Spectral cut-off regularisation for density estimation under multiplicative measurement errors

Sergio Brenner Miguel[*]

Ruprecht-Karls-Universität Heidelberg
e-mail: brennermiguel@math.uni-heidelberg.de

Fabienne Comte[†]

Université de Paris
e-mail: fabienne.comte@parisdescartes.fr

Jan Johannes[*]

Ruprecht-Karls-Universität Heidelberg
e-mail: johannes@math.uni-heidelberg.de

Abstract: We study the non-parametric estimation of an unknown density f with support on $\mathbb{R}_+$ based on an i.i.d. sample with multiplicative measurement errors. The proposed fully-data driven procedure is based on the estimation of the Mellin transform of the density f, a regularisation of the inverse of the Mellin transform by a spectral cut-off and a data-driven model selection in order to deal with the upcoming bias-variance trade-off. We introduce and discuss further *Mellin-Sobolev spaces* which characterize the regularity of the unknown density f through the decay of its Mellin transform. Additionally, we show minimax-optimality over *Mellin-Sobolev spaces* of the data-driven density estimator and hence its adaptivity.

Contents

[*]Institut für Angewandte Mathematik, Im Neuenheimer Feld 205, D-69120 Heidelberg, Germany

[†]CNRS, MAP5 UMR 8145 F-75006 Paris, France

1. Introduction

In this paper we are interested in estimating the unknown density $f : \mathbb{R}_+ \to \mathbb{R}_+$ of a positive random variable X given independent and identically distributed (i.i.d.) copies of $Y = XU$, where X and U are independent of each other and U has a known density $g : \mathbb{R}_+ \to \mathbb{R}_+$. In this setting the density $f_Y : \mathbb{R}_+ \to \mathbb{R}_+$ of Y is given by

$$f_Y(y) = [f * g](y) := \int_0^\infty f(x)g(y/x)x^{-1}dx \quad \forall y \in \mathbb{R}_+$$

such that $*$ denotes multiplicative convolution. The estimation of f using an i.i.d. sample $Y_1, \dots, Y_n$ from f_Y is thus an inverse problem called multiplicative deconvolution.

[16] and [17] introduce and study intensively *multiplicative censoring*, which corresponds to the particular multiplicative deconvolution problem with multiplicative error U uniformly distributed on $[0, 1]$. This model is often applied in survival analysis as explained and motivated in [15]. The estimation of the cumulative distribution function of X is considered in [17] and [2]. Series expansion methods are studied in [1] treating the model as an inverse problem. The density estimation in a multiplicative censoring model is considered in [7] using a kernel estimator and a convolution power kernel estimator. Assuming a uniform error distribution on an interval $[1 - \alpha, 1 + \alpha]$ for $\alpha \in (0, 1)$ [9] analyze a projection density estimator with respect to the Laguerre basis. [4] study a beta-distributed error U.

In this work, covering all those three variations of the multiplicative censoring model we consider a density estimator using the Mellin transform and a spectral cut-off regularization of its inverse, which borrows ideas from [5]. The key to the analysis of the multiplicative deconvolution problem is the convolution theorem of the Mellin transform $\mathcal{M}$, which roughly states $\mathcal{M}[f_Y] = \mathcal{M}[f]\mathcal{M}[g]$ for a density $f_Y = f * g$. Exploiting the convolution theorem [5] introduce a kernel density estimator of f allowing more generally X and U to be real-valued. Moreover, they point out that the following widely used naive approach is a special case of their estimation strategy. Transforming the data by applying the logarithm the model $Y = XU$ writes $\log(Y) = \log(X) + \log(U)$. In other words, multiplicative convolution becomes convolution for the log-transformed data. As a con-

sequence, the density of $\log(X)$ is eventually estimated employing usual strategies for non-parametric deconvolution problems (see for example [12]) and then transformed back to an estimator of f. However, it is difficult to interpret regularity conditions on the density of $\log(X)$. Furthermore, the analysis of a global risk of an estimator using this naive approach is challenging as [9] pointed out.

Our strategy differs in the following way. Making use of the convolution theorem of the Mellin transform and applying an additional spectral cut-off on the inversion of the Mellin-transform we define a density estimator. We measure the accuracy of the estimator by introducing a global risk in terms of a weighted $\mathbb{L}^2_{\mathbb{R}_+}$-norm. Exploiting properties of the Mellin transform we characterize the underlying inverse problem and natural regularity conditions which borrow ideas from the inverse problems community ([10]). The regularity conditions expressed in the form of *Mellin-Sobolev spaces* and their relations to the analytical properties of the density f are discussed in more details. The proposed estimator, however, involves a tuning parameter which is selected by a data-driven method. We establish an oracle inequality for the fully-data driven spectral cut-off estimator under fairly mild assumptions on the error density g. Moreover we show that uniformly over *Mellin-Sobolev spaces* the proposed data-driven estimator is minimax-optimal. Precisely, we state both an upper bound for the mean weighted integrated squared error of the fully-data driven spectral cut-off estimator and a general lower bound for estimating the density f based on i.i.d. copies from $f_Y = f * g$.

The paper is organized as follows. In section 2 we collect properties of the Mellin transform. We explain our general estimation strategy by first introducing and analyzing an estimator based on direct observations $X_1, \ldots, X_n$ from f. The estimator relies on an inversion of the Mellin transform which we stabilize using a spectral cut-off. Exploiting then the multiplication theorem of the Mellin-transform we propose a fully-data driven estimator of f based on the sample $Y_1, \ldots, Y_n$. We derive an oracle type upper bound for its mean weighted integrated squared error. We finish the theoretical part by showing in section 3 that our fully-data driven estimator is minimax optimal over *Mellin-Sobolev spaces* for a large class of error densities g. Finally, results of a simulation study are reported in section 4 which visualize the reasonable finite sample performance of our estimators. The proof of section 2 and section 3 are postponed to the appendix.

2. Adaptive weighted $\mathbb{L}^2_{\mathbb{R}_+}$ estimation

In this section we introduce the Mellin transform and collect some of its properties. For a more detailed introduction we refer to [13] and [3].

2.1. Mellin transform

Let $\mathbb{L}^{1,\text{loc}}_{\mathbb{R}_+}$ denote the set of locally integrable real-valued functions. For $h \in \mathbb{L}^{1,\text{loc}}_{\mathbb{R}_+}$ the Mellin transform of h in the point $c + it \in \mathbb{C}$ is defined by

$$\mathcal{M}_c[h](t) := \mathcal{M}[h](c + it) := \int_0^\infty x^{c-1+it} h(x)\,dx \qquad (2.1)$$

provided that the integral is absolutely convergent. If there exists a $c \in \mathbb{R}$ such that the mapping $x \mapsto x^{c-1} h(x)$ is integrable over $\mathbb{R}_+$ then the region $\Xi_h \subseteq \mathbb{C}$ of absolute convergence of the integral in Eq. 2.1 is either a vertical strip $\{s + it \in \mathbb{C} : s \in (a,b), t \in \mathbb{R}\}$ for $a < b$ with $c \in (a,b)$ or a vertical line $\{c + it \in \mathbb{C} : t \in \mathbb{R}\}$. In the case that Ξ_h is a vertical strip the function $s + it \mapsto \mathcal{M}_s[h](t)$ is analytical on this strip. In the literature Ξ_h is often called *strip of analyticity*. In the following illustration we give some techniques to determine Ξ_h.

Note that for any density $h_1 \in \mathbb{L}_{\mathbb{R}_+}^{1,\mathrm{loc}}$ the vertical strip $\{1 + it : t \in \mathbb{R}\}$ belongs to Ξ_{h_1}, and hence the Mellin transform $\mathcal{M}_1[h_1]$ is well-defined. Furthermore the region Ξ_{h_2} of $h_2 \in \mathbb{L}_{\mathbb{R}_+}^{1,\mathrm{loc}}$ is a superset of the vertical strip $\{c + it : c \in (a,b), t \in \mathbb{R}\}$ if $h_2(x) = \mathrm{O}(x^{-a+\varepsilon})$ for $x \downarrow 0$ and $h_2(x) = \mathrm{O}(x^{-b-\varepsilon})$ for $x \to \infty$ and for all $\varepsilon > 0$ small enough.

Illustration 2.1. *Now let us give a few examples of Mellin transforms of commonly considered distribution families.*

1. *Beta Distribution: Consider the family $(g_b)_{b \in \mathbb{N}}$, $g_b(x) := \mathbb{1}_{(0,1)}(x)b(1 - x)^{b-1}$ for a $b \in \mathbb{N}$ and $x \in \mathbb{R}_+$. Obviously we see that $\mathcal{M}_c[g_b]$ is well-defined for $c > 0$ and it holds*

$$\mathcal{M}_c[g_b](t) = \prod_{j=1}^b \frac{j}{c - 1 + j + it}, \quad t \in \mathbb{R}.$$

2. *Gamma Distribution: Consider the family $(g_d)_{d \in \mathbb{R}_+}$, where for $d, x \in \mathbb{R}^+$ $g_d(x) = \frac{x^{d-1}}{\Gamma(d)} \exp(-x)\mathbb{1}_{\mathbb{R}^+}(x)$. Obviously we see that $\mathcal{M}_c[g_d]$ is well-defined for $c > -d + 1$ and it holds*

$$\mathcal{M}_c[g_d](t) = \frac{\Gamma(c + d - 1 + it)}{\Gamma(d)}, \quad t \in \mathbb{R}.$$

3. *Scaled Log-Gamma Distibution: Consider the family $(g_{\mu,a,\lambda})_{(\mu,a,\lambda) \in \mathbb{R} \times \mathbb{R}^+ \times \mathbb{R}^+}$, $g_{\mu,a\lambda}(x) = \frac{\exp(\lambda\mu)\lambda^a}{\Gamma(a)} x^{-\lambda-1}(\log(x) - \mu)^{a-1}\mathbb{1}_{(e^\mu,\infty)}(x)$ for $a, \lambda, x \in \mathbb{R}^+$ and $\mu \in \mathbb{R}$. Then for $c < \lambda + 1$ holds*

$$\mathcal{M}_c[g_{\mu,a,\lambda}](t) = \lambda^a \exp(\mu(c - 1 + it))(\lambda - c + 1 - it)^{-a}, \quad t \in \mathbb{R}.$$

   If $a = 1$ then $g_{\mu,1,\lambda}$ is the density of a Pareto distribution with parameter e^μ and λ. If $\mu = 0$ we have that $g_{0,a,\lambda}$ is the density of a Log-Gamma distribution.

4. *Weibull Distribution: Consider the family $(g_m)_{m \in \mathbb{R}_+}$, where for $m, x \in \mathbb{R}^+$ $g_m(x) = mx^{m-1}\exp(-x^m)\mathbb{1}_{\mathbb{R}^+}(x)$. Obviously we see that $\mathcal{M}_c[g_m]$ is well-defined for $c > -m + 1$ and it holds*

$$\mathcal{M}_c[g_m](t) = \frac{(c - 1 + it)}{m}\Gamma\left(\frac{c - 1 + it}{m}\right), \quad t \in \mathbb{R}.$$

Let us now come back to the collection of important properties of the Mellin transform. For $c \in \Xi_h$ the inversion formula of the Mellin transform is given by

$$h(x) = \frac{1}{2\pi} \int_{-\infty}^{\infty} x^{-c-it} \mathcal{M}_c[h](t) dt, \text{ for } x > 0, \tag{2.2}$$

whenever Ξ_h is not a vertical line (c.f. [13]) or alternatively if the function $t \mapsto \mathcal{M}_c[h](t)$ is square integrable over $\mathbb{R}$ (c.f. [3]). Considering Illustration 2.1 we see that the assumption on Ξ_h not being a vertical line is rather weak. It is fulfilled for almost all functions in the upcoming theory. However, in all the other cases we make use of the second assumption, that $\mathcal{M}_c[h]$ is square integrable, without further reference.

