}$ are related to moments of the pile-up distribution of Z. Namely, for any measurable bounded function h, we have

$$
(6) \qquad\qquad \mathbb{E}(h(Y_{1,1} + \eta_{1,1})) = \mathbb{E}(w_\theta \circ G(Z_1)h(Z_1)) .
$$

Since $Y_{1,1}$ and $\eta_{1,1}$ are independent, we have $f_{Y+\eta} = f \star f_\eta$. By taking the Fourier transform and using that f_η^* does not vanish, we get for all $t \in \mathbb{R}$

$$
f^*(t) = \frac{f_{Y+\eta}^*(t)}{f_\eta^*(t)} ,
$$

and by the Fourier inverse formula

$$
f(z) = \frac{1}{2\pi} \int e^{izt} \frac{f_{Y+\eta}^*(t)}{f_\eta^*(t)} dt .
$$

Therefore, we propose the following estimator of f

$$
(7) \qquad\qquad \hat{f}_m(z) = \frac{1}{2\pi} \int_{-\pi m}^{\pi m} e^{izt} \frac{\widehat{f_{Y+\eta}^*}(t)}{f_\eta^*(t)} dt ,
$$

where the cutoff πm is required to ensure convergence of the integral. We use (5) and more precisely (6) to find an estimator of $f_{Y+\eta}^*$. Indeed, we have

$$
f_{Y+\eta}^*(t) = \mathbb{E}(e^{-it(Y_{1,1}+\eta_{1,1})}) = \mathbb{E}(e^{-itZ_1} w_\theta \circ G(Z_1)) ,
$$

yielding

$$
(8) \qquad \widehat{f_{Y+\eta}^*}(t) = \frac{1}{n} \sum_{k=1}^{n} w_\theta \circ \hat{G}_n(Z_k)e^{-itZ_k} , \quad \text{with} \quad \hat{G}_n(z) = \frac{1}{n} \sum_{k=1}^{n} \mathbb{1}_{\{Z_k \le z\}} .
$$

The estimate $\hat{G}_n$ is the standard empirical c.d.f.. We note that, since $w_\theta \circ \hat{G}_n(Z_{(k)}) = w_\theta(k/n)$ when $Z_{(k)}$ denotes the k-th order statistic associated with $(Z_1, \ldots, Z_n)$ satisfying $Z_{(1)} \leq \cdots \leq Z_{(n)}$, we can equivalently write

$$\widehat{f^*_{Y+\eta}}(t) = \frac{1}{n} \sum_{k=1}^{n} w_\theta(k/n) e^{-itZ_{(k)}} \ .$$

In the literature such weighted sums of order statistics are known as L-statistics.

Now, gathering (7) and (8), we get the estimator

$$(9) \qquad \hat{f}_m(z) = \frac{1}{2\pi} \int_{-\pi m}^{\pi m} e^{izt} \frac{\frac{1}{n} \sum_{k=1}^{n} w_\theta \circ \hat{G}_n(Z_k) e^{-itZ_k}}{f^*_\eta(t)} dt$$

$$= \frac{1}{2\pi n} \sum_{k=1}^{n} w_\theta \circ \hat{G}_n(Z_k) \int_{-\pi m}^{\pi m} \frac{e^{it(z-Z_k)}}{f^*_\eta(t)} dt$$

$$= \frac{1}{2\pi n} \sum_{k=1}^{n} w_\theta(k/n) \int_{-\pi m}^{\pi m} \frac{e^{it(z-Z_{(k)})}}{f^*_\eta(t)} dt \ .$$

The last two expressions show clearly why the cutoff is necessary: it is known that the Fourier Transform of a density tends to zero near infinity, i.e. here $f^*_\eta(t)$ tends to zero when t gets large and thus, the integral over $\mathbb{R}$ can not be defined. It is worth noting that the estimate given by formula (9) is real-valued: indeed, taking its conjugate leads to the same formula by virtue of the symmetry of the integration interval.

We can also see why the estimator $\hat{f}_m$ is going to involve technical difficulties for the theoretical study. Indeed, we apply simultaneously three approximations:

(1) the unknown c.d.f. G is approached by its empirical version,
(2) we deal with the pile-up effect by using weights $w_\theta \circ \hat{G}_n(Z_k)$, which can, by the way, be related to bias corrections in survival analysis,
(3) we deal with measurement errors by applying a deconvolution operator involving a cut-off parameter, $\cdot \mapsto \frac{1}{2\pi} \int_{-\pi m}^{\pi m} e^{it\cdot}/f^*_\eta(t)dt$.

It is interesting to mention that the estimator can be seen as a weighted kernel deconvolution estimator. Indeed, by setting $h = 1/m$, we have

$$(10) \quad \hat{f}_m(z) = \frac{1}{n} \sum_{k=1}^{n} w_\theta \circ \hat{G}_n(Z_k) \tilde{K}_h(z - Z_k), \quad \text{with} \quad \tilde{K}_h(z) = \frac{1}{2\pi h} \int e^{izt/h} \frac{K^*(t)}{f^*_\eta(t/h)} dt \ ,$$

where the kernel K is in this case the particular sinus cardinal kernel satisfying $K^*(t) = \mathbb{1}_{|t| \leq \pi}$. This makes a link with many other works in the kernel deconvolution setting, see Diggle and Hall (1993), Fan (1991), Delaigle and Gijbels (2004).

3. STUDY OF THE ESTIMATOR

3.1. **Risk bound on the estimator.** In addition to assumptions [M1]-[M4], we require

[M5] $\mathbb{P}(N = 1) \neq 0$ and $\mathbb{E}(N^2) < +\infty$.

The assumption $\mathbb{P}(N = 1) \neq 0$ is required to ensure that $\dot{M}_\theta(0) \neq 0$ and thus $w_\theta(u)$ is well defined for $u \in [0, 1]$. More specifically, the method does not work if $\mathbb{P}(N = 1) = 0$: we

can see from (5), that then, the link between the densities would fail for $z = 1$.

Note that, if $\mathbb{P}(N = 1) = 1$, then there is no pile-up effect and the problem becomes a pure deconvolution question. The assumption $\mathbb{E}(N^2) < +\infty$ ensures that $\ddot{M}_\theta$ is well defined. In other words, under [M5], we can see that for all $u \in [0, 1]$,

$$(11) \qquad 0 < \dot{M}_\theta(0) = \mathbb{P}(N = 1) < \dot{M}_\theta(u) < \dot{M}_\theta(1) = \mathbb{E}(N) < +\infty \,,$$

and

$$0 \leq \ddot{M}_\theta(0) = 2\mathbb{P}(N = 2) \leq \ddot{M}_\theta(u) \leq \ddot{M}_\theta(1) = \mathbb{E}[N(N - 1)] < +\infty \,.$$

This implies that for all $u \in [0, 1]$,

$$(12) \qquad 0 < w_\theta(0) = \frac{1}{\dot{M}_\theta(1)} = \frac{1}{\mathbb{E}(N)} \leq w_\theta(u) \leq w_\theta(1) = \frac{1}{\dot{M}_\theta(0)} = \frac{1}{\mathbb{P}(N = 1)} < \infty \,.$$

Moreover, noting that

$$\dot{w}_\theta(u) = \ddot{M}_\theta \circ M_\theta^{-1}(1 - u) / [\dot{M}_\theta \circ M_\theta^{-1}(1 - u)]^3 \,,$$

we also have, for all $u \in [0, 1]$,

$$0 \leq \dot{w}_\theta(u) \leq \frac{\sup_{u \in [0,1]} \ddot{M}_\theta(u)}{\inf_{u \in [0,1]} [\dot{M}_\theta(u)]^3} \leq \frac{\mathbb{E}[N(N - 1)]}{[\mathbb{P}(N = 1)]^3} < +\infty \,.$$

Lastly, it is convenient to mention that, still under [M5], w_θ is Lipschitz continuous, i.e.

$$(13) \qquad \text{there exists } c_{w,\theta} > 0 \text{ such that } \forall x, y \in [0, 1], \quad |w_\theta(x) - w_\theta(y)| \leq c_{w,\theta}|x - y| \,,$$

and clearly, we can take

$$c_{w,\theta} = \frac{\mathbb{E}[N(N - 1)]}{[\mathbb{P}(N = 1)]^3} \,.$$

Remark 3.1. In the more general nonlinear transformation model the function $M : [0, 1] \to [0, 1]$ in (3) is not necessarily a probability generating function, but any function M such that G given by (3) is a c.d.f. (Tsodikov, 2003). That is G is still the result of a distortion of the target distribution F, but the interpretation as a minimum is no longer valid. Those models are studied in survival analysis. The estimators proposed in this paper for the pile-up model are also applicable for nonlinear transformation models.

We now provide a risk bound for the estimator defined in (9).

Proposition 3.1. *Consider the model given by (1)-(2) under Assumptions* [M1]–[M5]. *Let f_m denote the function verifying $f_m^* = f^* \mathbb{1}_{[-\pi m, \pi m]}$. Then*

$$(14) \qquad \mathbb{E}(\|\hat{f}_m - f\|^2) \leq \|f - f_m\|^2 + C\frac{\Delta_\eta(m)}{n} \quad \text{where} \quad \Delta_\eta(m) = \frac{1}{2\pi} \int_{-\pi m}^{\pi m} \frac{\mathrm{d}u}{|f_\eta^*(u)|^2} \,,$$

and

$$C = 2 \left(\int_0^1 w_\theta^2(u) \mathrm{d}u + 2c_{w,\theta}^2 \right) \leq 2 \left(\frac{1}{\mathbb{P}(N = 1)} + 2 \left(\frac{\mathbb{E}[N(N - 1)]}{[\mathbb{P}(N = 1)]^3} \right)^2 \right) \,.$$

In the evaluation of C, we use that $\int_0^1 w_\theta^2(u) du = \int_0^1 (1/\dot{M}_\theta(u)) du$. In the bound on C, we can see that the smaller $\mathbb{P}(N = 1)$, the larger the bound. On the opposite, if $\mathbb{P}(N = 1) = 1$, then $c_{w,\theta} = 0$ and $\int_0^1 w_\theta^2(u) du = 1$ and C is minimal.

Moreover, when reading the estimator under the form (10), it is clear that the Central Limit Theorem for L-statistics proved in Rebafka et al. (2010) (see Appendix B, Theorem 5 and Proposition 1 therein) can be applied to obtain asymptotic normality results for

$\sqrt{n}(\hat{f}_m(z) - f_m(z))$ with computable variance, this is illustrated in Section 5. We will also use in the simulations the proposals given in Bissantz et al. (2007).

Note that $\|f - f_m\|^2 = (2\pi)^{-1} \int_{|u| \geq \pi m} |f^*(u)|^2 \mathrm{d}u$ and thus this quantity is clearly decreasing when m increases. On the contrary, the variance term $\Delta_\eta(m)/n$ is increasing with m. Hence (14) is a bias-variance decomposition, and a compromise is required for the choice of m.

Another remark is in order. Obviously, the variance depends crucially on the rate of decrease to 0 of f_η^* near infinity. For instance, if f_η is the standard normal density, the variance is proportional to $\int_{|u| \leq \pi m} \mathrm{e}^{u^2} \mathrm{d}u/n$. Whereas for the Laplace distribution where $f_\eta(x) = \mathrm{e}^{-|x|}/2$, we have $1/f_\eta^*(u) = 1 + u^2$ and a variance of order $O(m^4/n)$.

3.2. **Other ways to view the estimator.** The estimator $\hat{f}_m$ can also be obtained in a different way. It is classical to define a density estimator as the minimizer over a set of functions h of a contrast $\gamma_n(h)$ which is an approximation of $\|h - f\|^2 - \|f\|^2 = \|h\|^2 - 2\mathbb{E}[h(Y)]$. Writing $\mathbb{E}[h(Y)] = \langle h, f \rangle = (2\pi)^{-1} \langle h^*, f_Y^* \rangle = (2\pi)^{-1} \langle h^*, f_{Y_1+\eta_1}^*/f_\eta^* \rangle$ suggests to consider functions h in

$$S_m = \{h, \text{ support}(h^*) \subset [-\pi m, \pi m]\}$$

and the contrast

$$\gamma_n(h) = \|h\|^2 - \frac{1}{\pi} \int h^*(-u) \frac{\widehat{f_{Y_1+\eta_1}^*}(u)}{f_\eta^*(u)} \mathrm{d}u \, ,$$

where $\widehat{f_{Y_1+\eta_1}^*}$ is given by (8). Now we can see that the estimator $\hat{f}_m$ minimizes the contrast γ_n. Indeed, note that $\hat{f}_m^*(u) = \widehat{f_{Y_1+\eta_1}^*}/f_\eta^*(u) \, \mathbb{1}_{[-\pi m, \pi m]}(u)$ and thus $\hat{f}_m \in S_m$. By Parseval's formula $\langle h, \hat{f}_m \rangle = (2\pi)^{-1} \langle h^*, \hat{f}_m^* \rangle$. This yields that $\gamma_n(h) = \|h\|^2 - 2\langle h, \hat{f}_m \rangle = \|h - \hat{f}_m\|^2 - \|\hat{f}_m\|^2$. Therefore,

$$\hat{f}_m = \arg \min_{h \in S_m} \gamma_n(h).$$

Another expression of the estimator is obtained by describing more precisely the functional spaces S_m on which the minimization is performed. To that aim, let us define the sinc function and its translated-dilated versions by

$$(15) \qquad \varphi(x) = \frac{\sin(\pi x)}{\pi x} \quad \text{and} \quad \varphi_{m,j}(x) = \sqrt{m}\varphi(mx - j) \, ,$$

where m is an integer that can be taken equal to 2^ℓ. It is well known that $\{\varphi_{m,j}\}_{j \in \mathbb{Z}}$ is an orthonormal basis of the space of square integrable functions having Fourier transforms with compact support in $[-\pi m, \pi m]$ (Meyer, 1990, p.22). Indeed, as $\varphi^*(u) = \mathbb{1}_{[-\pi, \pi]}(u)$, an elementary computation yields that $\varphi_{m,j}^*(x) = m^{-1/2}\mathrm{e}^{-ixj/m}\mathbb{1}_{[-\pi m, \pi m]}(x)$. Thus, the functions $\varphi_{m,j}$ are such that $S_m = \text{Span}\{\varphi_{m,j}, \ j \in \mathbb{Z}\} = \{h \in \mathbb{L}_2(\mathbb{R}), \text{supp}(h^*) \subset [-m\pi, m\pi]\}$. For any function $h \in \mathbb{L}_2(\mathbb{R})$, let $\Pi_m(h)$ denote the orthogonal projection of h on S_m given by $\Pi_m(h) = \sum_{j \in \mathbb{Z}} a_{m,j}(h)\varphi_{m,j}$ with $a_{m,j}(h) = \int_{\mathbb{R}} \varphi_{m,j}(x)h(x)\mathrm{d}x$. As $a_{m,j}(h) = (2\pi)^{-1}\langle \varphi_{m,j}^*, h^* \rangle$, it follows that $\Pi_m(h)^* = h^*\mathbb{1}_{[-\pi m, \pi m]}$, and thus $f_m = \Pi_m(f)$. Since $\hat{f}_m$ minimizes γ_n, this yields that the estimator $\hat{f}_m$ can be written in the following convenient way

$$(16) \qquad \hat{f}_m = \sum_{j \in \mathbb{Z}} \hat{a}_{m,j}\varphi_{m,j} \quad \text{with } \hat{a}_{m,j} = \frac{1}{2\pi} \int \varphi_{m,j}^*(-u) \frac{\widehat{f_{Y_1+\eta_1}^*}(u)}{f_\eta^*(u)} \mathrm{d}u \, .$$

Consequently, $\|\hat{f}_m\|^2 = \sum_j |\hat{a}_{m,j}|^2$.