It can be shown that $\int_0^{\infty} |h(x)|^2 x^{2c-1} dx < \infty$ for $c \in \mathbb{R}$ implies that $c \in \Xi_h$ and also that $\int_0^{\infty} |h(x)|^2 x^{2c-1} dx = (2\pi)^{-1} \int_{-\infty}^{\infty} |\mathcal{M}_c[h](t)|^2 dt$. This result combined with the Mellin inversion formula implies an isometry in the following way. For $c \in \mathbb{R}$ define the weight function $\omega_c(x) := x^{2c-1}, x \in \mathbb{R}$, and the corresponding weighted norm by $\|h\|_{\omega_c}^2 := \int_0^{\infty} |h(x)|^2 \omega_c(x) dx$ for a measurable function. Denote by $\mathbb{L}_{\mathbb{R}_+}^2(\omega_c)$ the set of all measurable functions with finite $\| . \|_{\omega_c}$-norm and by $\langle h_1, h_2 \rangle_{\omega_c} := \int_0^{\infty} h_1(x) \overline{h_2(x)} \omega_c(x) dx$ for $h_1, h_2 \in \mathbb{L}_{\mathbb{R}_+}^2(\omega_c)$ the corresponding weighted scalar product. Similarly, define $\mathbb{L}_{\mathbb{R}}^2 := \{h : \mathbb{R} \to \mathbb{C}$ measurable $: \|h\|_{\mathbb{R}}^2 := \int_{-\infty}^{\infty} h(t) \overline{h(t)} dt < \infty\}$. Now, following [3] both operators $\mathcal{M}_c : \mathbb{L}_{\mathbb{R}_+}^2(\omega_c) \to \mathbb{L}_{\mathbb{R}}^2$ and

$$\mathcal{M}_c^{-1} : \mathbb{L}_{\mathbb{R}}^2 \to \mathbb{L}_{\mathbb{R}_+}^2(\omega_c), \quad h \mapsto (x \mapsto \mathcal{M}_c^{-1}[h](x) := (2\pi)^{-1} \int_{-\infty}^{\infty} x^{-c-it} h(t) dt) \tag{2.3}$$

are isomorphism. We first construct an estimator for f given an i.i.d. sample $X_1, \ldots, X_n$, that is the direct case, and afterwards we construct an estimator based on the i.i.d. sample $Y_1, \ldots, Y_n$, which constitutes the censored case.

2.2. *Case of direct observation*

In this paragraph we define the estimator of $f \in \mathbb{L}_{\mathbb{R}_+}^2(\omega_c)$ given the direct observations $X_1, \ldots, X_n$ by using the Mellin transform and spectral cut-off regularised inverse. Since $f \in \mathbb{L}_{\mathbb{R}_+}^2(\omega_c)$ the Mellin transform $\mathcal{M}_c[f]$ is well-defined and a natural unbiased estimator of $\mathcal{M}_c[f][t]$ for each $t \in \mathbb{R}$ is given by $\widehat{\mathcal{M}}_c(t) := n^{-1} \sum_{j=1}^n X_j^{c-1+it}$. It is worth pointing out that this estimator is bounded in $t \in \mathbb{R}$ by $|\widehat{\mathcal{M}}_c(t)| \leq |\widehat{\mathcal{M}}_c(0)|$ which is finite almost surely. Thus $\mathbb{1}_{[-k,k]} \widehat{\mathcal{M}}_c \in \mathbb{L}_{\mathbb{R}}^2$ for all $k \in \mathbb{R}_+$ which allows us to apply the operator in Eq. 2.3 to define

$$\widehat{f}_k(x) := \mathcal{M}_c^{-1}[\mathbb{1}_{[-k,k]} \widehat{\mathcal{M}}_c](x) = \frac{1}{2\pi} \int_{-k}^{k} x^{-c-it} \widehat{\mathcal{M}}_c(t) dt, \text{ for } x \in \mathbb{R}_+, \tag{2.4}$$

as an unbiased estimator of $f_k := \mathcal{M}_c^{-1}[\mathbb{1}_{[-k,k]}\mathcal{M}_c[f]]$. Additionally, we see that $\|f - f_k\|_{\omega_c}^2 = \pi^{-1}\int_k^\infty |\mathcal{M}_c[f](t)|^2 dt$ tends to zero for k going to infinity, that is f_k approximates f in the weighted $\mathbb{L}_{\mathbb{R}_+}^2$ sense.

Furthermore, we have $\|\widehat{f}_k\|_{\omega_c}^2 = (2\pi)^{-1}\int_{-k}^k |\widehat{\mathcal{M}}_c(t)|^2 dt$. By application of Fubini-Tonelli theorem, we deduce that $\mathbb{E}_f^n(\|\widehat{f}_k\|_{\omega_c}^2) < \infty$ if and only if $\mathbb{E}_f^1(X^{2(c-1)}) < \infty$. We state the following proposition which implies the consistency of the estimator for a suitable choice of the cut-off parameter $k \in \mathbb{R}_+$.

Proposition 2.1. *Assume that $f \in \mathbb{L}_{\mathbb{R}_+}^2(\omega_c)$ and that $\sigma_c^2 := \mathbb{E}_f(X^{2(c-1)}) < \infty$. Then, for all $k \in \mathbb{R}^+$ we have*

$$\mathbb{E}_f^n(\|f - \widehat{f}_k\|_{\omega_c}^2) \le \|f - f_k\|_{\omega_c}^2 + \pi^{-1}\sigma_c^2 k n^{-1}.$$

By choosing $k = k_n$ such that $n^{-1}k_n \to 0$ and $k_n \to \infty$, $\widehat{f}_{k_n}$ is a consistent estimator of f.

The proof of Proposition 2.1 can be found in B. Let us comment the condition $\mathbb{E}_f^1(X^{2(c-1)}) < \infty$. Note that, if one would like to consider the case $c = 1/2$, that is $\omega_c = 1$, it is necessary to assume a finite first moment of X^{-1}. Considering the examples of Illustration 2.1 we see that the assumption $\mathbb{E}_f(X^{-1}) < \infty$ holds not true for (i) and for (ii) and (iv) if $k \le 1$.

On the other hand, a less restrictive situation would be to consider the case of $c = 1$ which needs no additional moment condition on f (respectively on g) since in this case $\sigma_c^2 = 1$ holds automatically. Note that in this case $\omega(x) := \omega_1(x) = x$ for $x \in \mathbb{R}_+$ is the corresponding weight function. Furthermore, when estimating a density without a compact support assumption, intervals far away from zero are often of special interest. A choice of the model parameter $c > 1/2$ could model this and may allow us to capture more interesting characteristics of the density f like being heavy-tailed or compactly supported.

In the next part we define an estimator of f based on the censored observation $Y_1, \ldots, Y_n$. Since $Y = XU$, in the multiplicative measurement error model we would need to assume that both X and U have a finite -1-moment to consider the unweighted norm. Especially the latter scenario would exclude several interesting examples, for instance the multiplicative censoring model where U is uniformly distributed on $[0, 1]$.

2.3. Case of censored observation

Coming back to the original model of multiplicative censored observations the key property which makes the Mellin transform useful for multiplicative deconvolution is that for two functions $h_1, h_2 \in \mathbb{L}_{\mathbb{R}_+}^{1,\mathrm{loc}}$ with $c \in \Xi_{h_1} \cap \Xi_{h_2}$,

$$\mathcal{M}_c[h_1 * h_2](t) = \mathcal{M}_c[h_1](t) \cdot \mathcal{M}_c[h_2](t) \quad \text{for } t \in \mathbb{R}. \tag{2.5}$$

We will refer to it from now on as the convolution theorem. In the context of deconvolution a similar property adressing the convolution and its Fourier

transform is frequently used to construct deconvolution estimators. Since f and g are both densities we have $1 \in \Xi_f \cap \Xi_g$ which implies that for all $t \in \mathbb{R}$, $\mathcal{M}_1[f_Y](t) = \mathcal{M}_1[f](t)\mathcal{M}_1[g](t)$. More generally assuming that $c \in \Xi_f \cap \Xi_g$ implies that for $t \in \mathbb{R}, \mathcal{M}_c[f_Y](t) = \mathcal{M}_c[f](t)\mathcal{M}_c[g](t)$.

Under the mild assumption that $\mathcal{M}_c[g](t) \neq 0$ for all $t \in \mathbb{R}$, which we do in the upcoming theory without further reference, we see that for all $t \in \mathbb{R}$ holds $\mathcal{M}_c[f](t) = \mathcal{M}_c[f_Y](t)/\mathcal{M}_c[g](t)$ which allows us to express the functions $(f_k)_{k \in \mathbb{R}_+}$ in the following way

$$f_k(x) = \frac{1}{2\pi} \int_{-k}^{k} x^{-c-it} \mathcal{M}_c[f](t) dt = \frac{1}{2\pi} \int_{-k}^{k} x^{-c-it} \frac{\mathcal{M}_c[f_Y](t)}{\mathcal{M}_c[g](t)} dt$$

for $k, x \in \mathbb{R}_+$. Similar to the direct case we define our estimator by replacing $\mathcal{M}_c[f_Y](t)$ with its empirical counterpart $\widehat{\mathcal{M}}_c(t) = \frac{1}{n} \sum_{j=1}^{n} Y_j^{c-1+it}$ to define the estimator

$$\widehat{f}_k(x) := \frac{1}{2\pi} \int_{-k}^{k} x^{-c-it} \frac{\widehat{\mathcal{M}}_c(t)}{\mathcal{M}_c[g](t)} dt \quad \text{for } x, k \in \mathbb{R}^+. \tag{2.6}$$

Here and subsequently we assume that the error density g belongs to a function class $\mathbb{G}_{0,c} \subset \mathbb{L}^2_{\mathbb{R}_+}(\omega_c)$, which ensures that the estimator is well-defined, where we set

$$\mathbb{G}_{0,c} := \{g \in \mathbb{L}^2_{\mathbb{R}_+}(\omega_c) : \forall t \in \mathbb{R} : \mathcal{M}_c[g](t) \neq 0, \forall k \in \mathbb{R}_+ : \frac{\mathbb{1}_{[-k,k]}}{\mathcal{M}_c[g]} \in \mathbb{L}^2_{\mathbb{R}}\}. \tag{2.7}$$

Note that $\mathcal{M}_c[\widehat{f}_k](t) = \mathbb{1}_{[-k,k]}(t) \frac{\widehat{\mathcal{M}}(t)}{\mathcal{M}_c[g](t)}$ by definition of $\widehat{f}_k$, and hence the estimators defined in Eq. 2.6 and Eq. 2.4 coincide when setting $\mathcal{M}_c[g](t) = 1$ for $t \in \mathbb{R}$. The proof of the next proposition is very similar to the proof of Proposition 2.1 and thus omitted.

Proposition 2.2. *Assume that $f \in \mathbb{L}^2_{\mathbb{R}_+}(\omega_c)$, $g \in \mathbb{G}_{0,c}$ and $\sigma_c^2 := \mathbb{E}_f(Y^{2(c-1)}) < \infty$. Then for all $k \in \mathbb{N}$ we have*

$$\mathbb{E}^n_{f_Y}(\|f - \widehat{f}_k\|^2_{\omega_c}) \leq \|f - f_k\|^2_{\omega_c} + \frac{\sigma_c^2}{2\pi n} \int_{-k}^{k} |\mathcal{M}_c[g](t)|^{-2} dt.$$

By choosing $k = k_n$ such that $n^{-1} \int_{-k_n}^{k_n} |\mathcal{M}_c[g](t)|^{-2} dt \to 0$ and $k_n \to \infty$, $\widehat{f}_{k_n}$ is a consistent estimator of f.

From now on we will restrict ourselves to the case of a model parameter $c = 1$, since in that case $\sigma_1^2 = 1$ which simplifies the presentation of the upcoming theory. Nevertheless we want to stress out that the theory extends straight forward for arbitrary model parameters $c \in \mathbb{R}$. Further let us use the abbreviation $\omega := \omega_1$.

Let us now have a closer look at the second summand in Proposition 2.2 which bounds the variance term of the estimator. In fact, the growth of the second summand in $k \in \mathbb{R}^+$ is determined by the decay of the Mellin transform

of the error density g. In analogy to the usual deconvolution setting (compare [11]) and to the work of [5] we define the function class $\mathbb{G}_{1.\gamma} \subset \mathbb{G}_{0,1}$ of *smooth error densities with decay* $\gamma \in \mathbb{R}_+$ through

$$\mathbb{G}_{1,\gamma} := \{g \in \mathbb{G}_{0,1} : \exists c, C \in \mathbb{R}^+ : c|t|^{-\gamma} \leq |\mathcal{M}_1[g](t)| \leq C|t|^{-\gamma} \text{ as } |t| \to \infty\}. \quad (2.8)$$

Remark 2.1. *For the family of Beta distributions* $(g_b)_{b \in \mathbb{N}}$ *and the family of Scaled Log-Gamma distributions* $(g_{(\mu,a,\lambda)})_{(\mu,a,\lambda) \in \mathbb{R} \times \mathbb{R}_+ \times \mathbb{R}_+}$ *from Illustration 2.1 we see that both* $g_b \in \mathbb{G}_{1,b}$ *for any* $k \in \mathbb{N}$ *and* $g_{\mu,a,\lambda} \in \mathbb{G}_{1,a}$ *for any* $\mu \in \mathbb{R}, a \in \mathbb{R}, \lambda \in \mathbb{R}_+$. *By application of the Stirling formula (compare [5]) for the examples (ii) and (iv) we see that the Mellin transform has exponential decay. This implies that these families do not belong to* $\mathbb{G}_{1,\gamma}$ *for any choice* $\gamma > 0$.

Now $g \in \mathbb{G}_{1,\gamma}$ implies that $\int_{-k}^{k} |\mathcal{M}_1[g](t)|^{-2}dt \leq C_g k^{2\gamma+1}$ where $C_g > 0$ is a positive constant. This consideration and the result of Proposition 2.2 directly imply the following corollary whose proof is thus omitted.

Corollary 2.1. *Assume that* $f \in \mathbb{L}^2_{\mathbb{R}_+}(\omega)$ *and that* $g \in \mathbb{G}_{1,\gamma}$. *Then for all* $k \in \mathbb{N}$ *we have*

$$\mathbb{E}^n_{f_Y}(\|f - \widehat{f}_k\|^2_\omega) \leq \|f - f_k\|^2_\omega + C_g(2\pi n)^{-1}k^{2\gamma+1}$$

where C_g *is a constant only dependent on g. By choosing* k_n *such that* $k_n \to \infty$ *and* $n^{-1}k_n^{2\gamma+1} \to 0$, $\widehat{f}_{k_n}$ *is a consistent estimator of f.*

Under the assumptions of Corollary 2.1 it is natural to restrict the set of suitable parameters k to $\mathcal{K}_n := [\![1, K_n]\!]$ with $K_n := \lfloor n^{1/(2\gamma+1)} \rfloor$ and to choose $k_n :\in \arg\min\{\|f - f_k\|^2_\omega + C_g(2\pi n)^{-1}k^{2\gamma+1} : k \in \mathcal{K}_n\}$. Unfortunately, this choice is not feasible since it depends on the unknown density f itself. We note that the bias $\|f - f_k\|^2_\omega = \|f\|^2_\omega - \|f_k\|^2_\omega$ behaves like $-\|f_k\|^2_\omega$. Exchanging $-\|f_k\|^2_\omega$ with its empirical counterpart $-\|\widehat{f}_k\|^2_\omega$ we define a fully data-driven model selection $\widehat{k}$ by

$$\widehat{k} \in \arg\min\{-\|\widehat{f}_k\|^2_\omega + \text{pen}(k) : k \in \mathcal{K}_n\} \quad \text{where} \quad \text{pen}(k) := \chi k^{2\gamma+1}n^{-1} \quad (2.9)$$

for $\chi > 0$. The following theorem shows that this procedure is adaptive up to a negligeable term.

Theorem 2.1. *Assume that* $f \in \mathbb{L}^2_{\mathbb{R}_+}(\omega)$, $g \in \mathbb{G}_{1,\gamma}$ *and that* $\|\omega f_Y\|_\infty := \sup_{y>0}|yf_Y(y)| < \infty$. *Then for* $\chi > 12C_g\pi^{-1}$

$$\mathbb{E}^n_{f_Y}(\|f - \widehat{f}_{\widehat{k}}\|^2_\omega) \leq 4\inf_{k \in \mathcal{K}_n}\left(\|f - f_k\|^2_\omega + \text{pen}(k)\right) + C(\|\omega f_Y\|_\infty, g)n^{-1}$$

where $C(\|\omega f_Y\|_\infty, g) > 0$ *is a constant depending on* $\|\omega f_Y\|_\infty$ *and g.*

The proof of Theorem 2.1 is postponed to B. The assumption that $\|\omega f_Y\|_\infty < \infty$ is rather weak. In fact, since $1 \in \Xi_f \cap \Xi_g$ we are able to write for all $y \in \mathbb{R}_+$

$$|yf_Y(y)| = |y\frac{1}{2\pi}\int_{-\infty}^{\infty} y^{-1-it}\mathcal{M}_1[f](t)\mathcal{M}_1[g](t)dt| \leq \|f\|_\omega\|g\|_\omega < \infty.$$

The last assertion establishes an oracle inequality assuming a smooth error density. For a *super smooth* error density with exponentially decay of its Mellin transform (see [5]) a result similar to Theorem 2.1 can be derived from B.4 in the appendix provided the upper bound K_n and the penalty terms $\text{pen}(k)$, $k \in [\![1, K_n]\!]$ are choosen accordingly. However, we omit the details, since the minimax theory presented in the next chapter does not cover a *super smooth* error density.

3. Minimax theory

In this section we develop the minimax theory for the proposed estimator in section 2. Over the *Mellin-Sobolev* spaces we derive an upper and lower bound for the mean weighted integrated squared error, which are equal up to a multiplicative constant, showing that our estimator is minimax-optimal over these spaces.

A natural condition which allows a more sophisticated study on the bias, is to assume that the Mellin-transform decays with a polynomial rate, that is for $s \geq 0$,

$$f \in \mathbb{W}^s := \{ h \in \mathbb{L}^2_{\mathbb{R}_+} : |h|_s^2 := \int_{-\infty}^{\infty} |\mathcal{M}_1[h](t)|^2 (1+t^2)^s dt < \infty \}. \qquad (3.1)$$

The definition of these spaces strongly resembles to the frequently considered Fourier-Sobolev spaces which are defined as $W^s := \{ H \in \mathbb{L}^2_{\mathbb{R}} : \int_{-\infty}^{\infty} |\mathcal{F}[H](t)|^2 (1+t^2)^s dt < \infty \}$ for $s \geq 0$ where $\mathcal{F}[H]$ denotes the usual Fourier-Plancherel-transform. If additionally $H \in \mathbb{L}^1(\mathbb{R})$ we can express $\mathcal{F}[H]$ explicitly through $\mathcal{F}[H](t) := \int_{-\infty}^{\infty} H(x) \exp(-ixt) dx$ for any $t \in \mathbb{R}$. In analogy we refer $\mathbb{W}^s$ as *Mellin-Sobolev space.*

But not just the general motivation seems to be similar. In fact there is a strong connection between the *Mellin-Sobolev spaces* and *Fourier-Sobolev spaces* we will use to characterize this function space through regularity assumptions which we exploit in the following paragraph.

3.1. Comparison between Mellin-Sobolev and Fourier-Sobolev spaces

For $h \in \mathbb{L}^2_{\mathbb{R}_+}(\omega)$ we have $\mathcal{M}_1[h] = \mathcal{M}_0[\omega h] = \mathcal{F}[(\omega h) \circ \varphi]$ with $\varphi : \mathbb{R} \to \mathbb{R}_+, x \mapsto \exp(-x)$. Setting $H := (\omega h) \circ \varphi$ we see that the conditions $h \in \mathbb{W}^s$ and $H \in W^s$ are equivalent. Therefore, it does not seem to be a suprise that it is possible to characterise the *Mellin-Sobolev spaces* via analytical properties as done in the case of the *Fourier-Sobolev spaces.* This is stated in the following proposition while its proof is postponed to the appendix.

Proposition 3.1. *Let $s \in \mathbb{N}$. Then $f \in \mathbb{W}^s$ if and only if f is $(s-1)$-times continuously differentiable where $f^{(s-1)}$ is locally absolutely continuous with derivative $f^{(s)}$ and $\omega^j f^{(j)} \in \mathbb{L}^2_{\mathbb{R}_+}(\omega)$ for all $j \in [\![0, s]\!]$.*

If a positive random variable R has a density h, then the real-valued random variable $T = \log(R)$ admits a density $f_T(y) = H(-y) = (\omega h) \circ \varphi(-y)$ for $y \in \mathbb{R}$. Again, we observe the strong connection between the Mellin transform of h and the Fourier transform of H. This has the following interesting implication for the multiplicative measurement error model.