Finally, one can see that $\sum_{j\in\mathbb{Z}} \varphi^*_{m,j}(u)\varphi_{m,j}(x) = e^{-ixu}\mathbb{1}_{|x|\leq\pi m}$. This is another way to see that (16) and (7) actually define the same estimator.

Remark 3.2. An interesting remark follows from equation (16). In the case where no noise has to be taken into account, i.e. $f^*_\eta(u) \equiv 1$, the integral in (16) becomes $\int \varphi^*_{m,j}(-u)e^{-iuZ_k}du = 2\pi\varphi_{m,j}(Z_k)$. Hence, $\hat{a}_{m,j} = (1/n)\sum_{k=1}^n \varphi_{m,j}(Z_{(k)})w(k/n)$.

3.3. **Semi-parametric Poisson setting.** Clearly, Assumption [M5] is fulfilled in our main Poisson example. In this case, the weight function w_θ writes

$$(17) \qquad\qquad w_\theta(u) = \frac{1 - e^{-\theta}}{\theta(1 - u(1 - e^{-\theta}))} \ ,$$

with corresponding constant $c_{w,\theta} = (e^\theta - 1)^2/\theta$ in (13). It is interesting to note that $\mathbb{P}(N = 1) = \theta/(e^\theta - 1)$ is a decreasing function of θ. As we have seen that many important bounds depend on $1/\mathbb{P}(N = 1)$, we conclude that the smaller θ, the better the estimation procedure. We shall see in Section 5, that this remark is confirmed by the simulation experiments.

In practice, even if physicists consider θ as known, it is in fact estimated as follows.
In the fluorescence setting, N has indeed a classical Poisson distribution on $\{0, 1, \dots\}$ with parameter θ. That means, that there are excitations that are not followed by the emission of photons. In this case, we set the variable to a default value, here $+\infty$. More precisely, observations $Z_1, \dots, Z_n$ are i.i.d. and defined by

$$Z_j = \begin{cases} \min\{Y_{1,j} + \eta_{1,j}, \dots, Y_{N_j,j} + \eta_{N_j,j}\} \ , & \text{if } N_j > 0 \\ +\infty \ , & \text{if } N_j = 0. \end{cases} \ ,$$

Therefore, one can use the proportion of observations Z_j equal to $+\infty$ to estimate the Poisson parameter θ. As $\mathbb{P}(Z_1 = +\infty) = \mathbb{P}(N_1 = 0) = e^{-\theta}$, a natural estimator of θ is given by

$$(18) \qquad \hat{\theta} = -\log\left(\frac{1}{n+1}\sum_{j=1}^n \mathbb{1}\{Z_j = +\infty\} + \frac{1}{n+1}\right) = \log\left(\frac{n+1}{\#\{Z_j = +\infty\}+1}\right) \ .$$

Now the following question naturally arises. Let $\hat{\theta}$ be a consistent estimator of θ, we can wonder what conditions ensure a satisfactory behavior of

$$(19) \quad \hat{f}_{m,\hat{\theta}}(z) = \frac{1}{2\pi}\int_{-\pi m}^{\pi m} e^{izt}\frac{\frac{1}{n}\sum_{k=1}^n w_{\hat{\theta}} \circ \tilde{G}_n(Z_k)e^{-itZ_k}}{f^*_\eta(t)}dt \ , \quad \tilde{G}_n(z) = \frac{1}{n+1}\sum_{k=1}^n \mathbb{1}_{\{Z_k \leq z\}} \ .$$

In fact, we can extend the result of Proposition 3.1 under the following condition:
[M6] (i) $\theta \in [0, \theta_{max}]$, $\mathbb{E}[(\hat{\theta} - \theta)^{2k}] \leq C_1/n^k$ for $k = 1, 2, 1 + 2a$,

(ii) $\displaystyle\sup_{u\in[0,1]}\sup_{\theta\in[0,2\theta_{max}]}\left|\frac{\partial w_\theta(u)}{\partial\theta}\right| \leq C_2(\theta_{max}) < +\infty$,

(iii) $\displaystyle\sup_{u\in[0,\frac{n}{n+1}]}\sup_{\theta\in\mathbb{R}^+}\left|\frac{\partial w_\theta(u)}{\partial\theta}\right| \leq C_3 n^a$, for some integer $a \geq 1$.

Then, the following result holds.

Proposition 3.2. *Consider the model given by (1)-(2) under Assumptions [M1]–[M6]. Then*

$$\mathbb{E}\left(\|\hat{f}_{m,\hat{\theta}} - f\|^2\right) \le \|f - f_m\|^2 + C_4 \frac{\Delta_\eta(m)}{n} \ , \tag{20}$$

where

$$C_4 = 3\left(\int_0^1 w_\theta^2(u)du + 4c_{w,\theta}^2 + C_1 C_2(\theta_{max}) + \frac{C_1 C_3}{\theta_{max}^6}\right) \ ,$$

and C_1, C_2, C_3 are defined in [M6].

Moreover, in the Poisson case, if we assume that θ belongs to $[0, \theta_{max}]$, then $\hat{\theta}$ defined by (18) satisfies [M6] with $C_2(\theta_{max}) = e^{2\theta_{max}}$, $C_3 = 2$ and $a = 1$.

We can see that inequality (20) extends inequality (14) for known θ with simply increased multiplicative constant.
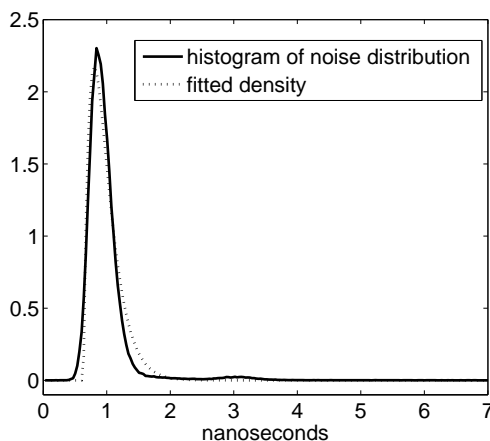

FIGURE 1. Normalized histogram based on a sample of the noise distribution (solid line) and the fitted density (dashed line) having the form of (21) with $\hat{\alpha} = 0.961$, $\hat{\beta} = 0.941$, $\hat{\nu} = 5.74$, $\hat{\tau} = 5.89$.

3.4. **Discussion on the type of noise.** To determine the rate of convergence of the MISE, it is necessary to specify the type of the noise distribution. Let us consider two examples.

First, the noise distribution can be exponential with density given by $f_\eta(x) = \theta e^{-\theta x} \mathbb{1}_{x>0}$, for some $\theta > 0$. Then we have $f_\eta^*(u) = \theta/(\theta + iu)$, $|f_\eta^*(u)|^2 = 1/(1 + u^2/\theta^2)$ and $\Delta_\eta(m) = m + \pi^2 m^3/(3\theta^2)$.

In the fluorescence setting, we found that TCSPC noise distributions can be approximated by densities of the following form

$$f_\eta(x) = \left(\frac{\alpha\nu}{\alpha - \beta}e^{-\nu x} - \frac{\beta\tau}{\alpha - \beta}e^{-\tau x}\right)\mathbb{1}_{\{x>0\}} \ , \tag{21}$$

with constraints $\alpha > \beta$, $\nu < \tau$, $\beta\tau/(\alpha\nu) \ge 1$. Figure 1 presents a dataset with 259,260 measurements from the noise distribution of a TCSPC instrument (independently from the fluorescence measurements) and the corresponding estimated density having form (21) obtained by least squares fitting. Even though the fit is not perfect, the estimated density captures the main features of the dataset. Thus densities of the form (21) can be considered

as a good approximative model of the noise distribution in the fluorescence setting. In the general case of (21) we have

$$f_\eta^*(u) = \frac{\alpha \nu}{\alpha - \beta} \frac{1}{\nu + iu} - \frac{\beta \tau}{\alpha - \beta} \frac{1}{\tau + iu} \ .$$

In the simulation study we will consider a noise distribution of the form (21) with parameters $\alpha = 2, \beta = 1, \nu = 1, \tau = 2$. In this case we get

$$(22) \qquad |f_\eta^*(u)|^2 = \frac{4}{(1 + u^2)(4 + u^2)} \qquad \text{and} \qquad \Delta_\eta(m) = m + \frac{5}{12}\pi^2 m^3 + \frac{1}{20}\pi^4 m^5 \ .$$

¿From the application viewpoint it is hence interesting to consider the class of noise distributions η whose characteristic functions decrease in the *ordinary smooth* way of order γ, denoted by $\eta \sim OS(\gamma)$, defined by

$$(23) \qquad c_0(1 + u^2)^{-\gamma} \leq |f_\eta^*(u)|^2 \leq C_0(1 + u^2)^{-\gamma}.$$

Clearly, we find that $\Delta_\eta(m) = O(m^{2\gamma+1})$.

3.5. **Rates of convergence on Sobolev spaces.** In classical deconvolution the regularity spaces used for the functions to estimate are Sobolev spaces defined by

$$\mathcal{C}(a, L) = \left\{ g \in (\mathbb{L}^1 \cap \mathbb{L}^2)(\mathbb{R}), \ \int (1 + u^2)^a |g^*(u)|^2 du \leq L \right\} \ .$$

If f belongs to $\mathcal{C}(a, L)$, then

$$\begin{aligned} 2\pi \|f - f_m\|^2 &= \int_{|u| \geq \pi m} |f^*(u)|^2 du = \int_{|u| \geq \pi m} (1 + u^2)^a |f^*(u)|^2 / (1 + u^2)^a du \\ &\leq (1 + (\pi m)^2)^{-a} L \leq L(\pi m)^{-2a} \ . \end{aligned}$$

Therefore, if $f \in \mathcal{C}(a, L)$ and $\eta \sim OS(\gamma)$, Proposition 3.1 implies that

$$\mathbb{E}(\|\hat{f}_m - f\|^2) \leq C_1 m^{-2a} + C_2 n^{-1} m^{2\gamma+1}.$$

The optimization of this upper bound provides the optimal choice of m by $m_{opt} = O\left(n^{1/(2a+2\gamma+1)}\right)$ with resulting rate $\mathbb{E}(\|\hat{f}_m - f\|^2) = O\left(n^{-2a/(2a+2\gamma+1)}\right)$. More formally, one can show the following result.

Proposition 3.3. *Assume that the assumptions of Proposition 3.1 are satisfied and that $f \in \mathcal{C}(a, L)$ and $\eta \sim OS(\gamma)$ (see (23)). Then for $m_{opt} = O(n^{1/(2a+2\gamma+1)})$, we have*

$$\mathbb{E}(\|\hat{f}_{m_{opt}} - f\|^2) = O(n^{-2a/(2a+2\gamma+1)}) \ .$$

Obviously, in practice the optimal choice m_{opt} is not feasible since a and part of the constants involved in the order are unknown. Therefore, another model selection device is required to choose a relevant $\hat{f}_m$ in the collection.

4. Automatic cutoff selection

4.1. **Adaptive result.** The general method consists in finding a data driven penalty pen(.) such that the following model

$$(24) \qquad \hat{m} = \arg \min_{m \in \mathcal{M}_n} \{\gamma_n(\hat{f}_m) + \text{pen}(m)\}$$

achieves a bias-variance trade-off, where $\mathcal{M}_n$ has to be specified. Usually, the penalty has the same order as the variance term, while $\gamma_n(\hat{f}_m) = -\|\hat{f}_m\|^2$ approximate the squared-bias $\|f - f_m\|^2 = \|f\|^2 - \|f_m\|^2$ up to the constant $\|f\|^2$.

Here in contrast to this general approach our result involves an additional $\log(n)$-factor in the penalty compared to the variance order, which implies a loss with respect to the expected rate derived in Section 3.5. This is certainly due to the complexity of the problem which involves three sources of errors.