Remark 3.1. *As already mentioned the application of the logarithm to the random variable Y, $Z := \log(Y) = \log(X) + \log(U) =: V + \varepsilon$, where V and ε are independent, can be used to transfer the model to a regular deconvolution setting. This technique was used for instance by [9]. Given an estimator $\widehat{f}_V$ of f_V it is possible to derive an estimator of f through $\widehat{f} := \omega^{-1}\widehat{f}_V \circ \log$. In fact, one can show that $\|\widehat{f}_V - f_V\|_{\mathbb{R}}^2 = \|\widehat{f} - f\|_{\omega}^2$. This change of risk illustrates the difficulties which arise when using this technique to consider the global risk. Furthermore, in the deconvolution approach via a Fourier transformation one usually assumes that the densities $f_V, f_\varepsilon \in \mathbb{L}_{\mathbb{R}}^2$ which again corresponds to the fact that $f, g \in \mathbb{L}_{\mathbb{R}_+}^2(\omega)$ using the considerations above.*

3.2. Regularity assumptions

Let us define for $s \geq 0$ and the ellipsoids $\mathbb{W}^s(L) := \{h \in \mathbb{W}^s : |h|_s^2 \leq L\}$ for any $L \geq 0$ which correspond to the *Mellin-Sobolev spaces* defined in Eq. 3.1. We see that for any $f \in \mathbb{W}^s(L)$ we have $\int_{[-k,k]^c} |\mathcal{M}_1[f](t)|^2 dt \leq Lk^{-2s}$ and $\|f\|_\omega^2 = (2\pi)^{-1} \int_{-\infty}^{\infty} |\mathcal{M}_1[f](t)|^2 dt \leq L(2\pi)^{-1}$. We denote the subset of densities by

$$\mathbb{D}_{\mathbb{R}_+}^{s,L} := \{f \in \mathbb{W}^s(L) : f \text{ is a density}\}. \tag{3.2}$$

Again, assuming $g \in \mathbb{G}_{1,\gamma}$ implies that $\int_{-k}^{k} |\mathcal{M}_1[g](t)|^{-2} dt \leq C_g k^{2\gamma+1}$ where $C_g > 0$ is a constant only dependent on the error density g. These considerations imply the following theorem whose proof is omitted.

Theorem 3.1. *Assume that $g \in \mathbb{G}_{1,\gamma}$. Then for $k_o := n^{1/(2s+2\gamma+1)}$,*

$$\sup_{f \in \mathbb{D}_{\mathbb{R}_+}^{s,L}} \mathbb{E}_{f_Y}^n(\|f - \widehat{f}_{k_o}\|_\omega^2) \leq C(g,L,s)n^{-2s/(2s+2\gamma+1)}.$$

As mentioned before, for $\gamma > 1/2$, we have $\|\omega f_Y\|_\infty \leq \|f\|_\omega \|g\|_\omega$. Thus, we can state the following corollary which is a direct consequence of Theorem 2.1 and Theorem 3.1.

Corollary 3.1. *Assume that $g \in \mathbb{G}_{1,\gamma}$ for $\gamma > 1/2$. Then for $\chi > 12C_g\pi^{-1}$,*

$$\sup_{f \in \mathbb{D}_{\mathbb{R}_+}^{s,L}} \mathbb{E}_{f_Y}^n(\|f - \widehat{f}_{\widehat{k}}\|_\omega^2) \leq C(g,L,s)n^{-2s/(2s+2\gamma+1)},$$

where χ is defined in Eq. 2.9.

Remark 3.2. *For $g_b(x) = b(1-x)^{b-1}\mathbb{1}_{(0,1)}(x)$ with $b \in \mathbb{N}$, $x > 0$ and $c > 0$ its Mellin transform is given by $\mathcal{M}_c[g_b](t) = \prod_{j=1}^{b} \frac{j}{c-1+j+it}$ and satisfies $g \in \mathbb{G}_{1,b}$. In fact this covers the model considered by [4] as a generalisation of the multiplicative censoring where we consider a uniform distributed error density, that is $b = 1$. Again, setting $\gamma = 0$ covers the case of direct observations, getting a rate of $n^{-2s/(2s+1)}$ over the ellipsoid.*

To prove that the rate of Theorem 3.1 is minimax-optimal over the ellipsoids under certain assumptions on g we finish this section by stating a lower bound result. We want to emphasize that up to now we had no constraints on the support of g. To prove the lower bound we will need to assume that g has a bounded support. For the sake of simplicity we will assume that g has a support in $[0, 1]$ and define the function class $\mathbb{G}'_{1,\gamma} \subset \mathbb{G}_{1,\gamma}$ through

$$\mathbb{G}'_{1,\gamma} := \{g \in \mathbb{G}_{1,\gamma} : \forall x > 1 : g(x) = 0 \text{ and }$$
$$\exists c, C \in \mathbb{R}^+ : c|t|^{-\gamma} \leq |\mathcal{M}_{1/2}[g](t)| \leq C|t|^{-\gamma} \text{ as } |t| \to \infty\}.$$

Theorem 3.2. *Let $s, \gamma \in \mathbb{N}$, assume that $g \in \mathcal{G}'_{1,\gamma}$. Then there exist constants $c_g, L_{s,g}, n_{s,\gamma} > 0$ such that for all $L \geq L_{s,g}$, $n \geq n_{s,\gamma}$ and for any estimator $\widehat{f}$ of f based on an i.i.d. sample $Y_1, \ldots, Y_n$,*

$$\sup_{f \in \mathbb{D}_{\mathbb{R}_+}^{s,L}} \mathbb{E}_{f_Y}^n (\|f - \widehat{f}\|_\omega^2) \geq c_g n^{-2s/(2s+2\gamma+1)}.$$

We want to emphasize that the error densities $(g_b)_{b \in \mathbb{N}}$, with $g_b(x) = b(1 - x)^{b-1}\mathbb{1}_{(0,1)}, x \in \mathbb{R}_+$, lies in $\mathbb{G}'_{1,b}$. Thus, in this situation our estimation strategy is minimax-optimal. The proof of the lower bound can be extended to the case of directly observed $X_1, \ldots, X_n$ or for different weight functions $\omega_c, c \in \mathbb{R}^+$.

Remark 3.3. *In the work of [5] one can find a similar lower bound result. The authors considered in their work the analysis of the pointwise risk while we are studying the global risk. Additionally their regularity assumptions are more adpated to the pointwise risk while ours are natural in context of global density estimation. Our assumptions seem also to be similar but still different from the work of [5] which makes a comparison of the results difficult. Nevertheless both lower bound results show the same rate which is typical in the context of nonparametric inverse problems.*

4. Numerical study

Let us illustrate the performance of the estimator $\widehat{f}_{\widehat{k}}$ defined in Eq. 2.6 and Eq. 2.9 in the cases $U \sim \mathrm{U}_{[0,1]}$, $U \sim \mathrm{U}_{(0.5,1.5)}$ and $U \sim \mathrm{Beta}_{(1,2)}$. For the density g of an uniform distribution on $[0.5, 1.5]$ we get that $\mathcal{M}_1[g](t) = (1 + it)^{-1}(1.5^{1+it} - 0.5^{1+it}), t \in \mathbb{R}$, which corresponds to the case $g \in \mathbb{G}_{1,1}$. We consider the densities
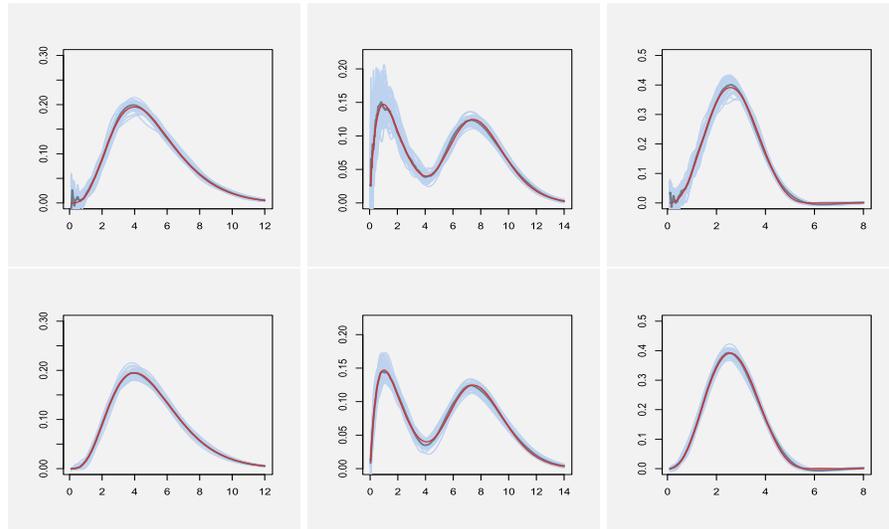
FIG 1. *Considering the estimators $\widehat{f}_{\widehat{k}}$ (top) and $\widetilde{f}$ (bottom) are depict for 50 Monte-Carlo simulations with sample size $n = 1000$ in the case 1 (left), 2 (middle) and 3 (right) with direct observations. The true density f is given by the red curve while the dark blue curve is the point-wise empirical median of the 50 estimates.*

1. *Gamma Distribution*: $f(x) = \frac{x^4}{4!}\exp(-x)$,
2. *Gamma Mixture:* $f(x) = 0.4 \cdot 3.2^2 x \exp(-3.2x) + 0.6 \cdot \frac{6.8^{16}x^{15}}{15!}\exp(-6.8x)$,
3. *Beta Distribution*: $f(x) = \frac{1}{560}(0.5x)^3(1-0.5x)^4 \mathbb{1}_{[0,1]}(0.5x)$ and
4. *Weibull Distribution*: $f(x) = 2x\exp(-x^2)$.

By minimising an integrated weighted squared error over a family of histogram densities with randomly drawn partitions and weights we select $\chi = 1.2$, $\chi = 0.8$ and $\chi = 0.01$ for the cases $\gamma = 0$, $\gamma = 1$ and $\gamma = 2$, respectively, where χ is the penalty constant, see Eq. 2.9.

In the direct case we compare the estimator $\widehat{f}_{\widehat{k}}$ with the data-driven density estimator $\widetilde{f}$ from the work of [6] which is based on the adaptive aggregation of projection estimators with respect to the Laguerre basis.

4.1. Comment

Since the estimator $\widehat{f}_{\widehat{k}}$ is built to minimize the weighted global risk, it seems natural that the estimator $\widehat{f}_{\widehat{k}}$ behaves worse in the region close to zero then the estimator $\widetilde{f}$ which is built to minimize the unweighted global risk. This effect is observable in Figure 1. Furthermore, the developed minimax theory suggests that the cases of $U \sim \mathrm{U}_{[0,1]}$ and $U \sim \mathrm{U}_{[0.5,1.5]}$ are of similiar complexity which is reflected in the plots of Figure 2. For the case $U \sim \beta(1,2)$, and thus $g \in \mathbb{G}_{1,2}$, both theory and simulation imply that the recovering of the density f based on
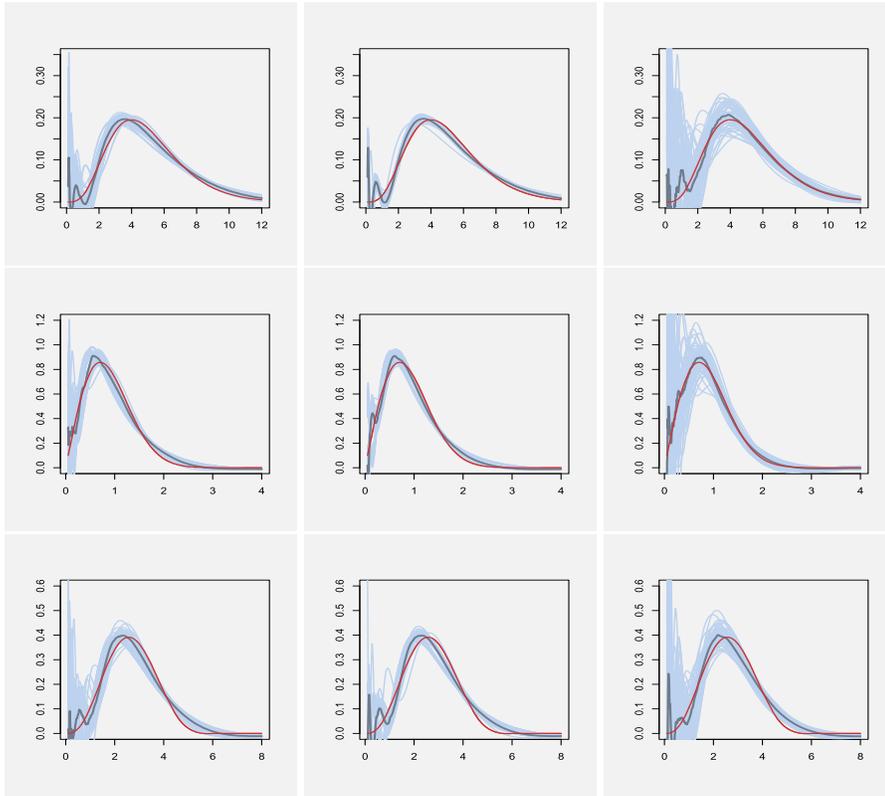
FIG 2. *Considering the estimator $\widehat{f}_{\widehat{k}}$ and a sample size $n = 2000$ the adaptive estimators are depicted for 50 Monte-Carlo simulations with $U \sim U_{[0,s]}$ (left), $U \sim U_{[1/2,3/2]}$ (middle) and $U \sim \beta(1,2)$ (right) in the cases 1 (first row), 4 (second row) and 3 (third row). The true density f is given by the red curve while the dark blue curve is the point-wise empirical median of the 50 estimates.*

the noise sample $Y_1, \ldots, Y_n$ leads to a more difficult inverse problem than in the other cases.

Appendix A: Preliminaries

A.1. Properties of the Mellin transform

By assuming that $h \in \mathbb{L}_{\mathbb{R}_+}^{1,\text{loc}}$ is a at least b-time differentiable function h, where $h^{(b)}$ denotes its b-th derivative, $b \in \mathbb{N}$, and that $c-b \in \Xi_h$ and $c+a \in \Xi_h, a \in \mathbb{N}$, we get that

$$\mathcal{M}_c[x^a h](t) = \mathcal{M}_{c+a}[h](t) \text{ respec. } \mathcal{M}_c[h^{(b)}](t) = (-1)^b \frac{\Gamma(c+it)}{\Gamma(c-b+it)} \mathcal{M}_{c-b}[h](t) \tag{A.1}$$

Combining both results in Eq. A.1 we get that $\mathcal{M}_c[xh^{(1)}](t) = (-c-it)\mathcal{M}_c[h](t)$ if $c+1 \in \Xi_h$ and h differentiable. Further for $h_1, h_2 \in \mathbb{L}_{\mathbb{R}_+}^2$ with $c \in \Xi_{h_1}$ and $1-c \in \Xi_{h_2}$ we get that $\int_0^\infty h_1(x)h_2(x)dx = \frac{1}{2\pi}\int_{-\infty}^\infty \mathcal{M}_c[h_1](t)\overline{\mathcal{M}_{1-c}[h_2](t)}dt$. Combining this and Eq. A.1 we conclude that for $h_1, h_2 \in \mathbb{L}_2(\omega_c)$ it holds $\langle h_1, h_2 \rangle_{\omega_c} = \frac{1}{2\pi}\int_{-\infty}^\infty \mathcal{M}_c[h_1](t)\overline{\mathcal{M}_c[h_2](t)}dt$.

Lemma A.1. *(Talagrand's inequality) Let $X_1, \ldots, X_n$ be independent $\mathcal{Z}$-valued random variables and let $\bar{\nu}_h = n^{-1}\sum_{i=1}^n [\nu_h(X_i) - \mathbb{E}(\nu_h(X_i))]$ for ν_h belonging to a countable class $\{\nu_h, h \in \mathcal{H}\}$ of measurable functions. Then,*

$$\mathbb{E}(\sup_{h\in\mathcal{H}} |\bar{\nu}_h|^2 - 6\Psi^2)_+ \leq C \left[\frac{\tau}{n}\exp\left(\frac{-n\Psi^2}{6\tau}\right) + \frac{\psi^2}{n^2}\exp\left(\frac{-Kn\Psi}{\psi}\right) \right] \quad (A.2)$$

with numerical constants $K = (\sqrt{2}-1)/(21\sqrt{2})$ and $C > 0$ and where

$$\sup_{h\in\mathcal{H}}\sup_{z\in\mathcal{Z}} |\nu_h(z)| \leq \psi, \qquad \mathbb{E}(\sup_{h\in\mathcal{H}}|\bar{\nu}_h|) \leq \Psi, \qquad \sup_{h\in\mathcal{H}}\frac{1}{n}\sum_{i=1}^n \mathbb{V}\mathrm{ar}(\nu_h(X_i)) \leq \tau.$$

Remark A.1. *Keeping the bound Eq. A.2 in mind, let us specify particular choices K, in fact $K \geq \frac{1}{100}$. The next bound is now an immediate consequence,*

$$\mathbb{E}(\sup_{h\in\mathcal{H}} |\bar{\nu}_h|^2 - 6\Psi^2)_+ \leq C \left(\frac{\tau}{n}\exp\left(\frac{-n\Psi^2}{6\tau}\right) + \frac{\psi^2}{n^2}\exp\left(\frac{-n\Psi}{100\psi}\right) \right) \quad (A.3)$$

In the sequel we will make use of the slightly simplified bounds Eq. A.3 rather than Eq. A.2.

Appendix B: Proofs of Section 2

Proof of Proposition 2.1. Since $\mathcal{M}_c[f - f_k](t) = 0$ for $|t| \leq k$ we get that $\langle f - f_k, f_k - \widehat{f}_k \rangle_{\omega_c} = \frac{1}{2\pi}\int_{-k}^k \mathcal{M}_c[f-f_k](t)\overline{\mathcal{M}_c[f_k-\widehat{f}_k](t)}dt = 0$ and thus $\|f - \widehat{f}_k\|_{\omega_c}^2 = \|f - f_k\|_{\omega_c}^2 + \|f_k - \widehat{f}_k\|_{\omega_c}^2$.

Finally we see $\|\widehat{f}_k - f_k\|_{\omega_c}^2 = \frac{1}{2\pi}\int_{-k}^k |\mathcal{M}_c[f](t) - \widehat{\mathcal{M}}_c(t)|^2 dt$ and $\mathbb{E}_f^n(|\mathcal{M}_c[f](t) - \widehat{\mathcal{M}}_c(t)|^2) = \mathbb{V}\mathrm{ar}(\widehat{\mathcal{M}}_c(t)) \leq \frac{1}{n}\mathbb{E}_f^1(|X_1^{c-1+it}|^2) = \sigma_c^2 n^{-1}$. Now by using Fubini we get $\mathbb{E}_f^n(\|\widehat{f}_k - f_k\|_{\omega_c}^2) \leq \sigma_c^2\pi^{-1}kn^{-1}$. $\qquad\square$

Proof of Proposition 3.1. The main strategy of this proofs relies on the well-known fact that

$$W^s = \{H \in \mathbb{L}_{\mathbb{R}}^2 : H \text{ weakly differentiable up to the order } s, H^{(i)} \in \mathbb{L}_{\mathbb{R}}^2, i \in [\![0,s]\!]\}$$

for $s \in \mathbb{N}$ and the already discussed connection between the Mellin transform and the Fourier transform. Further we want to stress out that for a function $H \in \mathbb{L}_\Omega^2, \Omega \subset \mathbb{R}$ open, being weakly differentiable corresponds to being locally absolutely continuous on Ω, that is h is absolutely continuous on all compact intervals $[a,b] \subset \Omega$, $a < b \in \Omega$.

Let us start by assuming $f \in \mathbb{W}^s$ then $F := (\omega f) \circ \varphi \in W^s$ which means that F is s-times weakly differentiable with $F^{(i)} \in \mathbb{L}^2_{\mathbb{R}}$ for $i \in [\![0, s]\!]$. Thus we get that F is a $s - 1$-times continously differentiable function, more precisely there exists a representant of the equivalence class of F such that it is $s - 1$-times continuously differentiable. Now since $f = \omega^{-1} F(\varphi^{-1})$ we can deduce that f itself is $s - 1$-times continuously differentiable. Let us define the operator $\mathcal{T} : C^1(\mathbb{R}_+) \to C^0(\mathbb{R}_+), f \mapsto (\omega f)^{(1)}$, $\mathcal{T}^0 :=$ Id denote the identity and $\mathcal{T}^j = \mathcal{T} \circ \mathcal{T}^{j-1}$ for $j \in \mathbb{N}$. Since $\mathcal{T}[\omega^j f^{(j)}] = (j+1)\omega^j f^{(j)} + \omega^{j+1} f^{(j+1)}$ for $j \in [\![0, s-1[\![$ we conclude $\mathcal{T}^j[f] = \sum_{i=0}^{j} b_{i,j} \omega^i f^{(i)}$, $b_{i,j} \geq 1$. Now we can use the following Lemma to deduce that $\|\mathcal{T}^j[f]\|^2_\omega = \|(\omega \mathcal{T}^j[f]) \circ \varphi\|^2_{\mathbb{R}} = \|F^{(j)}\|^2_{\mathbb{R}} < \infty$.

Lemma B.1. *For $h \in C^k(\mathbb{R}_+)$ and all $j \in [\![0, k]\!]$, it holds $(\omega \mathcal{T}^j[h]) \circ \varphi = (-1)^j (\varphi h(\varphi))^{(j)}$.*

From this follows directly that $\omega^j f^{(j)} \in \mathbb{L}^2_{\mathbb{R}_+}(\omega)$ for $j \in [\![0, s[\![$. As the next step we show that $f^{(s-1)}$ is locally absolutely continuous. To do so, we see first for $j \in [\![0, s[\![$ we have that $\mathcal{T}^j[f] \in \mathbb{W}^1$. Now using the following lemma implies that $\omega^{s-1} f^{(s-1)}$ is absolutely continuous as linear combination of $\{\mathcal{T}^j[f] : j \in [\![0, s[\![\}$.

Lemma B.2. *Let $h \in \mathbb{W}^1$. Then h is a locally absolutely continuous function with derivative $h' : \mathbb{R}_+ \to \mathbb{R}$ and $h, \omega h' \in \mathbb{L}^2_{\mathbb{R}_+}(\omega)$.*

Let now δ denote the derivative of $\omega^{s-1} f^{(s-1)}$. Then by Lemma B.2 we see that $\omega \delta \in \mathbb{L}^2_{\mathbb{R}_+}(\omega)$. Defining now $f^{(s)} := \omega^{-s+1} \delta - (s-1)\omega^{-1} f^{(s-1)} \in \mathbb{L}^1_{\text{loc}}(\mathbb{R}_+)$ we get for any $a, b \in \mathbb{R}_+$, $a < b$, $\int_a^b f^{(s)}(x) dx = f^{(s-1)}(b) - f^{(s-1)}(a)$ using the integration by part rule for absolutely continuous function (see [8]). Finally we have that $\omega^s f^{(s)} = \delta \omega - (s-1)\omega^{s-1} f^{(s-1)} \in \mathbb{L}^2_{\mathbb{R}_+}(\omega)$.