Theorem 4.1. *Consider the model given by (1)-(2) under Assumptions* [M1]–[M5]. *Assume that f is square integrable on $\mathbb{R}$ and $\eta \sim OS(\gamma)$. Consider the estimator $\hat{f}_{\hat{m}}$ with model $\hat{m}$ defined by (24) with penalty*

$$\text{(25)} \qquad \text{pen}(m) = \kappa \left(\int_0^1 w_\theta^2(u)\mathrm{d}u + \kappa' c_{w,\theta}^2 \log(n) \right) \frac{\Delta_\eta(m)}{n} \,,$$

where κ and κ' are numerical constants. Assume moreover that η is ordinary smooth, i.e. $\eta \sim OS(\gamma)$, and that the model collection is described by

$$\mathcal{M}_n = \{m \in \mathbb{N}, \Delta_\eta(m) \leq n\} = \{1, \ldots, m_n\} \,.$$

Then, there exist constants κ, κ' such that

$$\text{(26)} \qquad \mathbb{E}\left(\|\hat{f}_{\hat{m}} - f\|^2 \right) \leq C \inf_{m \in \mathcal{M}_n} \left(\|f - f_m\|^2 + \text{pen}(m) \right) + C'\frac{\log(n)}{n} \,,$$

where C is a numerical constant and C' depends on $c_{w,\theta}$ and the bounds on w_θ.

The numerical constants κ and κ' are calibrated via simulations. In practice, to compute $\hat{m}$ by (24), we approximate $\gamma_n(\hat{f}_m) = -\|\hat{f}_m\|^2 = -\sum_{j\in\mathbb{Z}} |\hat{a}_{m,j}|^2$ by $-\sum_{|j|\leq K_n} |\hat{a}_{m,j}|^2$, where the sum is truncated to K_n of order n. We refer to Comte et al. (2006) for theoretical justifications of this truncation, see also Bissantz et al. (2005).

A better result can be obtained with additional technicalities in the proof and under slightly stronger assumptions: the price to pay for avoiding the log-loss in the penalty and thus in the rate.

Theorem 4.2. *Assume that all assumptions of Theorem 4.1 hold. In addition, assume that $\mathbb{E}(N^3) < +\infty$ and that $\hat{G}_n(z)$ is estimated with a sample $(Z_{-j})_{1\leq k\leq n}$ independent of $(Z_k)_{1\leq k\leq n}$ and from the same distribution G. Then inequality (26) can be obtained with*

$$\text{(27)} \qquad \text{pen}(m) = \tilde{\kappa} \left(\int_0^1 w_\theta^2(u)\mathrm{d}u + \tilde{\kappa}' \int_0^1 \dot{w}_\theta^2(u)\mathrm{d}u \right) \frac{\Delta_\eta(m)}{n}$$

for some numerical constants $\tilde{\kappa}, \tilde{\kappa}'$.

Of course, the result in Theorem 4.2 requires to split the sample, but in the fluorescence context, this is feasible since very large samples are available. The assumption that $\mathbb{E}(N^3) < +\infty$ is a weak constraint and is fulfilled in our main Poisson example. Then we can see that there is no longer the $\log(n)$ factor in the penalty, so that the estimator can reach the optimal rate without loss. Indeed, following steps analogous to Section 3.5, we can easily see that, if f belongs to a Sobolev space $\mathcal{C}(a, L)$, then the order of the right-hand side of inequality (26) is $O\left((n/\log(n))^{-2a/(2a+2\gamma+1)}\right)$ in Theorem 4.1 and of order $O\left(n^{-2a/(2a+2\gamma+1)}\right)$ in Theorem 4.2.

4.2. **Adaptive estimation in the semi-parametric Poisson case.** Here we consider that N follows a Poisson distribution with parameter θ, which we estimate by $\hat{\theta}$ defined in (18). Then we consider $\hat{f}_{m,\hat{\theta}}$ as given by (19). From the theoretical point of view, the adaptive procedure gets slightly complicated by the fact that the penalty becomes a random quantity. In this case, we can prove the following result.

Theorem 4.3. *Consider the model given by (1)-(2) and $N \sim \mathcal{P}(\theta)$, under Assumptions [M1]–[M5] and $\theta \in [0, \theta_{max}]$. Assume that f is square integrable on $\mathbb{R}$ and $\eta \sim OS(\gamma)$. Consider the estimator $\hat{f}_{\hat{m},\hat{\theta}}$ given by (19) with model $\hat{m}$ defined as $\hat{m} = \arg\min_{m \in \mathcal{M}_n}\{\gamma_n(\hat{f}_{m,\hat{\theta}}) + \widehat{\mathrm{pen}}(m, \hat{\theta})\}$ with $\hat{\theta}$ given by (18) and penalty*

$$(28) \qquad \widehat{\mathrm{pen}}(m, \hat{\theta}) = \tilde{\kappa}\left(\int_0^1 w_{\hat{\theta}}^2(u)\mathrm{d}u + \tilde{\kappa}'(c_{w,\hat{\theta}}^2 + \mathrm{e}^{3\theta_{max}})\log(n)\right)\frac{\Delta_\eta(m)}{n},$$

where κ and κ' are some numerical constants. Assume moreover that η is ordinary smooth, i.e. $\eta \sim OS(\gamma)$, and that the model collection is described by

$$\mathcal{M}_n = \{m \in \mathbb{N}, \Delta_\eta(m) \le n\} = \{1, \ldots, m_n\}.$$

Then, there exist constants $\tilde{\kappa}, \tilde{\kappa}'$ such that

$$(29) \qquad \mathbb{E}\left(\|\hat{f}_{\hat{m},\hat{\theta}} - f\|^2\right) \le C \inf_{m \in \mathcal{M}_n}\left(\|f - f_m\|^2 + \widehat{\mathrm{pen}}(m, \theta)\right) + C'\frac{\log(n)}{n},$$

where C is a numerical constant and C' depends on $c_{w,\theta}$ and the bounds on w_θ.

The bound given by (29) has the same order as in the one-sample result for known θ. The penalty is now given by (28) which looks like (25) in the known θ case, with only a slight increase. Therefore, the estimation strategies are robust to a missing parameter estimation.

4.3. **Discussion about unknown noise density.** In the fluorescence set-up, the noise distribution f_η is generally unknown. However, independent, large samples of the noise distribution are available. Hence one may still use the procedure proposed above by replacing f_η^* with the estimate $\hat{f}_\eta^*(u) = \sum_{k=1}^M \mathrm{e}^{-iu\eta_{-k}}/M$, where $(\eta_{-k})_{1 \le k \le M}$ denotes the independent noise sample. In Comte and Lacour (2011) the same substitution is considered for deconvolution methods. It is shown that for ordinary smooth noise this leads to a risk bound exactly analogous to the one given in (26). The main constraint given in Comte and Lacour (2011) is that $M \ge n^{1+\epsilon}$, for some $\epsilon > 0$. As the noise samples provided in the fluorescence setting have huge size, this condition is certainly fulfilled in our practical examples. In the following numerical study we consider the estimator with both the exact f_η^* and an estimated $\hat{f}_\eta^*$. We refer to Comte and Lacour (2011) for details on the procedure. Theoretical justifications are beyond the scope of this paper, but would clearly require a considerable amount of work due to our additional measurement errors (pile-up effect and c.d.f. estimation).

5. NUMERICAL RESULTS FOR SIMULATED AND REAL DATA

In this section we first give details on the practical implementation of the estimation method. Then a simulation study is conducted to test the performance of the method in different settings. Finally, an application to a sample of fluorescence data shows that the estimation method gives satisfying results on real measurements.

5.1. **Practical computation of estimators.** For practical computation of the estimator $\hat{f}_m$ proposed in Section 3, the presentation of the estimator in the sinc basis given by (16) is convenient. So one has to compute the coefficients $\hat{a}_{m,j}$. For $j \geq 0$, they can be approximated as follows

$$\hat{a}_{m,j} = \frac{1}{2\pi} \int \varphi^*_{m,j}(-u) \frac{\widehat{f^*_{Y_1+\eta_1}}(u)}{f^*_\eta(u)} \mathrm{d}u = (-1)^j \frac{\sqrt{m}}{2} \int_0^2 e^{i\pi jv} \frac{\widehat{f^*_{Y_1+\eta_1}}(\pi m(v-1))}{f^*_\eta(\pi m(v-1))} \mathrm{d}v$$

$$\approx (-1)^j \frac{\sqrt{m}}{T} \sum_{t=0}^{T-1} e^{i2\pi jt/T} \frac{\widehat{f^*_{Y_1+\eta_1}}(\pi m(\frac{2t}{T}-1))}{f^*_\eta(\pi m(2t/T-1))} = (-1)^j \sqrt{m}(\mathrm{IFFT}(H))_j = \breve{a}_{m,j} \ ,$$

where IFFT(H) is the inverse fast Fourier transform of the T-vector H whose t-th entry equals $\widehat{f^*_{Y_1+\eta_1}}(\pi m(2t/T - 1))/f^*_\eta(\pi m(2t/T - 1))$. Similarly, for $j < 0$ the coefficients $\hat{a}_{m,j}$ are approximated by $\breve{a}_{m,j} = (-1)^j \sqrt{m}(\mathrm{IFFT}(\overline{H}))_j$.

The integral $\Delta_\eta(m)$ appearing in the penalty term pen(m) defined in (25) is explicitly known if f_η is known (see Section 3.4). In the case where we only have an estimator $\hat{f}_\eta$, $\Delta_\eta(m)$ can be approximated by a Riemann sum of the form

$$(m/S) \sum_{s=0}^{S} |\hat{f}^*_\eta(-\pi m(1 - \frac{2s}{S}))|^{-2}.$$

Then the best model $\hat{m}$ is selected as the point of minimum of the criterion given in (24). Finally, we obtain the estimator $\hat{f}_{\hat{m}} = \sum_{j=-T}^{T} \breve{a}_{\hat{m},j} \varphi_{\hat{m},j}$ with the sinc functions $\varphi_{m,j}$ defined in (15).

Figure 2 presents the visual summary of our simulation results. We implemented the estimation method when f has one of the following forms:

(1) a Gamma(3, 3) p.d.f given by $1/(2!3^3)x^2 \exp(-x/3)\mathbb{1}_{x>0}$, to have a benchmark with a smooth distribution,
(2) an exponential p.d.f. given by $(1/3) \exp(-x/3)\mathbb{1}_{x>0}$,
(3) a Pareto(1/4, 1, 0) p.d.f. given by $(1 + x/4)^{-5}\mathbb{1}_{x>0}$,
(4) a Weibull(1/4, 3/4) p.d.f. given by $(3/4)(1/4)^{-3/4}x^{-1/4} \exp(-(4x)^{3/4})\mathbb{1}_{x>0}$.

The last two densities are inspired by chemical results about fluorescence phenomena given in Berberan-Santos et al. (2005a,b).

5.2. **Simulation study.**

Implementation of the estimation method. The adaptive estimator described in Section 3 is tested with the numerical constants $\kappa = 1$ and $\kappa' = 0.001$ in (25) and with N following a Poisson distribution. The value of κ' is very small and makes the logarithmic term in general negligible except when $c^2_{w,\theta}$ is large (for instance $c^2_{w,\theta} \approx 416$ for $\theta = 2$). The results are given in Figure 2. The observations are such that $Y + \eta$ with $\eta = \sigma\varepsilon$. In the first row, the pile-up effect is almost negligible ($\theta = 0.01$), but σ is rather large. That is, the first row illustrates the performance of the deconvolution step of the estimation procedure. In contrast, for the last row, σ is taken to be small, but the pile-up effect is significant ($\theta = 2$), in order to see how the estimator copes with the pile-up effect. The second row is an intermediate situation, illustrating how the estimator performs when the variance of the noise and the pile-up effect are both non negligible.

Gamma	Exponential	Pareto	Weibull
3.36 (0.49)	14.2 (2.1)	22.4 (1.8)	43.1 (5.9)

Estimation with $\sigma = 0.7$, $\theta = 0.01$, $n = 2000$

| 3.36 (0.57) | 15.2 (1.6) | 27.1 (3.6) | 48.9 (6.3) |

Estimation with $\sigma = 0.5$, $\theta = 1$, $n = 2000$

| 3.00 (0.00) | 15.3 (2.0) | 33.6 (3.5) | 81.0 (7.6) |

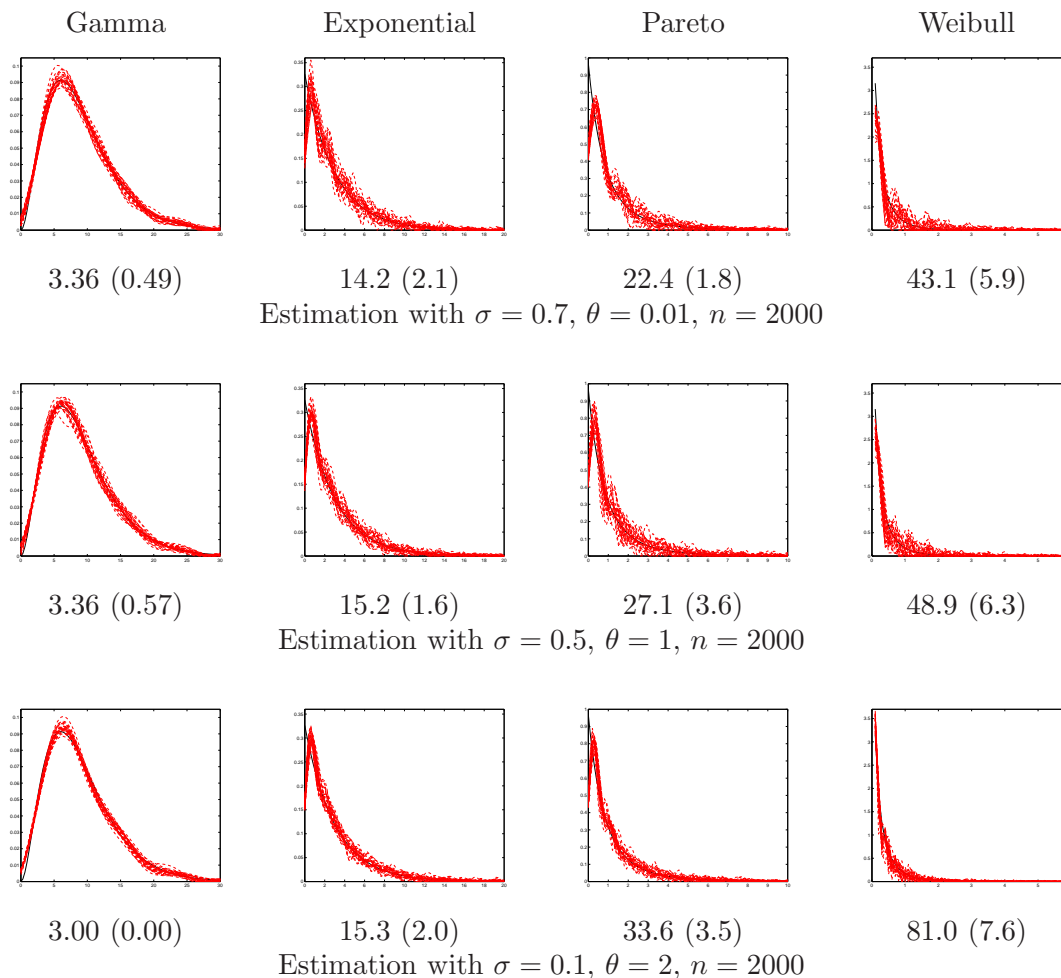Estimation with $\sigma = 0.1$, $\theta = 2$, $n = 2000$

FIGURE 2. True density and 25 estimated curves. Estimation by deconvolution with sinc basis for different noise levels and different levels of the pile-up effect. Numbers below the figures indicate mean and standard deviation of the selected model $\hat{m}$.