Let us now show the other direction, indeed let us assume that f is $s - 1$-times continuously differentiable, $f^{(s-1)}$ is locally absolutely continuous with derivative $f^{(s)}$ and $\omega^j f^{(j)} \in \mathbb{L}^2_{\mathbb{R}_+}(\omega)$ for $j \in [\![0, s]\!]$. Thus for $\mathcal{T}^j[f] = \sum_{i=0}^{j} c_{i,j} \omega^i f^{(i)} \in \mathbb{L}^2_{\mathbb{R}_+}(\omega)$ with $j \in [\![0, s-1[\![$ we have that $\omega(\mathcal{T}^j[f])^{(1)} = \mathcal{T}^{j+1}[f] - \mathcal{T}^j[f] \in \mathbb{L}^2_{\mathbb{R}_+}(\omega)$. We can conclude that $\mathcal{T}^j[f] \in \mathbb{W}^1$ for $j \in [\![0, s-1[\![$ applying the following lemma.

Lemma B.3. *Let $h : \mathbb{R}_+ \to \mathbb{R}$ be locally absolutely continous function with derivative $h' : \mathbb{R}_+ \to \mathbb{R}$ and $h, \omega h' \in \mathbb{L}^2_{\mathbb{R}_+}(\omega)$. Then $h \in \mathbb{W}^1$.*

Now setting $\delta = (s-1)\omega^{s-2} f^{(s-1)} + \omega^{s-1} f^{(s)} \in \mathbb{L}^1_{\text{loc}}(\mathbb{R}_+)$ we have that $\int_a^b \delta(x) dx = b^{s-1} f^{(s-1)}(b) - a^{s-1} f^{(s-1)}(a)$ and $\omega \delta \in \mathbb{L}^2_{\mathbb{R}_+}(\omega)$. Thus again applying Lemma B.3 on $\omega^s f^{(s)}$ again shows that $\omega^s f^{(s)} \in \mathbb{W}^1$ and thus $\mathcal{T}^{s-1}[f] \in \mathbb{W}^1$. Now we use that $\mathcal{M}_1[\mathcal{T}^{s-1}](t) = (-1)^{s-1} \mathcal{F}[F^{(s-1)}](t) = (-it)^{s-1} \mathcal{F}[F](t) = (-it)^{s-1} \mathcal{M}_1[f](t)$ which implies that $f \in \mathbb{W}^s$. $\square$

Proof of Lemma B.1. For $j = 0$ the claim is trivially correct. Assume that the claim hold for $j \in [\![1, k[\![$ then

$$(\omega \mathcal{T}^{j+1}[f]) \circ \varphi = (\omega \mathcal{T}^j[f + \omega f^{(1)}]) \circ \varphi = (-1)^{j-1}(\varphi f(\varphi) + \varphi^2 f^{(1)}(\varphi))^{(j)}$$

S. Brenner Miguel et al.

and thus $(\varphi f(\varphi))^{(1)} = -\varphi f(\varphi) - \varphi^2 f^{(1)}(\varphi)$ implies the claim. $\qquad\square$

Proof of Lemma B.2. Since $h \in \mathbb{W}^1$ we have that $H = (\omega h) \circ \varphi$ lies in the Sobolev space of order 1. In equal H is locally absolutely continuous with derivative H' and $H, H' \in \mathbb{L}^2_{\mathbb{R}}$. From this we can conclude that h is locally absolutely continuous. Indeed for $h' := -\omega^{-2}(H' + H) \circ \varphi^{-1} \in \mathbb{L}^1_{\mathrm{loc}}(\mathbb{R}_+)$ and $a, b \in \mathbb{R}_+, a < b$ holds

$$\int_a^b -x^{-2} H'(\varphi^{-1})(x) dx = \int_{\varphi^{-1}(a)}^{\varphi^{-1}(b)} \exp(x) H'(x) dx$$

$$= h(b) - h(a) + \int_a^b x^{-2} H(\varphi^{-1})(x) dx$$

applying the integration by part rule for absolutely continuous function. Further we have that $\|\omega h'\|_\omega = \|\omega^{-1}(H' + H) \circ \varphi^{-1}\|_\omega \leq \|H'\|_{\mathbb{R}} + \|H\|_{\mathbb{R}} < \infty.$ $\qquad\square$

Proof of Lemma B.3. Since $h \in \mathbb{L}^2_{\mathbb{R}_+}(\omega)$ we have for $H := (\omega h) \circ \varphi$ that $\|H\|_{\mathbb{R}} = \|h\|_\omega < \infty$. Further H is locally absolutely continuous with derivative $-\varphi h(\varphi) - \varphi^2 h'(\varphi)$ since for $a, b \in \mathbb{R}$ with $a < b$ holds

$$\int_a^b -\varphi^2(x) h'(\varphi(x)) dx = \int_{\varphi(a)}^{\varphi(b)} x h'(x) dx = [\varphi h \circ \varphi]_a^b - \int_{\varphi(a)}^{\varphi(b)} h(x) dx.$$

Now since $\|\varphi^2 h'(\varphi)\|_{\mathbb{R}} = \|\omega f'\|_\omega < \infty$ we deduce that H is in the Sobolev space of order 1 and thus $(1 + t^2)^{1/2} \mathcal{M}_1[h] = (1 + t^2)^{1/2} \mathcal{F}[H] \in \mathbb{L}^2(\mathbb{C})$ and thus $h \in \mathbb{W}^1$. $\qquad\square$

Proof of Theorem 2.1. Let us define the nested subspaces $(S_k)_{k \in \mathbb{R}_+}$ by $S_k := \{h \in \mathbb{L}^2_{\mathbb{R}_+}(\omega) : \forall |t| \geq k : \mathcal{M}_1[h](t) = 0\}$. For any $h \in S_k$ we consider the empirical contrast

$$\gamma_n(h) := \|h\|_\omega^2 - 2\frac{1}{2\pi} \int_{-\infty}^\infty \widehat{\mathcal{M}}(t) \frac{\mathcal{M}_1[h](-t)}{\mathcal{M}_1[g](t)} dt = \|h\|_\omega^2 - 2n^{-1} \sum_{j=1}^n \nu_h(Y_j)$$

with $\nu_h(Y_j) := \frac{1}{2\pi} \int_{-\infty}^\infty Y_j^{it} \frac{\mathcal{M}_1[h](-t)}{\mathcal{M}_1[g](t)} dt$. One can easily see that $\widehat{f}_k = \arg\min\{\gamma_n(h) : h \in S_k\}$ with $\gamma_n(\widehat{f}_k) = -\|\widehat{f}_k\|_\omega^2$. For $h \in S_k$ define the empirical process $\bar{\nu}_h := n^{-1} \sum_{j=1}^n \nu_h(Y_j) - \langle h, f \rangle_\omega$. Then we have that for $h_1, h_2 \in S_k$ that

$$\gamma_n(h_1) - \gamma_n(h_2) = \|h_1 - f\|_\omega^2 - \|h_2 - f\|_\omega^2 - 2\bar{\nu}_{h_1 - h_2}. \tag{B.1}$$

Now since $\gamma_n(\widehat{f}_k) \leq \gamma_n(f_k)$ and by the definition of $\widehat{k}$ we have that $\gamma_n(\widehat{f}_{\widehat{k}}) - \mathrm{pen}(\widehat{k}) \leq \gamma_n(\widehat{f}_k) - \mathrm{pen}(k) \leq \gamma_n(f_k) - \mathrm{pen}(k)$ for any $k \in \mathcal{K}_n$. Now using Eq. B.1 we get that

$$\|f - \widehat{f}_{\widehat{k}}\|_\omega^2 \leq \|f - f_k\|_\omega^2 + 2\bar{\nu}_{\widehat{f}_{\widehat{k}} - f_k} + \mathrm{pen}(k) - \mathrm{pen}(\widehat{k}).$$

First we note that $S_{k_1} \subseteq S_{k_2}$ for $k_1 \leq k_2$. Let us now denote by $a \vee b := \max(a, b)$ and define for all $k \in \mathcal{K}_n$ the unit balls $B_k := \{h \in S_k : \|h\|_\omega \leq 1\}$. Next we deduce from $2ab \leq a^2 + b^2$ that $2\bar{\nu}_{\widehat{f}_{\widehat{k}} - f_k}^2 \leq 4^{-1}\|\widehat{f}_{\widehat{k}} - f_k\|_\omega^2 + 4\sup_{h \in B_{\widehat{k} \vee k}} \bar{\nu}_h^2$. Further we see that $4^{-1}\|\widehat{f}_{\widehat{k}} - f_k\|_\omega^2 \leq 2^{-1}(\|\widehat{f}_{\widehat{k}} - f\|_\omega^2 + \|f - f_k\|_\omega^2)$. Putting all this together and define

$$p(\widehat{k} \vee k) := 6(2\pi n)^{-1}\Delta_g(\widehat{k} \vee k) \text{ where } \Delta_g(k) := \int_{-k}^{k} |\mathcal{M}_1[g](t)|^{-2}dt \quad \text{(B.2)}$$

we get

$$\|f - \widehat{f}_{\widehat{k}}\|_\omega^2 \leq 3\|f - f_k\|_\omega^2 + 8\Big(\sup_{h \in B_{\widehat{k} \vee k}} \bar{\nu}_h^2 - p(k \vee \widehat{k})\Big)_+ + 8p(\widehat{k} \vee k)$$
$$+ 2\mathrm{pen}(k) - 2\mathrm{pen}(\widehat{k})$$

Assuming now that $\chi \geq 12C_g\pi^{-1}$ we get that $4p(\widehat{k} \vee k) \leq \mathrm{pen}(k) + \mathrm{pen}(\widehat{k})$ and thus

$$\|f - \widehat{f}_{\widehat{k}}\|_\omega^2 \leq 4\big(\|f - f_k\|_\omega^2 + \mathrm{pen}(k)\big) + 8 \max_{k' \in \mathcal{K}_n} \Big(\sup_{h \in B_{k'}} \bar{\nu}_h^2 - p(k')\Big)_+$$

We will use the following lemma which we will be proven afterwards.

Lemma B.4. *Assuming that $\|\omega f_Y\|_\infty < \infty$ and that for all $k \in \mathcal{K}_n$ the function $G_k : \mathbb{R} \to \mathbb{R}, t \mapsto \mathbb{1}_{[-k,k]}(t)|\mathcal{M}_1[g](t)|^{-2}$ is bounded we have*

$$\mathbb{E}_{f_Y}^n \Big(\sup_{h \in B_k} \bar{\nu}_h^2 - p(k)\Big)_+ \leq \frac{C}{n}\left(\|G_k\|_\infty\|\omega f_Y\|_\infty \exp(-\frac{\Delta_g(k)}{12\pi\|\omega f_Y\|_\infty\|G_k\|_\infty}) \right.$$
$$\left. + \frac{\Delta_g(k)}{(2\pi)^2 n} \exp(-\frac{\sqrt{n}}{50}) \right),$$

where Δ_g is defined in Eq. B.2.