The 25 curves indicate variability bands for the estimation procedure. They show that the estimator is quite stable, especially in the last rows. Moreover, the selected model order $\hat{m}$ is different from one example to the other. Globally the selected cutoff $\hat{m}$ increases when going from example 1 to 4. That means that the estimator adapts to the peaks that are more and more difficult to recover.

In Table 1 the MISE of the estimation procedure is analyzed. The table gives the empirical mean and standard deviation of the MISE obtained over 100 simulated datasets. This is done for the same four examples of distributions as above. We compare the error for the estimator using the exact noise distribution to the estimator based on an approximation of the noise distribution based on an independent noise sample of size 500. Moreover, we study the influence of the noise distribution on the estimator. Therefore, we consider, on the one hand exponential noise with variances $\sigma^2 \in \{0.2, 1\}$, and on the other hand density (21) with $\alpha = 2$, $\beta = 1$, $\nu = 1$, $\tau = 2$ (multiplied with adequate constants to have the same variance σ^2 as for the exponential distributions).

	Exponential noise					
(σ^2, θ)	$(0.2, 0.5)$	$(0.2, 1.5)$	$(0.2, 2)$	$(1, 0.5)$	$(1, 1.5)$	$(1, 2)$
Gamma	.063 (.042)	.081 (.045)	.112 (.026)	.061 (.039)	.088 (.040)	.115 (.028)
	.063 (.042)	.081 (.045)	.112 (.026)	.061 (.039)	.087 (.040)	.115 (.028)
Exponential	1.11 (0.22)	1.20 (0.26)	1.45 (0.21)	1.36 (0.26)	1.40 (0.24)	1.67 (0.27)
	1.11 (0.22)	1.19 (0.25)	1.46 (0.21)	1.36 (0.27)	1.40 (0.24)	1.67 (0.27)
Pareto	4.25 (0.82)	4.55 (0.58)	5.45 (0.84)	6.62 (1.5)	6.58 (0.95)	8.09 (1.2)
	4.23 (0.83)	4.56 (0.61)	5.47 (0.83)	6.62 (1.6)	6.58 (1.0)	8.09 (1.2)
Weibull	10.6 (6.7)	9.46 (5.0)	9.22 (2.7)	21.4 (4.1)	26.7 (5.6)	39.5 (5.9)
	8.54 (4.7)	9.40 (4.8)	9.30 (2.3)	22.1 (4.8)	26.7 (5.7)	40.1 (5.7)

	Bi-exponential noise					
(σ^2, θ)	$(0.2, 0.5)$	$(0.2, 1.5)$	$(0.2, 2)$	$(1, 0.5)$	$(1, 1.5)$	$(1, 2)$
Gamma	.060 (.032)	.075 (.040)	.113 (.023)	.061 (.048)	.088 (.043)	.114 (.025)
	.060 (.032)	.075 (.040)	.113 (.023)	.062 (.048)	.089 (.043)	.114 (.025)
Exponential	1.06 (0.20)	1.14 (0.17)	1.49 (0.26)	1.23 (0.27)	1.37 (0.28)	1.62 (0.28)
	1.06 (0.20)	1.14 (0.16)	1.48 (0.25)	1.25 (0.26)	1.37 (0.28)	1.62 (0.27)
Pareto	4.15 (0.76)	4.31 (0.69)	5.08 (0.71)	6.08 (1.5)	6.41 (1.1)	7.43 (1.0)
	4.14 (0.77)	4.30 (0.69)	5.07 (0.72)	6.11 (1.6)	6.49 (1.2)	7.45 (1.1)
Weibull	10.2 (6.1)	8.89 (5.6)	8.29 (2.1)	24.7 (3.9)	29.4 (4.5)	40.1 (5.2)
	8.25 (4.3)	8.75 (5.4)	8.31 (2.2)	24.9 (4.3)	29.5 (4.9)	40.4 (5.3)

TABLE 1. $100 \times$ mean MISE and standard deviation in parentheses, sample size $n = 1000$. First lines correspond to exact noise distribution, second lines give results obtained with estimated noise distribution with additional noise sample of size 500.

¿From Table 1 it is clear that increasing the variance of the noise distribution increases the error. Furthermore, changing the type of the noise does not influence a lot the estimation procedure. Indeed, the second case (21) is just slightly less favorable than the exponential distribution. This difference is in accordance with Proposition 3.3 that holds with $\gamma = 1$ for the exponential and with $\gamma = 2$ for the other density. The comparison with the results based on an approximated noise distribution (second lines) reveals that there is rarely a difference between the two methods. Indeed, using an approximation of the noise does not corrupt the results, in some cases we even observe an improvement of the error. We show in Figure 3 that it is indispensable to take into account both the pile-up correction (which is omitted in (b) where $w(k/n)$ is replaced by 1) and the deconvolution correction (which is omitted in (c) where the estimation is done with the method of Brunel et al. (2009) with the trigonometric basis, se! e also Remark 3.2). Thus, we conclude from these simulation results for the fluorescence setting that it is justified to use an estimate of the noise instead of the theoretical distribution.

Influence of the distribution of N. Table 2 illustrates the effect on the MISE of different laws of N, namely a Poisson distribution, a geometric distribution $\mathrm{Geo}(p)$ and a uniform distribution on $\{1, \ldots, k_0\}$. The first rows give the MISE and associated variance
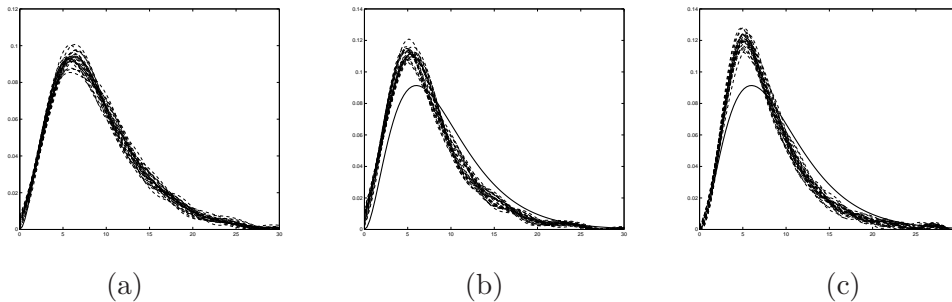
FIGURE 3. (a) Estimation with pile-up correction and deconvolution. (b) No pile-up correction. (c) No deconvolution.

$\mathbb{E}[N] = 4$, $\Gamma(2,2)$, $n = 1.000$				
			MISE	
	parameter	$\mathbb{P}(N=1)$	mean$\times 10^3$	variance$\times 10^3$
Poisson distribution	$\theta = 3.92$	0.079	5.91	0.0101
Geometric distribution	$p = 0.25$	0.25	3.85	0.0046
Uniform distribution	$k_0 = 7$	0.14	4.91	0.0066

$\mathbb{P}(N=1) = 0.10$, $\mathrm{Par}(4,4)$, $n = 2.000$				
			MISE	
	parameter	$\mathbb{E}[N]$	mean$\times 10^3$	variance$\times 10^3$
Poisson distribution	$\theta = 3.62$	3.72	4.51	0.0019
Geometric distribution	$p = 0.10$	10	4.64	0.0025
Uniform distribution	$k_0 = 10$	5.5	4.21	0.0017

TABLE 2. Comparison of the MISE and associated variance based on 300 repetitions for different laws of N, namely Poisson distribution $\mathrm{Poi}(\theta)$, geometric distribution $\mathrm{Geo}(p)$ and uniform distribution on $\{1, \ldots, k_0\}$. The target distribution f is a Gamma $\Gamma(2,2)$ and a Pareto $\mathrm{Par}(4,4)$, respectively. The additive noise η is exponentially distributed with $\sigma = 0.7$.

over 300 repetitions when the expected mean $\mathbb{E}[N]$ is fixed. We see that the geometric distribution corresponds to the smallest MISE, whereas the Poisson distribution has the worst performance. We note that at the same time the geometric distribution has the highest probability $\mathbb{P}(N = 1)$ and the Poisson distribution the lowest. We have seen that this probability plays a key role in the theoretical study of the estimator. The last rows of Table 2 refer to the case where the probability $\mathbb{P}(N = 1)$ is the same for the three distributions of N. Despite the differences of the three distributions (see e.g. the different associated values of $\mathbb{E}[N]$), the performance of the estimator in terms of the MISE is quite the same. This confirms that the probability $\mathbb{P}(N = 1)$ is the decisive characteristic of the law of N for the performance of the estimator.

About confidence intervals. Lastly, Figure 4 shows confidence intervals with confidence level 0.90 obtained by different procedures for data ($n = 1000$ and $n = 4000$) from a Gamma $\Gamma(2,2)$-distribution in the Poisson case (with $\theta = 0.55$). Figure 4 (a) represents asymptotic confidence intervals for $f_m(z)$ based on a result on the asymptotic normality of $\sqrt{n}(\hat{f}_m(z) - f_m(z))$ and a consistent estimator of the limit variance (see Rebafka et al.
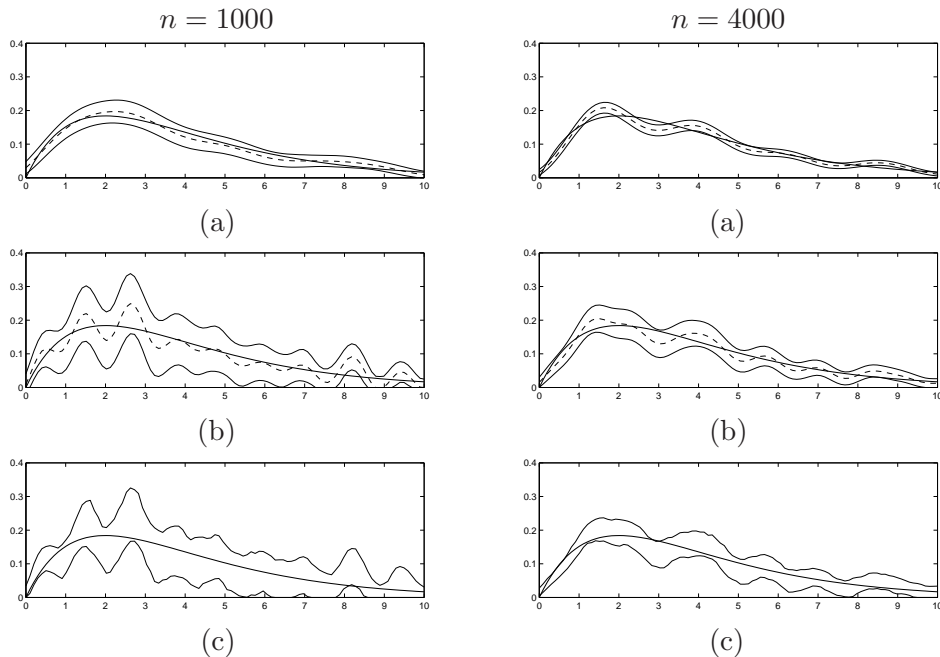
$$n = 1000 \qquad\qquad n = 4000$$

FIGURE 4. (a) Asymptotic confidence intervals for $f_m(z)$. (b) Asymptotic confidence intervals for $f(z)$. (c) Bootstrap confidence intervals for $f(z)$.

(2010)). To obtain confidence intervals of $f(z)$, we adapted the approach of Bissantz et al. (2007) for the construction of confidence bands in a different context. The procedure consists in computing asymptotic confidence intervals from the result on the asymptotic normality of $\sqrt{n}(\hat{f}_m(z) - f_m(z))$, however, by selecting a larger m than the one proposed by our data-driven model selection tool $\hat{m}$. This kind of under smoothing leads to larger intervals that contain $f(z)$ with the required confidence level when the parameters of the procedure are well tuned. The same technique of under smoothing in combination with bootstrapping the quantity $\sqrt{n}(\hat{f}_m(z) - f_m(z))$ yields the confidence intervals of $f(z)$ in Figure 4 (c). The construction of confidence bands for f is beyond the scope of this paper, as it requires a non trivial analysis of a process involving L-statistics.