Now since $g \in \mathbb{G}_{1,\gamma}$ we have that $\Delta_g(k) \geq c_g k^{2\gamma+1}$ and for all $t \in \mathbb{R}$ holds $|G_k(t)| \leq C_g k^{2\gamma}$ thus we have that the first summand is bounded by $C_g k^{2\gamma}\|\omega f_Y\|_\infty \exp(-\frac{c_g k}{12\pi\|\omega f_Y\|_\infty C_g})$ which is summable over $\mathbb{N}$. For the second summand we use that $n^{-1}\Delta_g(k) \leq C_g n^{-1} k^{2\gamma+1} \leq C_g$ and thus bounded in $\mathbb{N}$. We can therefore deduce that

$$\mathbb{E}_{f_Y}^n \Big(\max_{k' \in \mathcal{K}_n} \Big(\sup_{h \in B_{k'}} \bar{\nu}_h^2 - p(k')\Big)_+\Big) \leq \sum_{k \in \mathcal{K}_n} \mathbb{E}_{f_Y}^n \Big(\Big(\sup_{h \in B_{k'}} \bar{\nu}_h^2 - p(k')\Big)_+\Big)$$
$$\leq C(\|\omega f_Y\|_\infty, g)n^{-1}$$

applying the Lemma. We get that

$$\mathbb{E}_{f_Y}^n (\|f - \widehat{f}_{\widehat{k}}\|_\omega^2) \leq 4\big(\|f - f_k\|_\omega^2 + \mathrm{pen}(k)\big) + C(\|\omega f_Y\|_\infty, g)n^{-1}.$$

Since this inequality holds for all $k \in \mathcal{K}_n$ this implies the claim. $\qquad\square$

Proof of Lemma B.4. We will use the Talagrand inequality Eq. A.3 to show the claim. We want to emphasize that we are able to apply the Talagrand inequality on the sets B_k since B_k has a dense countable subset and due to continuity arguments. To do so we start to determine the constant Ψ^2. We have for any $h \in B_k$ that $\bar{\nu}_h^2 = \langle h, \widehat{f}_k - f_k \rangle_\omega^2 \leq \|h\|_\omega^2 \|\widehat{f}_k - f_k\|_\omega^2$. Since $\|h\|_\omega \leq 1$ we get

$$\mathbb{E}_{f_Y}^n \big(\sup_{h \in B_k} \bar{\nu}_h^2 \big) \leq \mathbb{E}_{f_Y}^n \big(\|\widehat{f}_k - f_k\|_\omega^2 \big) \leq (2n\pi)^{-1} \Delta_g(k) =: \Psi^2.$$

Thus $6\Psi^2 = p(k)$. Next we consider ψ. Let $y > 0$ and $h \in B_k$ then using the Cauchy Schwartz inequality we get $|\nu_h(y)|^2 = (2\pi)^{-2} |\int_{-k}^k y^{it} \frac{\mathcal{M}_1[h](-t)}{\mathcal{M}_1[g](t)} dt|^2 \leq (2\pi)^{-1} \int_{-k}^k |\mathcal{M}_1[g](t)|^{-2} dt \leq (2\pi)^{-1} \Delta_g(k) =: \psi^2$ since $|y^{it}| = 1$ for all $t \in \mathbb{R}$.

Next we consider τ. For $h \in B_k$ we conclude $\mathbb{V}\text{ar}(\nu_h(Y_1)) \leq \mathbb{E}_{f_Y}^n(\nu_h(Y_1)^2) \leq \|\omega f_Y\|_\infty \int_0^\infty y^{-1} \nu_h(y)^2 dt = \|\omega f_Y\|_\infty \|\nu_h\|_{\omega_0}^2$ with $\nu_h(y) = \frac{1}{2\pi} \int_{-k}^k y^{it} \frac{\mathcal{M}_1[h](-t)}{\mathcal{M}_1[g](t)} dt$ for $y > 0$. Thus

$$\|\nu_h\|_{\omega_0}^2 = \frac{1}{2\pi} \int_{-k}^k \left| \frac{\mathcal{M}_1[h](t)}{\mathcal{M}_1[g](t)} \right|^2 dt \leq \frac{\|G_k\|_\infty}{2\pi} \int_{-\infty}^\infty |\mathcal{M}_1[h](t)|^2 dt$$

where $\frac{1}{2\pi} \int_{-\infty}^\infty |\mathcal{M}_1[h](t)|^2 dt = \|h\|_\omega^2 \leq 1$. Thus we set $\tau = \|\omega f_Y\|_\infty \|G_k\|_\infty$. Hence we have that $\frac{n\Psi^2}{6\tau} = \frac{\Delta_g(k)}{12\pi \|\omega f_Y\|_\infty \|G_k\|_\infty}$ and $\frac{n\Psi}{\psi} = \sqrt{n}$. We deduce

$$\mathbb{E}_{f_Y}^n \big(\sup_{h \in B_k} \bar{\nu}_h^2 - p(k) \big)_+ \leq \frac{C}{n} \left(\|G_k\|_\infty \|\omega f_Y\|_\infty \exp(-\frac{\Delta_g(k)}{12\pi \|\omega f_Y\|_\infty \|G_k\|_\infty}) \right.$$
$$\left. + \frac{\Delta_g(k)}{(2\pi)^2 n} \exp(-\frac{\sqrt{n}}{50}) \right). \qquad \square$$

Appendix C: Proofs of Section 3

Proof of Theorem 3.2. First we outline here the main steps of the proof. We will construct a family of functions in $\mathbb{D}_{\mathbb{R}_+}^{s,L}$ by a perturbation of the density $f_o : \mathbb{R}_+ \to \mathbb{R}_+$ with small bumps, such that their $\mathbb{L}_{\mathbb{R}_+}^2(\omega)$-distance and the Kullback-Leibler divergence of their induced distributions can be bounded from below and above, respectively. The claim follows then by applying Theorem 2.5 in [14]. We use the following construction, which we present first.

Denote by $C_c^\infty(\mathbb{R})$ the set of all smooth functions with compact support in $\mathbb{R}$ and let $\psi \in C_c^\infty(\mathbb{R})$ be a function with support in $[0,1]$ and $\int_0^1 \psi(x)dx = 0$. For each $K \in \mathbb{N}$ (to be selected below) and $k \in [\![0, K[\![$ we define the bump-functions $\psi_{k,K}(x) := \psi(xK - K - k)$, $x \in \mathbb{R}$. and define for $j \in \mathbb{N}_0$ the finite constant $C_{j,\infty} := \max(\|\psi^{(l)}\|_\infty, l \in [\![0,j]\!])$. Let us further define the operator $\mathcal{S} : C_c^\infty(\mathbb{R}) \to C_c^\infty(\mathbb{R})$ with $\mathcal{S}[f](x) = xf^{(1)}(x)$ for all $x \in \mathbb{R}$ and define $\mathcal{S}^1 := \mathcal{S}$ and $\mathcal{S}^n := \mathcal{S} \circ \mathcal{S}^{n-1}$ for $n \in \mathbb{N}, n \geq 2$. Now, for $j \in \mathbb{N}$, we define the function $\psi_{k,K,j}(x) := \mathcal{S}^j[\psi_{k,K}](x) = \sum_{i=1}^j c_{i,j} x^i K^i \psi^{(i)}(xK - K - k)$ for $x \in \mathbb{R}_+$ and $c_{i,j} \geq 1$ and let $c_j := \sum_{i=1}^j c_{i,j}$

For a bump-amplitude $\delta > 0, \gamma \in \mathbb{N}$ and a vector $\boldsymbol{\theta} = (\theta_1, \ldots, \theta_K) \in \{0, 1\}^K$ we define

$$f_{\boldsymbol{\theta}}(x) = f_o(x) + \delta K^{-s-\gamma} \sum_{k=0}^{K-1} \theta_{k+1} \psi_{k,K,\gamma}(x) \text{ where } f_o(x) := \exp(-x). \quad \text{(C.1)}$$

Until now, we did not give a sufficient condition to ensure that our constructed functions $\{f_{\boldsymbol{\theta}} : \boldsymbol{\theta} \in \{0, 1\}^K\}$ are in fact densities. This condition is given by the following lemma.

Lemma C.1. *Let* $0 < \delta < \delta_o(\psi, \gamma) := \exp(-2)2^{-\gamma}(C_{\gamma,\infty} c_\gamma)^{-1}$. *Then for all* $\boldsymbol{\theta} \in \{0, 1\}^K$, $f_{\boldsymbol{\theta}}$ *is a density.*

Further one can show that these densities all lie inside the ellipsoids $\mathbb{D}_{\mathbb{R}_+}^{s,L}$ for L big enough. This is captured in the following lemma.

Lemma C.2. *Let* $s \in \mathbb{N}$. *Then, there is* $L_{s,\gamma,\delta} > 0$ *such that* f_o *and any* $f_{\boldsymbol{\theta}}$ *as in Eq. C.1 with* $\boldsymbol{\theta} \in \{0, 1\}^K$, $K \in \mathbb{N}$, *belong to* $\mathbb{D}_{\mathbb{R}_+}^{s,L_{s,\gamma,\delta}}$.

For sake of simplicity we denote for a function $\varphi \in \mathbb{L}_{\mathbb{R}_+}^2$ the multiplicative convolution with g by $\widetilde{\varphi} := [\varphi * g]$. Futher we see that for $y_2 \geq y_1 > 0$ holds

$$\widetilde{f}_o(y_1) = \int_0^\infty g(x) x^{-1} \exp(-y_1/x) dx \geq \int_0^\infty g(x) x^{-1} \exp(-y_2/x) dx = \widetilde{f}_o(y_2)$$

and thus $\widetilde{f}_o$ is monotone decreasing. Further we have that $\widetilde{f}_o(2) > 0$ since otherwise $g = 0$ almost everywhere. Exploiting *Varshamov-Gilbert's lemma* (see [14]) in Lemma C.3 we show further that there is $M \in \mathbb{N}$ with $M \geq 2^{K/8}$ and a subset $\{\boldsymbol{\theta}^{(0)}, \ldots, \boldsymbol{\theta}^{(M)}\}$ of $\{0, 1\}^K$ with $\boldsymbol{\theta}^{(0)} = (0, \ldots, 0)$ such that for all $j, l \in [\![0, M]\!]$, $j \neq l$ the $\mathbb{L}_{\mathbb{R}_+}^2(\omega)$-distance and the Kullback-Leibler divergence are bounded for $K \geq K_o(\gamma, \psi)$.