5.3. **Application to Fluorescence Measurements.** We finally apply the estimation procedure to real fluorescence lifetime measurements obtained by TCSPC. The data analyzed here are graphically presented in Figure 5 (a) by the histogram of the fluorescence lifetime measurements and the histogram of the noise distribution based on a sample obtained independently from the fluorescence measurements. The sample size of the fluorescence measurements is $n = 1,743,811$. The same sample of the noise distribution has already been considered in Figure 1, where it is compared to the parameterized density given by (21). In this setting the true density is known to be an exponential distribution with mean 2.54 nanoseconds and the Poisson parameter equals 0.166. The knowledge of the true density allows to evaluate the performance of our estimator. More details on the data and their acquisition can be found in Patting et al. (2007).

We apply the estimator of Section 3 with the sinc basis to this dataset. We recall that the numerical constants are $\kappa = 1$ and $\kappa' = 0.001$. Figure 5 (b) shows the estimation result in comparison to the exponential density with mean 2.54. We observe that the estimated
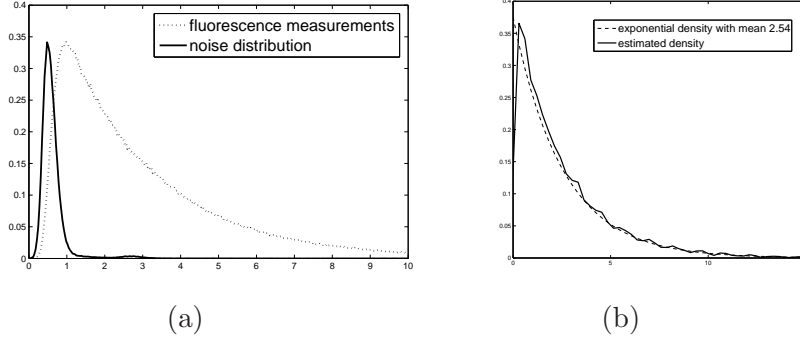
FIGURE 5. (a) Fluorescence lifetime measurements (solid line) and independent sample of the noise distribution (dashed). (b) Density estimator (solid) and 'true' exponential density with mean 2.54 (dashed).

function is quite close to the 'true' one. This indicates that the estimation procedure takes the errors present in the real data adequately into account and that the modeling by the pile-up distortion and additive measurement errors is appropriate.

We conclude that the estimation method proposed in this paper has a satisfactory behavior in various settings and give rather good results on both synthetic and real data.

6. PROOFS

6.1. **Proof of Proposition 3.1.** We have $\|\hat{f}_m - f\|^2 = (2\pi)^{-1}\|\hat{f}_m^* - f^*\|^2 = (2\pi)^{-1}(\|\hat{f}_m^* - f_m^*\|^2 + \|f_m^* - f^*\|^2)$ since they have disjoint supports. Next, by the Parseval formula and triangular inequality, we get

$$\|\hat{f}_m^* - f_m^*\|^2 = \int_{-\pi m}^{\pi m} \frac{\mathrm{d}u}{|f_\eta^*(u)|^2} \frac{1}{n^2} \left| \sum_{k=1}^n \left[\mathrm{e}^{-iuZ_k} w_\theta \circ \hat{G}_n(Z_k) - \mathbb{E}(\mathrm{e}^{-iuZ_k} w_\theta \circ G(Z_k)) \right] \right|^2$$

$$\leq 2 \int_{-\pi m}^{\pi m} \frac{\mathrm{d}u}{|f_\eta^*(u)|^2} \frac{1}{n^2} \left| \sum_{k=1}^n \left[\mathrm{e}^{-iuZ_k} w_\theta \circ \hat{G}_n(Z_k) - \mathrm{e}^{-iuZ_k} w_\theta \circ G(Z_k) \right] \right|^2$$

$$(30) \qquad + 2 \int_{-\pi m}^{\pi m} \frac{\mathrm{d}u}{|f_\eta^*(u)|^2} \frac{1}{n^2} \left| \sum_{k=1}^n \left[\mathrm{e}^{-iuZ_k} w_\theta \circ G(Z_k) - \mathbb{E}(\mathrm{e}^{-iuZ_k} w_\theta \circ G(Z_k)) \right] \right|^2 .$$

The expectation of the first term on the right-hand side of (30) is less than or equal to

$$\frac{2}{n} \sum_{k=1}^n \int_{-\pi m}^{\pi m} \frac{\mathrm{d}u}{|f_\eta^*(u)|^2} \mathbb{E}(|w_\theta \circ \hat{G}_n(Z_k) - w_\theta \circ G(Z_k)|^2)$$

$$\leq c_{w,\theta}^2 \mathbb{E}\left(\|\hat{G}_n - G\|_\infty^2 \right) \int_{-\pi m}^{\pi m} \frac{\mathrm{d}u}{|f_\eta^*(u)|^2} \leq 2\pi c_1 c_{w,\theta}^2 \frac{\Delta_\eta(m)}{n} ,$$

by using $\mathbb{E}(\|\hat{G}_n - G\|_\infty^{2k}) \leq c_k/n^k$ (see e.g. Lemma 6.1 p. 462, Brunel and Comte (2005) which is a straightforward consequence of Massart (1990)). Here $c_2 = 2$ for $\hat{G}_n$ and $c_2 = 4$ for $\tilde{G}_n$ and more generally, c_k is a numerical constant that depends on k only. The expectation

of the second term on the right-hand side of (30) is a variance and less than or equal to

$$\frac{2}{n} \int_{-\pi m}^{\pi m} \frac{\mathrm{d}u}{|f_\eta^*(u)|^2} \mathrm{Var}(\mathrm{e}^{-iuZ_1} w_\theta \circ G(Z_1)) \le 4\pi \frac{\Delta_\eta(m)\mathbb{E}[(w_\theta \circ G(Z_1))^2]}{n} .$$

Gathering the terms completes the proof of Proposition 3.1. $\square$

6.2. **Proof of Proposition 3.2.** Equation (30) now writes

$$\begin{aligned}
\|\hat{f}_m^* - f_m^*\|^2 &\le 3 \int_{-\pi m}^{\pi m} \frac{\mathrm{d}u}{|f_\eta^*(u)|^2} \frac{1}{n^2} \left| \sum_{k=1}^{n} \left[\mathrm{e}^{-iuZ_k}(w_{\hat{\theta}} \circ \tilde{G}_n(Z_k) - w_\theta \circ \tilde{G}(Z_k)) \right] \right|^2 \\
&\quad + 3 \int_{-\pi m}^{\pi m} \frac{\mathrm{d}u}{|f_\eta^*(u)|^2} \frac{1}{n^2} \left| \sum_{k=1}^{n} \left[\mathrm{e}^{-iuZ_k} w_\theta \circ \tilde{G}_n(Z_k) - \mathrm{e}^{-iuZ_k} w_\theta \circ G(Z_k) \right] \right|^2 \\
&\quad + 3 \int_{-\pi m}^{\pi m} \frac{\mathrm{d}u}{|f_\eta^*(u)|^2} \frac{1}{n^2} \left| \sum_{k=1}^{n} \left[\mathrm{e}^{-iuZ_k} w_\theta \circ G(Z_k) - \mathbb{E}(\mathrm{e}^{-iuZ_k} w_\theta \circ G(Z_k)) \right] \right|^2 . \\
&:= 3(T_1 + T_2 + T_3).
\end{aligned}$$

The last two terms T_2 and T_3 of the right-hand-side are exactly the ones found in the proof of Proposition 3.1 and they already have been studied. Next we have

$$T_3 \le \Delta_\eta(m) \frac{1}{n} \sum_{k=1}^{n} \mathbb{E}\left[\left(w_{\hat{\theta}}\left(\frac{k}{n+1}\right) - w_\theta\left(\frac{k}{n+1}\right) \right)^2 \right] .$$

This term is split is two parts:

$$\begin{aligned}
\mathbb{E}\left[\left(w_{\hat{\theta}}\left(\frac{k}{n+1}\right) - w_\theta\left(\frac{k}{n+1}\right) \right)^2 \mathbb{1}_{\hat{\theta}\in[0,2\theta_{max}]} \right] &\le \sup_{u\in[0,1]} \sup_{\theta\in[0,2\theta_{max}]} \left| \frac{\partial w_\theta(u)}{\partial \theta} \right|^2 \mathbb{E}[(\hat{\theta} - \theta)^2] \\
&\le C_2(\theta_{max}) \frac{C_1}{n} ,
\end{aligned}$$

by using [M6] (i) for $k = 1$ and (ii). Moreover, using now [M6] (i) for $k = 2 + a$ and (iii), we get

$$\begin{aligned}
\mathbb{E}\left[\left(w_{\hat{\theta}}\left(\frac{k}{n+1}\right) - w_\theta\left(\frac{k}{n+1}\right) \right)^2 \mathbb{1}_{\hat{\theta}\ge 2\theta_{max}} \right] &\le C_3 n^{2a} \mathbb{E}\left(|\hat{\theta} - \theta|^2 \mathbb{1}_{\{|\hat{\theta}-\theta|>\theta_{max}\}} \right) \\
&\le C_3 n^{2a} \frac{\mathbb{E}[(\hat{\theta} - \theta)^{2+4a}]}{\theta_{max}^{4a}} \le \frac{C_1 C_3}{n} .
\end{aligned}$$

The two above bounds added to the ones of the proof of Proposition 3.1 implies inequality (20) and the first part of Proposition 3.2.

For the second part of the result, related to the Poisson case, if $\theta \in [0, \theta_{max}]$, then we have $\mathbb{E}[(\hat{\theta} - \theta)^{2k}] \le C_1/n^k$ for $k = 1, 2, 3$ and C_1 depends on θ, θ_{max}, which ensures (i). We skip the proof of this result which uses classical tools. Besides, for (ii), we compute

$$\frac{\partial w_\theta(u)}{\partial \theta} = \frac{-(1 - \mathrm{e}^{-\theta} - \theta\mathrm{e}^{-\theta})/\theta^2 + u(1 - \mathrm{e}^{-\theta})^2/\theta^2}{[1 - u(1 - \mathrm{e}^{-\theta})]^2} ,$$

and as $-(1 - \mathrm{e}^{-\theta} - \theta\mathrm{e}^{-\theta})/\theta^2$ takes values in $[-1/2, 0]$ and $(1 - \mathrm{e}^{-\theta})^2/\theta^2$ belongs to $[0, 1]$, we have

$$\frac{-1/2}{[1 - u(1 - \mathrm{e}^{-\theta})]^2} \le \frac{\partial w_\theta(u)}{\partial \theta} \le \frac{u}{[1 - u(1 - \mathrm{e}^{-\theta})]^2} .$$

Therefore, we get $|\partial w_\theta(u)/\partial\theta| \leq \max(u, 1/2)/[1 - u(1 - e^{-\theta})]^2$, which yields $C_2(\theta_{max}) = e^{2\theta_{max}}$ in (ii). Lastly,

$$\sup_{\theta\in\mathbb{R}^+}\left|\frac{\partial w_\theta(u)}{\partial\theta}\right| \leq \frac{1}{1-u} ,$$

and thus the sup for $u \in [0, n/(n+1)]$ is $n + 1$ so that $C_3 = 2$ and $a = 1$ in (iii) suits. $\square$

6.3. **Proof of Theorem 4.1.** We have the following decomposition of the contrast for functions s, t in S_m,

$$\tag{31} \gamma_n(t) - \gamma_n(s) = \|t - f\|^2 - \|s - f\|^2 - 2\nu_n(t - s) - 2R_n(t - s) ,$$

where

$$\tag{32} \nu_n(t) = \frac{1}{2\pi n}\sum_{k=1}^{n}\int \frac{t^*(-u)\left[e^{-iuZ_k}(w\circ G)(Z_k) - \mathbb{E}(e^{-iuZ_k}(w\circ G)(Z_k))\right]}{f_\eta^*(u)}\mathrm{d}u ,$$

and

$$\tag{33} R_n(t) = \frac{1}{2\pi n}\sum_{k=1}^{n}\int \frac{t^*(-u)e^{-iuZ_k}}{f_\eta^*(u)}\mathrm{d}u\,[(w\circ\hat{G}_n)(Z_k) - (w\circ G)(Z_k)] .$$

Now, the definition of $\hat{f}_{\hat{m}}$ implies that, $\forall m \in \mathcal{M}_n$,

$$\gamma_n(\hat{f}_{\hat{m}}) + \mathrm{pen}(\hat{m}) \leq \gamma_n(f_m) + \mathrm{pen}(m) .$$

Thus, with decomposition (31) where we take $t = \hat{f}_{\hat{m}}$ and $s = f_m$, this can be rewritten as follows

$$\|\hat{f}_{\hat{m}} - f\|^2 \leq \|f_m - f\|^2 + \mathrm{pen}(m) + 2\nu_n(\hat{f}_{\hat{m}} - f_m) - \mathrm{pen}(\hat{m}) + 2R_n(\hat{f}_{\hat{m}} - f_m) .$$

Using this and and that $2xy \leq x^2/\theta + \theta y^2$ for all nonnegative x, y, θ, we obtain

$$
\begin{aligned}
\|f - \hat{f}_{\hat{m}}\|^2 &\leq \|f - f_m\|^2 + \mathrm{pen}(m) + 2\nu_n(\hat{f}_{\hat{m}} - f_m) - \mathrm{pen}(\hat{m}) + 2R_n(\hat{f}_{\hat{m}} - f_m)\\
&\leq \|f - f_m\|^2 + \mathrm{pen}(m) + 2\|\hat{f}_{\hat{m}} - f_m\|\sup_{t\in S_{\hat{m}}+S_m, \|t\|=1}|\nu_n(t)| - \mathrm{pen}(\hat{m})\\
&\quad + 2\|\hat{f}_{\hat{m}} - f_m\|\sup_{t\in S_{\hat{m}}+S_m, \|t\|=1}|R_n(t)|\\
&\leq \|f - f_m\|^2 + \mathrm{pen}(m) + \frac{1}{4}\|\hat{f}_{\hat{m}} - f_m\|^2 + 4\sup_{t\in S_{\hat{m}}+S_m, \|t\|=1}[\nu_n(t)]^2\\
&\quad - \mathrm{pen}(\hat{m}) + \frac{1}{8}\|\hat{f}_{\hat{m}} - f_m\|^2 + 8\sup_{t\in S_{\hat{m}}+S_m, \|t\|=1}[R_n(t)]^2 .
\end{aligned}
$$

As $\|\hat{f}_{\hat{m}} - f_m\|^2 \leq 2(\|\hat{f}_{\hat{m}} - f\|^2 + \|f_m - f\|^2)$, this yields

$$
\begin{aligned}
\frac{1}{4}\mathbb{E}[\|f - \hat{f}_{\hat{m}}\|^2] &\leq \frac{7}{4}\|f - f_m\|^2 + \mathrm{pen}(m) + 4\mathbb{E}\left(\sup_{t\in B_{m,\hat{m}}}[\nu_n(t)]^2\right) - \mathbb{E}(\mathrm{pen}(\hat{m}))\\
\tag{34} &\quad + 8\mathbb{E}\left(\sup_{t\in B_{m,\hat{m}}}[R_n(t)]^2\right),
\end{aligned}
$$

where $\nu_n(t)$ and $R_n(t)$ are defined by (32) and (33) and $B_m = \{t \in S_m, \|t\| = 1\}$, and $B_{m,m'} = \{t \in S_m + S_{m'}, \|t\| = 1\}$. Following a classical application of Talagrand Inequality in the deconvolution context for ordinary smooth noise (Comte et al., 2006), we deduce the following Lemma.

Lemma 6.1. *Under the Assumptions of Theorem 4.1,*

$$\mathbb{E}\left(\sup_{t\in B_{m,\hat{m}}}[\nu_n(t)]^2 - p_1(m,\hat{m})\right)_+ \le \frac{c}{n}\,,$$

where $p_1(m,m') = 2\mathbb{E}((w_\theta \circ G)^2(Z_1))\Delta_\eta(m \vee m')/n = 2(\int_0^1 w_\theta^2(u)\mathrm{d}u)\Delta_\eta(m \vee m')/n.$

Moreover for the study $R_n(t)$ we have the following Lemma.

Lemma 6.2. *Under the assumptions of Theorem 4.1,*

$$\mathbb{E}\left(\sup_{t\in B_{m,\hat{m}}}[R_n(t)]^2 - p_2(m,\hat{m})\right) \le 0\,,$$

where $p_2(m,m') = c_{w,\theta}^2 \Delta_\eta(m \vee m')\log(n)/n.$

It follows from the definition of $p_i(m,m')$, $i = 1,2$, that there exist numerical constants κ and κ', namely $\kappa, \kappa' \ge 8$, such that $4p_1(m,m') + 8p_2(m,m') \le \mathrm{pen}(m) + \mathrm{pen}(m').$

Now, starting from (34), we get, by applying Lemmas 6.1 and 6.2,

$$\frac{1}{4}\mathbb{E}[\|f - \hat{f}_{\hat{m}}\|^2] \le \frac{7}{4}\|f - f_m\|^2 + \mathrm{pen}(m) + 4\mathbb{E}\left(\sup_{t\in B_{m,\hat{m}}}[\nu_n(t)]^2 - p_1(m,\hat{m})\right)_+$$

$$+ 8\mathbb{E}\left(\sup_{t\in B_{m,\hat{m}}}[R_n(t)]^2 - p_2(m,\hat{m})\right) + \mathbb{E}[4p_1(m,\hat{m}) + 8p_2(m,\hat{m}) - \mathrm{pen}(\hat{m})]$$

$$\le \frac{7}{4}\|f - f_m\|^2 + 2\mathrm{pen}(m) + \frac{c}{n}\,.$$

Therefore we get $(1/4)\mathbb{E}[\|f - \hat{f}_{\hat{m}}\|^2] \le (7/4)\|f - f_m\|^2 + 2\mathrm{pen}(m) + c/n.$ This completes the proof of Theorem 4.1. $\square$

Proof of Lemma 6.2. First we remark that, with Cauchy-Schwarz inequality, we have

$$|R_n(t)|^2 = \frac{1}{4\pi^2}\left|\int \frac{t^*(-u)}{f_\eta^*(u)}\left(\frac{1}{n}\sum_{k=1}^n e^{-iuZ_k}[(w_\theta \circ \hat{G}_n)(Z_k) - (w_\theta \circ G)(Z_k)]\right)\mathrm{d}u\right|^2$$

$$\le \frac{1}{4\pi^2}\int |t^*(u)|^2\mathrm{d}u\int_{-\pi(m\vee\hat{m})}^{\pi(m\vee\hat{m})} \frac{\mathrm{d}u}{|f_\eta^*(u)|^2}\left(\frac{1}{n}\sum_{k=1}^n |(w_\theta \circ \hat{G}_n)(Z_k) - (w_\theta \circ G)(Z_k)|^2\right).$$

Then Parseval Formula gives $\|t^*\|^2 = 2\pi\|t\|^2$ and we find

$$\sup_{t\in B_{m,\hat{m}}}|R_n(t)|^2 \le c_{w,\theta}^2\Delta_\eta(m\vee\hat{m})\left(\frac{1}{n}\sum_{k=1}^n |\hat{G}_n(Z_k) - G(Z_k)|^2\right) \le c_{w,\theta}^2\Delta_\eta(m\vee\hat{m})\|\hat{G}_n-\hat{G}\|_\infty^2\,.$$

We define Ω_G by

$$(35) \qquad \Omega_G = \{\sqrt{n}\|\hat{G}_n - G\|_\infty \le \sqrt{\log(n)}\}\,.$$

Now, we know from Massart (1990) that

$$(36) \qquad \mathbb{P}(\sqrt{n}\|\hat{G}_n - G\|_\infty \ge \lambda) \le 2e^{-2\lambda^2}\,.$$

This implies that $\mathbb{P}(\Omega_G^c) \le 2/n^2.$

Now, we write $\sup_{t \in B_{m,\hat{m}}} |R_n(t)|^2 = \mathcal{R}_1 + \mathcal{R}_2$ by inserting the indicator functions $\mathbb{1}_{\Omega_G}$ and $\mathbb{1}_{\Omega_G^c}$ where Ω_G is defined by (35). Therefore

$$\mathbb{E}\left(\sup_{t \in B_{m,\hat{m}}} [R_n(t)]^2 - p_2(m, \hat{m}) \right) \le \mathbb{E}(\mathcal{R}_1 - p_2(m, \hat{m})) + \mathbb{E}(\mathcal{R}_2)$$

$$(37) \qquad \le c_{w,\theta}^2 \mathbb{E}\left(\Delta_\eta(m \vee \bar{m})(\|\hat{G}_n - \hat{G}\|_\infty^2 \mathbb{1}_{\Omega_G} - \frac{\log(n)}{n}) \right)$$

$$+ c_{w,\theta}^2 \Delta_\eta(m_n) \mathbb{E}(\|\hat{G}_n - G\|_\infty^2 \mathbb{1}_{\Omega_G^c}) \,.$$

Next $(\|\hat{G}_n - \hat{G}\|_\infty^2 \mathbb{1}_{\Omega_G} - \log(n)/n)) \le 0$ by definition of Ω_G for the first right-hand-side term of (37). For the second term, $\Delta(m_n) \le n$ by the definition of m_n, $\|\hat{G}_n - G\|_\infty \le 1$ and it follows from (36) that $\mathbb{P}(\Omega_G^c) \le 2/n^2$. Therefore

$$\mathbb{E}\left(\sup_{t \in B_{m,\hat{m}}} [R_n(t)]^2 - p_2(m, \hat{m}) \right) \le c_{w,\theta}^2 n \mathbb{P}(\Omega_G^c) \le 2c_{w,\theta}^2/n \,.$$

Gathering the bounds gives the result of Lemma 6.2. □

6.4. **Proof of Theorem 4.2.** The additional assumptions are used to provide a new bound in Lemma 6.2, which is now replaced by

Lemma 6.3. *Assume that $\hat{G}_n$ is estimated with a sample $\mathcal{Z}^- = (Z_{-j})_{1 \le j \le n}$ independent of $(Z_j)_{1 \le j \le n}$ and that $\ddot{w}$ exists and is bounded. Then we have*

$$\mathbb{E}\left(\left(\sup_{t \in B_{m,\hat{m}}} [R_n(t)]^2 - \tilde{\kappa}" \int \dot{w}^2 \frac{\Delta_\eta(m \vee \hat{m})}{n} \right)_+ \right) \le \frac{c}{n} \,.$$

Proof of Lemma 6.3. By Taylor formula, we write

$$R_n^2(t) \le 2(R_{n,1}^2(t) + R_{n,2}^2(t)) \,,$$

where

$$R_{n,1}(t) = \frac{1}{n} \sum_{j=1}^n \int (\hat{G}_n(Z_j) - G(Z_j)) \dot{w}(G(Z_j)) e^{-iuZ_j} \frac{t^*(-u)}{f_\eta^*(u)} \mathrm{d}u$$

and

$$R_{n,2}(t) = \frac{1}{2n} \sum_{j=1}^n \int (\hat{G}_n(Z_j) - G(Z_j))^2 \ddot{w}(\theta_j) e^{-iuZ_j} \frac{t^*(-u)}{f_\eta^*(u)} \mathrm{d}u \,,$$

where θ_j is a random element in $(G(Z_j), \hat{G}_n(Z_j))$. Therefore,

$$\mathbb{E}\left(\sup_{t \in B_{m,\hat{m}}} [R_{n,2}(t)]^2 \right) \le \mathbb{E}\left(\sup_{t \in B_{m_n}} [R_{n,2}(t)]^2 \right) \le \|\ddot{w}\|_\infty^2 \mathbb{E}(\|\hat{G}_n - G\|_\infty^4) \Delta_\eta(m_n)$$

$$\le \|\ddot{w}\|_\infty^2 \frac{\Delta_\eta(m_n)}{n^2} \le \frac{\|\ddot{w}\|_\infty^2}{n} \,.$$

Now, the main part of the study is for $R_{n,1}(t)$. We split the process into two other processes:

$$\vartheta_{n,1}(t) = R_{n,1}(t) - \mathbb{E}(R_{n,1}(t)|\mathcal{Z}^-) \quad \text{and} \quad \vartheta_{n,2}(t) = \mathbb{E}(R_{n,1}(t)|\mathcal{Z}^-) \,.$$

We first study $\vartheta_{n,1}$, which is a centered empirical process given $\mathcal{Z}^-$ and which is studied given $\mathcal{Z}^-$. We apply a Talagrand inequality to this process after the transformation

$$\mathbb{E}^- \left(\left(\sup_{t \in B_{m,\hat{m}}} [R_{n,1}(t)]^2 - \tilde{\kappa}_1 \int \dot{w}^2 \frac{\Delta_\eta(m \vee \hat{m})}{n} \right)_+ \right)$$

$$(38) \qquad \leq \sum_{m' \in \mathcal{M}_n} \mathbb{E}^- \left(\left(\sup_{t \in B_{m,m'}} [R_{n,1}(t)]^2 - \tilde{\kappa}_1 \int \dot{w}^2 \frac{\Delta_\eta(m \vee m')}{n} \right)_+ \right).$$

Now, we apply Talagrand Inequality as recalled in Appendix. We have therefore three quantities to bound, denoted by H_1, b_1, v_1 with obvious reference to Theorem 7.1. We denote by $\mathbb{E}^-(.) = \mathbb{E}(.|\mathcal{Z}^-)$.

$$\mathbb{E}^- \left(\sup_{t \in B_{m,m'}} \vartheta_{n,1}^2(t) \right)$$

$$\leq \mathbb{E}^- \left(\int \left| \frac{1}{n} \sum_{j=1}^n [(\hat{G}_n - G)(Z_j)\dot{w} \circ G(Z_j)e^{-iuZ_j} - \mathbb{E}((\hat{G}_n - G)(Z_1)\dot{w} \circ G(Z_1)e^{-iuZ_1})] \right|^2 \frac{\mathbb{1}_{|u| \leq \pi(m \vee m')}}{|f_\eta^*(u)|^2} du \right)$$

$$= \frac{1}{n} \int \mathbb{1}_{|u| \leq \pi(m \vee m')} \text{Var}^- \left((\hat{G}_n(Z_1) - G(Z_1))\dot{w} \circ G(Z_1)e^{-iuZ_1} \right) \frac{du}{|f_\eta^*(u)|^2}$$

$$\leq \frac{1}{n} \int \mathbb{1}_{|u| \leq \pi(m \vee m')} \mathbb{E}^- \left((\dot{w} \circ G(Z_1))^2 \right) \frac{du}{|f_\eta^*(u)|^2} du = \frac{\int (\dot{w})^2 \Delta_\eta(m \vee m')}{n} := H_1^2,$$

since $(\hat{G}_n(z) - G(z))^2 \leq 1$.

Next it is rather straightforward that $b_1 = \|\dot{w}\|_\infty \sqrt{\Delta_\eta(m \vee m')}$ gives a second bound.

Lastly, the most tedious term is v_1.

$$\text{Var}^- \left(\int (\hat{G}_n(Z_1) - G(Z_1))\dot{w} \circ G(Z_1)e^{-iuZ_1} \frac{t^*(-u)}{f_\eta^*(u)} du \right)$$

$$\leq \mathbb{E}^- \left((\hat{G}_n(Z_1) - G(Z_1))^2 (\dot{w} \circ G(Z_1))^2 \left| \int e^{iuZ_1} \frac{t^*(u)}{f_\eta^*(u)} du \right|^2 \right)$$

$$\leq \|\dot{w}\|_\infty^2 \mathbb{E} \left(\left| \int e^{-iuZ_1} \frac{t^*(-u)}{f_\eta^*(u)} du \right|^2 \right)$$

$$\leq \|\dot{w}\|_\infty^2 \iint g^*(u-v) \frac{t^*(u)\overline{t^*(v)}}{f_\eta^*(u)\overline{f_\eta^*(v)}} du dv$$

$$\leq \|\dot{w}\|_\infty^2 \left(\iint |t^*(u)t^*(v)|^2 du dv \iint \mathbb{1}_{|u| \leq \pi(m \vee m')} \mathbb{1}_{|v| \leq \pi(m \vee m')} \frac{1}{|f_\eta^*(u)f_\eta^*(v)|^2} |g^*(u-v)| du dv \right)^{1/2}$$

$$\leq \|\dot{w}\|_\infty^2 \sup_{|v| \leq \pi(m \vee m')} \frac{1}{|f_\eta^*(v)|} \sqrt{\Delta_\eta(\pi(m \vee m'))} \int |g^*|$$

$$\leq \|\dot{w}\|_\infty^2 (\int |g^*|)^{1/2} \sqrt{\pi(m \vee m')} \Delta_\eta(\pi(m \vee m')) := v_1,$$

where the term $\sqrt{\pi(m \vee m')}$ is obtained thanks to the OS assumption stated in (23) which implies that there are two constants c_1, c_2 such that

$$\sup_{|v| \leq \pi m} \frac{1}{|f_\eta^*(v)|} \leq c_1 (1 + (\pi m)^2)^{\gamma/2} \quad \text{and} \quad \Delta_\eta(m) \geq c_2 (1 + (\pi m)^2)^{\gamma + 1/2} .$$

Now we use (38) and the Talagrand inequality recalled in Lemma 7.1, and we get

$$\mathbb{E}^- \left(\left(\sup_{t \in B_{m,\hat{m}}} \vartheta_{n,1}^2(t) - 4 \int (\dot{w})^2 \frac{\Delta_\eta(m \vee \hat{m})}{n} \right)_+ \right)$$

$$\leq \quad \frac{C_1 \|w'\|_\infty^2}{n} \sum_{m' \in \mathcal{M}_n} \left(\Delta_\eta(m \vee m') \exp \left(-c_2 \frac{\|\dot{w}\|^2}{\|\dot{w}\|_\infty^2} \sqrt{\pi(m \vee m')} \right) + \frac{\Delta(m \vee m')}{n} \exp \left(-c_3 \frac{\|\dot{w}\|}{\|\dot{w}\|_\infty} \sqrt{n} \right) \right) .$$

The bound is of order $1/n$ since we have

$$\sum_{m \in \mathcal{M}_n} \Delta_\eta(m) \exp(-c m^{1/2}) \leq \Sigma < +\infty .$$

Lastly, as nothing in the bound is depending on $\mathcal{Z}^-$, taking the expectation w.r.t. the global distribution of the Z_{-j}'s gives the same result with a usual expectation ($\mathbb{E}^-(.)$ replaced by $\mathbb{E}$).

Next, we study the second process and we write

$$\vartheta_{n,2}(t) = \mathbb{E}(R_{n,1}(t)|\mathcal{Z}^-) = \int \left(\int \hat{G}_n(z) - G(z)) \dot{w}(G(z)) e^{-iuz} g(z) dz \right) \frac{t^*(-u)}{f_\eta^*(u)} du$$

$$= \frac{1}{n} \sum_{j=1}^n \int \left(\int (\mathbb{1}_{(Z_{-j} \leq z)} - G(z)) \dot{w}(G(z)) e^{-iuz} g(z) dz \right) \frac{t^*(-u)}{f_\eta^*(u)} du .$$

We can see that $\vartheta_{n,2}(t)$ is another centered empirical process, to which we also apply the Talagrand inequality. We denote here by H_2, b_2 and v_2 the bounds that we have to exhibit for applying the inequality.

First we find that

$$\mathbb{E} \left(\sup_{t \in B_{m,m'}} \vartheta_{n,2}^2(t) \right)$$

$$\leq \quad \mathbb{E} \left[\int \left| \frac{1}{n} \sum_{j=1}^n \left(\int (\mathbb{1}_{(Z_{-j} \leq z)} - G(z)) \dot{w}(G(z)) e^{-iuz} g(z) dz \right) \right|^2 \frac{\mathbb{1}_{|u| \leq \pi(m \vee m')}}{|f_\eta^*(u)|^2} du \right]$$

$$\leq \quad \frac{1}{n} \int \text{Var} \left(\int (\mathbb{1}_{(Z_{-1} \leq z)} \dot{w}(G(z)) e^{-iuz} g(z) dz \right) \frac{\mathbb{1}_{|u| \leq \pi(m \vee m')'}}{|f_\eta^*(u)|^2} du$$

$$\leq \quad \frac{1}{n} \int_0^1 (\dot{w}(x))^2 dx \Delta_\eta(m \vee m') := H_2^2 ,$$

since $(\int_0^1 |\dot{w}|(x) dx)^2 \leq \int_0^1 \dot{w}^2(x) dx$. Clearly, $b_2 = \|\dot{w}\| \sqrt{\Delta_\eta(m \vee m')}$ suits.

Again, the search for v_2 is more difficult.

$$\mathrm{Var}\left(\int\left(\int(\mathbb{1}_{(Z_{-1}\leq z)}-G(z))\dot{w}(G(z))\mathrm{e}^{-iuz}g(z)\mathrm{d}z\right)\frac{t^*(-u)}{f_\eta^*(u)}\mathrm{d}u\right)$$

$$\leq \mathbb{E}\left[\left|\int\left(\int\mathbb{1}_{(Z_{-1}\leq z)}\dot{w}(G(z))\mathrm{e}^{-iuz}g(z)\mathrm{d}z\right)\frac{t^*(-u)}{f_\eta^*(u)}\mathrm{d}u\right|^2\right]$$

$$\leq \mathbb{E}\left[\iint\left(\iint\mathbb{1}_{(Z_{-1}\leq z)}\dot{w}\circ G(z)\mathrm{e}^{-iuz}g(z)\mathbb{1}_{(Z_{-1}\leq t)}\dot{w}\circ G(t)\mathrm{e}^{ivt}g(t)\mathrm{d}z\mathrm{d}t\right)\frac{t^*(-u)\overline{t^*(-v)}}{f_\eta^*(u)\overline{f_\eta^*(v)}}\mathrm{d}u\mathrm{d}v\right]$$

$$= \iint\iint G(z\wedge t)\dot{w}\circ G(z)\dot{w}\circ G(t)\mathrm{e}^{-i(uz-vt)}g(z)g(t)\mathrm{d}z\mathrm{d}t\frac{t^*(-u)\overline{t^*(-v)}}{f_\eta^*(u)\overline{f_\eta^*(v)}}\mathrm{d}u\mathrm{d}v$$

$$= \iint\theta^*(u,-v)\frac{t^*(-u)\overline{t^*(-v)}}{f_\eta^*(u)\overline{f_\eta^*(v)}}\mathrm{d}u\mathrm{d}v ,$$

where

$$\theta^*(u,-v) = \iint G(z\wedge t)\dot{w}\circ G(z)\dot{w}\circ G(t)\mathrm{e}^{-i(uz-vt)}g(z)g(t)\mathrm{d}z\mathrm{d}t .$$

Therefore we obtain the bound

$$\mathrm{Var}\left(\int\left(\int(\mathbb{1}_{(Z_{-1}\leq z)}-G(z))\dot{w}\circ G(z)\mathrm{e}^{-iuz}g(z)\mathrm{d}z\right)\frac{t^*(-u)}{f_\eta^*(u)}\mathrm{d}u\right)$$

$$\leq \sup_{|u|\leq\pi(m\vee m')}\frac{1}{|f_\eta^*(u)|^2}\left(\iint|\theta^*(u,-v)|^2\mathrm{d}u\mathrm{d}v\iint|t^*(-u)t^*(-v)|^2\mathrm{d}u\mathrm{d}v\right)^{1/2}$$

$$\leq \pi(m\vee m')\Delta_\eta(m\vee m')\|\theta\|^2 ,$$

by using the OS assumption (23) as previously and since $\|\theta^*\|^2 = 2\pi\|\theta\|^2$. But we have

$$\|\theta\|^2 = \iint[G(z\wedge t)\dot{w}\circ G(z)\dot{w}\circ G(t)g(z)g(t)]^2\mathrm{d}z\mathrm{d}t \leq \left(\int|\dot{w}|(x)\mathrm{d}x\right)^2 \leq \|\dot{w}\|^2 .$$

Thus

$$v_2 = c\|\dot{w}\|^2(m\vee m')\Delta(m\vee m') .$$

Therefore, we can apply Talagrand's inequality as previously and we get

$$\mathbb{E}\left(\left(\sup_{t\in B_{m,\hat{m}}}\vartheta_{n,2}^2(t) - 4\int(\dot{w})^2\frac{\Delta_\eta(m\vee m')}{n}\right)_+\right) \leq C/n ,$$

since

$$\sum_{m\in\mathcal{M}_n}\Delta_\eta(m)\exp(-cm) \leq \Sigma < +\infty .$$

6.5. **Proof of Theorem 4.3.** We start with decomposition (31) again, and following the same line as in the proof of Theorem 4.1 relation (34) is replaced by

$$\frac{1}{4}\mathbb{E}[\|f-\hat{f}_{\hat{m},\hat{\theta}}\|^2] \leq \frac{7}{4}\|f-f_m\|^2 + \mathbb{E}(\widehat{\mathrm{pen}}(m,\hat{\theta})) + 4\mathbb{E}\left(\sup_{t\in B_{m,\hat{m}}}[\nu_n(t)]^2\right) - \mathbb{E}(\widehat{\mathrm{pen}}(\hat{m},\hat{\theta}))$$

$$(39) \qquad\qquad +8\mathbb{E}\left(\sup_{t\in B_{m,\hat{m}}}[T_n(t)]^2\right),$$

where $\nu_n(t)$ is defined by (32) and $T_n(t) = R_n(t) + U_n(t)$ with $R_n(t)$ defined by (33) and

$$U_n(t) = \frac{1}{2\pi n}\sum_{k=1}^n \int \frac{t^*(-u)\mathrm{e}^{-iuZ_k}}{f_\eta^*(u)}\mathrm{d}u\,[(w_{\hat\theta}\circ\hat G_n)(Z_k) - (w_\theta\circ G)(Z_k)]\,.$$

Next, we note that if $\theta\in[0,\theta_{max}]$ and $\hat\theta\in[0,2\theta_{max}]$,

$$\int_0^1 (w_{\hat\theta}(u) - w_\theta(u))^2\mathrm{d}u \le C(\theta_{max})(\hat\theta - \theta)^2$$

and also $(c_{w,\hat\theta} - c_{w,\theta})^2 \le C(\theta_{max})(\hat\theta - \theta)^2$. Therefore, we denote by

$$\Omega_\theta = \{\omega\in\Omega, \hat\theta(\omega)\le 2\theta_{max}\} = \{\hat\theta \le 2\theta_{max}\}\,.$$

Using the assumption on the collection of models ensuring that $\Delta_\eta(m)/n \le 1$, we obtain that uniformly in m,

$$\begin{aligned}
\mathbb{E}\left(\widehat{\mathrm{pen}}(m,\hat\theta)\mathbb{1}_{\Omega_\theta}\right) &\le& \frac{C(\theta,\theta_{max})}{n} + 2\mathbb{E}\left(\widehat{\mathrm{pen}}(m,\theta)\mathbb{1}_{\Omega_\theta}\right)\,,\\
\mathbb{E}\left(\widehat{\mathrm{pen}}(m,\theta)\mathbb{1}_{\Omega_\theta}\right) &\le& \frac{C(\theta,\theta_{max})}{n} + 2\mathbb{E}\left(\widehat{\mathrm{pen}}(m,\hat\theta)\mathbb{1}_{\Omega_\theta}\right)\,,
\end{aligned}$$

(40)

by using that $\mathbb{E}[|\hat\theta - \theta|^{2k}] \le c/n^k$.

Now, we add to Lemmas 6.1 and 6.2, the following result for the study $U_n(t)$.

Lemma 6.4. *Under the assumptions of Theorem 4.1, there exists a numerical constant κ' such that*

$$\mathbb{E}\left(\sup_{t\in B_{m,\hat m}}[U_n(t)\mathbb{1}_{\Omega_\theta}]^2 - p_3(m,\hat m)\mathbb{1}_{\Omega_\theta}\right) \le \frac{c}{n}\,,$$

where

$$(41)\qquad p_3(m,m') = \kappa' C_2(\theta_{max})\mathrm{e}^{\theta_{max}}\frac{\Delta_\eta(m\vee m')}{n}\log(n)\,.$$

It follows from the definition of $p_i(m,m')$, $i = 1,2,3$, that there exist numerical constants κ and κ', namely $\kappa,\kappa'\ge 16$, such that

$$4p_1(m,m') + 16p_2(m,m') + 16p_3(m,m') \le \frac{1}{2}\widehat{\mathrm{pen}}(m,\theta) + \frac{1}{2}\widehat{\mathrm{pen}}(m',\theta)\,.$$

Note that p_1 and p_2 depend also on θ. Now, starting from (39), we get, by applying Lemmas 6.1, 6.2, and 6.4, that on Ω_θ, the following inequalities hold.

$$\begin{aligned}
\frac{1}{4}\mathbb{E}[\|f - \hat f_{\hat m,\hat\theta}\|^2\mathbb{1}_{\Omega_\theta}] &\le \frac{7}{4}\|f - f_m\|^2 + \frac{5}{2}\widehat{\mathrm{pen}}(m,\theta) + 4\mathbb{E}\left(\sup_{t\in B_{m,\hat m}}[\nu_n(t)]^2 - p_1(m,\hat m)\right)_+\\
&\quad + 16\mathbb{E}\left(\sup_{t\in B_{m,\hat m}}[R_n(t)]^2 - p_3(m,\hat m)\right) + 16\mathbb{E}\left(\sup_{t\in B_{m,\hat m}}[U_n(t)\mathbb{1}_{\Omega_\theta}]^2 - p_3(m,\hat m)\mathbb{1}_{\Omega_\theta}\right)\\
&\quad + \mathbb{E}[(4p_1(m,\hat m) + 16p_2(m,\hat m) + 16p_3(m,\hat m) - \frac{1}{2}\widehat{\mathrm{pen}}(\hat m,\theta))\mathbb{1}_{\Omega_\theta}]\\
&\quad + \mathbb{E}[(\frac{1}{2}\widehat{\mathrm{pen}}(\hat m,\theta) - \widehat{\mathrm{pen}}(\hat m,\hat\theta))\mathbb{1}_{\Omega_\theta}] + \frac{C(\theta,\theta_{max})}{n}\\
&\le \frac{7}{4}\|f - f_m\|^2 + \frac{5}{2}\widehat{\mathrm{pen}}(m,\theta) + \frac{c}{n}\,,
\end{aligned}$$

by using (40). Therefore we get

$$(42) \qquad \mathbb{E}[\|f - \hat{f}_{\hat{m}}\|^2 \mathbb{1}_{\Omega_\theta}] \le 7\|f - f_m\|^2 + 10\widehat{\text{pen}}(m, \theta) + c/n .$$

On the other hand, we have $\|\hat{f}_{\hat{m},\hat{\theta}} - f\|^2 \le 2\|\hat{f}_{\hat{m},\hat{\theta}}\|^2 + 2\|f\|^2$ and

$$\|\hat{f}_{\hat{m},\hat{\theta}}\|^2 \le \Delta_\eta(m \vee \hat{m}) \left(\frac{1}{n} \sum_{k=1}^{n} w_{\hat{\theta}} \left(\frac{k}{n+1} \right) \right)^2 .$$

We know that $\Delta_\eta(m \vee \hat{m}) \le n$ by definition of $\mathcal{M}_n$ and

$$0 \le \frac{1}{n} \sum_{k=1}^{n} w_{\hat{\theta}} \left(\frac{k}{n+1} \right) \le \int_{\frac{1}{n}}^{1} w_{\hat{\theta}}(u)\mathrm{d}u \le \int_{0}^{1} w_{\hat{\theta}}(u)\mathrm{d}u = 1.$$

Therefore $\|\hat{f}_{\hat{m},\hat{\theta}}\|^2 \le n$ and

$$\mathbb{E}(\|\hat{f}_{\hat{m},\hat{\theta}} - f\|^2 \mathbb{1}_{\Omega_\theta^c}) \le 2(n + \|f\|^2)\mathbb{P}(|\hat{\theta} - \theta| > \theta_{max})$$

$$(43) \qquad \le 2(n + \|f\|^2)\frac{\mathbb{E}(|\hat{\theta} - \theta|^4)}{\theta_{max}^4} \le \frac{C(\theta_{max}, \|f\|)}{n}.$$

Gathering (42) and (43) completes the proof of Theorem 4.3. $\square$

Proof of Lemma 6.4.
We follow the same line as in the proof of Lemma 6.2 and we obtain similarly:

$$\sup_{t \in B_{m,\hat{m}}} |U_n(t)|^2 \mathbb{1}_{\Omega_\theta} \le \Delta_\eta(m \vee \hat{m}) \frac{1}{n} \sum_{k=1}^{n} \left| w_{\hat{\theta}} \left(\frac{k}{n+1} \right) - w_\theta \left(\frac{k}{n+1} \right) \right|^2 \mathbb{1}_{\Omega_\theta}$$

$$\le C_2(\theta_{max})\Delta_\eta(m \vee \hat{m})(\hat{\theta} - \theta)^2.$$

We define Ω_U by

$$\Omega_U = \{|\hat{\theta} - \theta| \le u_{n,\theta}\} , \qquad \text{with } u_{n,\theta} = \kappa' \frac{v_\theta \sqrt{\log(n)}}{\mathrm{e}^{-\theta}\sqrt{n}} .$$

We obtain $\sup_{t \in B_{m,\hat{m}}} |U_n(t)|^2 \mathbb{1}_{\Omega_\theta} \le (\mathcal{U}_1 + \mathcal{U}_2)\mathbb{1}_{\Omega_\theta}$ by inserting in the above bound $\mathbb{1}_{\Omega_U}$ and $\mathbb{1}_{\Omega_U^c}$ so that

$$\mathbb{E} \left((\sup_{t \in B_{m,\hat{m}}} |U_n(t)|^2 - p_3(m, \hat{m}))\mathbb{1}_{\Omega_\theta} \right) \le \mathbb{E}((\mathcal{U}_1 - p_3(m, \hat{m}))\mathbb{1}_{\Omega_\theta}) + \mathbb{E}(\mathcal{U}_2\mathbb{1}_{\Omega_\theta})$$

$$\le 0 + nC_2(\theta_{max})\mathbb{E}((\hat{\theta} - \theta)^2 \mathbb{1}_{\Omega_\theta}\mathbb{1}_{\Omega_U^C})$$

$$(44) \qquad \le C_2(\theta_{max})\sqrt{C_1}\mathbb{P}^{1/2}(\Omega_U^c) ,$$

for $p_3(m, m')$ as in (41) and $\kappa' \ge 1$.

Since $\forall x \geq 0$, $e^x \geq 1 + x$ and $e^{-x} \geq 1 - x$, by denoting $\delta_k = \mathbb{1}(Z_k = +\infty)$, we get

$$
\begin{aligned}
\mathbb{P}(\Omega_U^c) &= \mathbb{P}\left(\left|\log\left(\frac{1}{n+1} + \frac{1}{n+1}\sum_{k=1}^n \delta_k\right) - \log\left(\mathbb{E}(\delta_1)\right)\right| > u_{n,\theta}\right) \\
&= \mathbb{P}\left(\frac{\frac{1}{n+1} + \frac{1}{n+1}\sum_{k=1}^n \delta_k}{\mathbb{E}(\delta_1)} \geq e^{u_{n,\theta}} \text{ or } \frac{\frac{1}{n+1} + \frac{1}{n+1}\sum_{k=1}^n \delta_k}{\mathbb{E}(\delta_1)} \leq e^{-u_{n,\theta}}\right) \\
&\leq \mathbb{P}\left(\left|\frac{\frac{1}{n+1} + \frac{1}{n+1}\sum_{k=1}^n \delta_k}{\mathbb{E}(\delta_1)} - 1\right| \geq u_{n,\theta}\right) \\
&= \mathbb{P}\left(\left|\frac{1}{n+1}(1 - \mathbb{E}(\delta_1)) + \frac{1}{n+1}\sum_{k=1}^n(\delta_k - \mathbb{E}(\delta_k))\right| \geq u_{n,\theta}\mathbb{E}(\delta_1)\right) \\
&\leq \mathbb{P}\left(\left|\frac{1}{n+1}\sum_{k=1}^n(\delta_k - \mathbb{E}(\delta_k))\right| \geq u_{n,\theta}\mathbb{E}(\delta_1) + \frac{2}{n}\right) \\
&\leq \mathbb{P}\left(\left|\frac{1}{n}\sum_{k=1}^n(\delta_k - \mathbb{E}(\delta_k))\right| \geq u_{n,\theta}\mathbb{E}(\delta_1)\right).
\end{aligned}
$$

The Bernstein inequality ensures now that for well chosen κ', we have the above probability less than C/n^2 which, inserted in (44) completes the proof of Lemma 6.4. $\square$

7. Appendix

The Talagrand inequality. The following result follows from the Talagrand concentration inequality given in Klein and Rio (2005) and arguments in Birgé and Massart (1998) (see the proof of their Corollary 2 page 354).

Lemma 7.1. *(Talagrand Inequality) Let $Y_1, \ldots, Y_n$ be independent random variables, let $\nu_{n,Y}(f) = (1/n)\sum_{i=1}^n [f(Y_i) - \mathbb{E}(f(Y_i))]$ and let $\mathcal{F}$ be a countable class of uniformly bounded measurable functions. Then for $\epsilon^2 > 0$*

$$
\mathbb{E}\left[\sup_{f\in\mathcal{F}} |\nu_{n,Y}(f)|^2 - 2(1 + 2\epsilon^2)H^2\right]_+ \leq \frac{4}{K_1}\left(\frac{v}{n}e^{-K_1\epsilon^2\frac{nH^2}{v}} + \frac{98b^2}{K_1 n^2 C^2(\epsilon^2)}e^{-\frac{2K_1 C(\epsilon^2)\epsilon}{7\sqrt{2}}\frac{nH}{b}}\right),
$$

with $C(\epsilon^2) = \sqrt{1 + \epsilon^2} - 1$, $K_1 = 1/6$, and

$$
\sup_{f\in\mathcal{F}} \|f\|_\infty \leq b, \quad \mathbb{E}\left[\sup_{f\in\mathcal{F}} |\nu_{n,Y}(f)|\right] \leq H, \quad \sup_{f\in\mathcal{F}} \frac{1}{n}\sum_{k=1}^n \mathrm{Var}(f(Y_k)) \leq v.
$$

By standard denseness arguments, this result can be extended to the case where $\mathcal{F}$ is a unit ball of a linear normed space, after checking that $f \mapsto \nu_n(f)$ is continuous and $\mathcal{F}$ contains a countable dense family.

References

Berberan-Santos, M. N., Bodunov, E. N., and Valeur, B. (2005a), "Mathematical functions for the analysis of luminescence decays with underlying distributions 1. Kohlrausch decay function (stretched exponential)," *Chemical Physics*, 515, 171–182.

Berberan-Santos, M. N., Bodunov, E. N., and Valeur, B. (2005b), "Mathematical functions for the analysis of luminescence decays with underlying distributions: 2. Becquerel (compressed hyperbola) and related decay functions," *Chemical Physics*, 317, 57–62.

Birgé, L. and Massart, P. (1998), "Minimum contrast estimators on sieves: exponential bounds and rates of convergence," *Bernoulli*, 4, 329–375.

Bissantz, N., Holzmann, H., and Munk, A. (2005), "Testing parametric assumptions on band- or time-limited signals under noise," *IEEE Trans. Inform. Theory*, 51, 3796–3805.

Bissantz, N., Dümbgen, L., Holzmann, H., and Munk, A. (2007), "Non-parametric confidence bands in deconvolution density estimation," *J. R. Stat. Soc. Ser. B Stat. Methodol.*, 69, 483–506.

Brunel, E. and Comte, F. (2005), "Penalized contrast estimation of density and hazard rate with censored data," *Sankhya*, 67, 441–475.

Brunel, E., Comte, F., and Guilloux, A. (2009), "Nonparametric density estimation in presence of bias and censoring," *Test*, 18, 166–194.

Comte, F. and Genon-Catalot, V. (2010), "Nonparametric adaptive estimation for pure jump Lévy processes," *Ann. Inst. Henri Poincaré Probab. Stat.*, 46, 595–617.

Comte, F. and Lacour, C. (2011), "Data-driven density estimation in the presence of additive noise with unknown distribution," *Journal of the Royal Statistical Society B*, 73, 601–627.

Comte, F., Rozenholc, Y., and Taupin, M.-L. (2006), "Penalized contrast estimator for adaptive density deconvolution," *Canadian Journal of Statistics*, 34, 431–452.

Delaigle, A. and Gijbels, I. (2004), "Practical bandwidth selection in deconvolution kernel density estimation," *Comput. Statist. Data Anal.*, 45, 249–267.

Delaigle, A. and Meister, A. (2011), "Nonparametric function estimation under Fourier-oscillating noise," *Statist. Sinica*, 21, 1065–1092.

Diggle, P. J. and Hall, P. (1993), "A Fourier approach to nonparametric deconvolution of a density estimate," *J. Roy. Statist. Soc. Ser. B*, 55, 523–531.

Fan, J. (1991), "On the optimal rates of convergence for nonparametric deconvolution problems," *Ann. Statist.*, 19, 1257–1272.

Johnstone, I. M., Kerkyacharian, G., Picard, D., and Raimondo, M. (2004), "Wavelet deconvolution in a periodic setting," *J. R. Stat. Soc. Ser. B Stat. Methodol.*, 66, 547–573.

Klein, T. and Rio, E. (2005), "Concentration around the mean for maxima of empirical processes," *Annals of Probability*, 33, 1060–1077.

Lakowicz, J. R. (1999), *Principles of Fluorescence Spectroscopy*, Academic/Plenum, New York.

Massart, P. (1990), "The tight constant in the Dvoretzky-Kiefer-Wolfowitz inequality," *Annals of Probability*, 18, 1269–1283.

Meyer, Y. (1990), *Ondelettes et opérateurs*, Hermann, Paris.

O'Connor, D. V. and Phillips, D. (1984), *Time-correlated single photon counting*, Academic Press, London.

Patting, M., Wahl, M., Kapusta, P., and Erdmann, R. (2007), "Dead-time effects in TCSPC data analysis," in *Proceedings of SPIE*, vol. 6583.

Pensky, M. and Vidakovic, B. (1999), "Adaptive wavelet estimator for nonparametric density deconvolution," *Ann. Statist.*, 27, 2033–2053.

Rebafka, T., Roueff, F., and Souloumiac, A. (2010), "A corrected likelihood approach for the pile-up model with application to fluorescence lifetime measurements using exponential mixtures," *The International Journal of Biostatistics*, 6.

Rebafka, T., Roueff, F., and Souloumiac, A. (2011), "Information bounds and MCMC parameter estimation for the pile-up model," *Journal of Statistical Planning and Inference*,

141, 1–16.

Tsodikov, A. (2003), "A Generalized Self-Consistency Approach," *Journal of the Royal Statistical Society. Series B (Statistical Methodology)*, 65, 759–774.

Valeur, B. (2002), *Molecular Fluorescence*, Wiley-VCH, Weinheim.

van Es, B., Gugushvili, S., and Spreij, P. (2007), "A kernel type nonparametric density estimator for decompounding," *Bernoulli*, 13, 672–694.