Lemma C.3. *Let* $K \geq K_o(\psi, \gamma) \vee 8$. *Then there exists a subset* $\{\boldsymbol{\theta}^{(0)}, \ldots, \boldsymbol{\theta}^{(M)}\}$ *of* $\{0, 1\}^K$ *with* $\boldsymbol{\theta}^{(0)} = (0, \ldots, 0)$ *such that* $M \geq 2^{K/8}$ *and for all* $j, l \in [\![0, M]\!]$, *with* $j \neq l$ *holds* $\|f_{\boldsymbol{\theta}^{(j)}} - f_{\boldsymbol{\theta}^{(l)}}\|_\omega^2 \geq \frac{\|\psi^{(\gamma)}\|^2 \delta^2}{16} K^{-2s}$ *and further* $KL(\widetilde{f}_{\boldsymbol{\theta}^{(j)}}, \widetilde{f}_{\boldsymbol{\theta}^{(0)}}) \leq \frac{C_1(g)\|\psi\|^2}{\widetilde{f}_o(2) \log(2)} \delta^2 \log(M) K^{-2s-2\gamma-1}$ *where* KL *is the Kullback-Leibler-divergence.*

Selecting $K = \lceil n^{1/(2s+2\gamma+1)} \rceil$, it follows

$$\frac{1}{M} \sum_{j=1}^M KL((\widetilde{f}_{\boldsymbol{\theta}^{(j)}})^{\otimes n}, (\widetilde{f}_{\boldsymbol{\theta}^{(0)}})^{\otimes n}) = \frac{n}{M} \sum_{j=1}^M KL(\widetilde{f}_{\boldsymbol{\theta}^{(j)}}, \widetilde{f}_{\boldsymbol{\theta}^{(0)}}) \leq C_{\psi,\delta,g,\gamma,f_o}^{(2)} \log(M)$$

where $C_{\psi,\delta,g,\gamma,f_o} < 1/8$ for all if $\delta \leq \delta_1(\psi, g, \gamma, f_o)$ and $M \geq 2$ for $n \geq n_{s,\gamma} := 8^{2s+1} \vee K_o(\gamma, \psi)^{2s+2\gamma+1}$. Thereby, we can use Theorem 2.5 of [14], which in turn for any estimator $\widehat{f}$ of f implies

$$\sup_{f \in \mathbb{D}_{\mathbb{R}_+}^{s,L}} \mathbb{P}\big(\|\widehat{f} - f\|_\omega^2 \geq \frac{C_{\psi,\delta,\gamma}^{(1)}}{2} n^{-\frac{2s}{2s+2\gamma+1}}\big) \geq \frac{\sqrt{M}}{1 + \sqrt{M}}\big(1 - 1/4 - \sqrt{\frac{1}{4\log(M)}}\big) \geq 0.07.$$

Note that the constant $C^{(1)}_{\psi,\delta,\gamma}$ does only depend on ψ, γ and δ, hence it is independent of the parameters s, L and n. The claim of Theorem 3.2 follows by using Markov's inequality, which completes the proof. □

Proofs of the lemmata

Proof of Lemma C.1. For any $h \in C^\infty_c(\mathbb{R})$ we can state that $\int_{-\infty}^{\infty} \mathcal{S}[h](x)dx = [xh(x)]_{-\infty}^{\infty} - \int_{-\infty}^{\infty} h(x)dx = -\int_{-\infty}^{\infty} h(x)dx$ and therefore $\int_{-\infty}^{\infty} \mathcal{S}^j[h](x)dx = (-1)^j \int_{-\infty}^{\infty} h(x)dx$ for $j \in \mathbb{N}$. Thus $\int_{-\infty}^{\infty} \psi_{k,K,\gamma}(x)dx = (-1)^\gamma \int_{-\infty}^{\infty} \psi_{k,K}(x)dx = 0$ which implies that for any $\delta > 0$ and $\boldsymbol{\theta} \in \{0,1\}^K$ we have $\int_0^\infty f_{\boldsymbol{\theta}}(x)dx = 1$.

Now due to the construction Eq. C.1 of the functions $\psi_{k,K}$ we easily see that the function $\psi_{k,K}$ has support on $[1 + k/K, 1 + (k+1)/K]$ which lead to $\psi_{k,K}$ and $\psi_{l,K}$ having disjoint supports if $k \neq l$. Here, we want to emphasize that $\mathrm{supp}(\mathcal{S}[h]) \subseteq \mathrm{supp}(h)$ for all $h \in C^\infty_c(\mathbb{R})$. Which implies that $\psi_{k,K,\gamma}$ and $\psi_{l,K,\gamma}$ have disjoint supports if $k \neq l$, too. For $x \in [1,2]^c$ we have $f_{\boldsymbol{\theta}}(x) = \exp(-x) \geq 0$. Now let us consider the case $x \in [1,2]$. In fact there is $k_o \in [\![0,K[\![$ such that $x \in [1 + k_o/K, 1 + (k_o+1)/K]$ and hence

$$f_{\boldsymbol{\theta}}(x) = f_o(x) + \theta_{k_o+1}\delta K^{-s-\gamma}\psi_{k_o,K,\gamma}(x) \geq \exp(-2) - \delta 2^\gamma C_{\gamma,\infty}c_\gamma$$

since $\|\psi_{k,K,j}\|_\infty \leq 2^j C_{j,\infty}c_j K^j$ for any $k \in [\![0,K[\![$ and $j \in \mathbb{N}$ where $c_j := \sum_{i=1}^{j} c_{i,j}$. Now choosing $\delta \leq \delta_o(\psi,\gamma) = \exp(-2)2^{-\gamma}(C_{\gamma,\infty}c_\gamma)^{-1}$ ensures $f_{\boldsymbol{\theta}}(x) \geq 0$ for all $x \in \mathbb{R}_+$. □

Proof of Lemma C.2. Our proof starts with the observation that for all $t \in \mathbb{R}$ we have $\mathcal{M}_1[f_o](t) = \Gamma(1 + it)$. Now by applying the Stirling formula (see also [5]) we get $|\Gamma(1 + it)| \sim |t|^{1/2}\exp(-\pi/2|t|)$, $|t| \geq 2$, thus for every $s \in \mathbb{N}$ there exists L_s such that $|f_o|_s^2 \leq L$ for all $L \geq L_s$.

Next we consider $|f_o - f_{\boldsymbol{\theta}}|_s$. Let us therefore define first $\Psi_K := \sum_{k=0}^{K-1} \theta_{k+1}\psi_{k,K}$ and $\Psi_{K,j} := \mathcal{S}^j[\Psi_K]$ for an $j \in \mathbb{N}$. Then we have $|f_o - f_{\boldsymbol{\theta}}|_s^2 = \delta^2 K^{-2s-2\gamma}|\Psi_{K,\gamma}|_s^2$ where $|.|_s$ is defined in Eq. 3.1. Now since for any $j \in \mathbb{N}$, it holds that $\mathrm{supp}(\Psi_{K,j}) \subset [1,2], \|\Psi_{K,j}\|_\infty < \infty$ we have that $(0,\infty)$ is a subset of the strip of analyticity of $\Psi_{K,j}$. By application of Eq. A.1 we deduce that $|\mathcal{M}_1[\Psi_{K,s+\gamma}](t)|^2 = (1 + t^2)^s|\mathcal{M}_1[\Psi_{K,\gamma}](t)|^2$ and thus

$$|\Psi_{K,\gamma}|_s^2 = \int_{-\infty}^{\infty} |\mathcal{M}_1[\Psi_{K,s+\gamma}](t)|^2 dt = 2\pi \int_0^\infty x|\Psi_{K,s+\gamma}(x)|^2 dx$$

by the Parseval formula. Since $\psi_{k,K}$ have disjoint support for different values of k we follow that $|\Psi_{k,\gamma}|_s^2 = 2\pi \sum_{k=0}^{K-1} \theta_{k+1}^2 \int_0^\infty x|\mathcal{S}^{\gamma+s}[\psi_{k,K}](x)|^2 dx$. Applying the Jensen inequality and the fact that $\mathrm{supp}(\psi_{k,K}) \subset [1,2]$ leads to

$$|\Psi_{k,\gamma}|_s^2 \leq 2\pi 2^{\gamma+s-1} \sum_{k=0}^{K-1} \sum_{j=1}^{\gamma+s} c_{j,\gamma+s}^2 \int_1^2 x^{2j+1}K^{2j}\psi^{(j)}(xK - K - k)^2 dx$$

$$\leq 2\pi K^{2(\gamma+s)}2^{\gamma+s} \sum_{k=0}^{K-1} \sum_{j=1}^{\gamma+s} c_{j,\gamma+s}^2 4^j C_{\psi,s,\gamma}^2 K^{-1} \leq C_{(\gamma,s)}K^{2(\gamma+s)}$$

Thus $|f_o - f_{\boldsymbol{\theta}}|_s^2 \le C_{(s,\gamma,\delta)}$ and $|f_{\boldsymbol{\theta}}|_s^2 \le 2(|f_o - f_{\boldsymbol{\theta}}|_s^2 + |f_o|_s^2) \le 2(C_{(s,\gamma,\delta)} + L_s) =: L_{s,\gamma,\delta}$. $\qquad\square$

Proof of Lemma C.3. Using that the functions $(\psi_{k,K,\gamma})_{k\in[\![0,K[\![}$ with different index k have disjoint supports we get

$$\|f_{\boldsymbol{\theta}} - f_{\boldsymbol{\theta}'}\|_{\omega}^2 = \delta^2 K^{-2s-2\gamma} \| \sum_{k=0}^{K-1} (\theta_{k+1} - \theta'_{k+1}) \psi_{k,K,\gamma} \|_{\omega}^2$$

$$= \delta^2 K^{-2s-2\gamma} \rho(\boldsymbol{\theta}, \boldsymbol{\theta}') \|\psi_{0,K,\gamma}\|_{\omega}^2$$

with $\rho(\boldsymbol{\theta}, \boldsymbol{\theta}') := \sum_{j=0}^{K-1} \mathbb{1}_{\{\boldsymbol{\theta}_{j+1} = \boldsymbol{\theta}'_{j+1}\}}$ the *Hamming distance*. Now the first claim follows by showing that by $\|\psi_{0,K,\gamma}\|_{\omega}^2 \ge \frac{K^{2\gamma-1}\|\psi^{(\gamma)}\|^2}{2}$ for K big enough. To do so we observe that $\|\psi_{0,K,\gamma}\|_{\omega}^2 = \sum_{i,j\in[\![1,\gamma]\!]} c_{j,\gamma} c_{i,\gamma} \int_0^{\infty} x^{j+i+1} \psi_{0,K}^{(j)}(x) \psi_{0,K}^{(i)}(x) dx$ and define $\Sigma := \|\psi_{0,K,\gamma}\|_{\omega}^2 - \int_0^{\infty} (x^{\gamma} \psi_{0,K}^{(\gamma)}(x))^2 x \, dx$

$$\|\psi_{0,K,\gamma}\|_{\omega}^2 = \Sigma + \int_0^{\infty} (x^{\gamma} \psi_{0,K}^{(\gamma)}(x))^2 x \, dx \ge \Sigma + K^{2\gamma-1}\|\psi^{(\gamma)}\|^2 \ge \frac{K^{2\gamma-1}\|\psi^{(\gamma)}\|^2}{2} \tag{C.2}$$

as soon as $|\Sigma| \le \frac{K^{2\gamma-1}\|\psi^{(\gamma)}\|^2}{2}$. This is obviously true as soon as $K \ge K_o(\gamma, \psi)$ and thus $\|f_{\boldsymbol{\theta}} - f_{\boldsymbol{\theta}'}\|_{\omega}^2 \ge \frac{\delta^2 \|\psi^{(\gamma)}\|^2}{2} K^{-2s-1} \rho(\boldsymbol{\theta}, \boldsymbol{\theta}')$ for $K \ge K_o(\psi, \gamma)$.

Now we use the *Varshamov-Gilbert Lemma* (see [14]) which states that for $K \ge 8$ there existes a subset $\{\boldsymbol{\theta}^{(0)}, \dots, \boldsymbol{\theta}^{(M)}\}$ of $\{0,1\}^K$ with $\boldsymbol{\theta}^{(0)} = (0, \dots, 0)$ such that $\rho(\boldsymbol{\theta}^{(j)}, \boldsymbol{\theta}^{(k)}) \ge K/8$ for all $j, k \in [\![0, M]\!], j \ne k$ and $M \ge 2^{K/8}$. Applying this leads to $\|f_{\boldsymbol{\theta}^{(j)}} - f_{\boldsymbol{\theta}^{(l)}}\|_{\omega}^2 \ge \frac{\|\psi^{(\gamma)}\|^2 \delta^2}{16} K^{-2s}$.

For the second part we have $f_o = f_{\boldsymbol{\theta}^{(0)}}$ and by using $\mathrm{KL}(\tilde{f}_{\boldsymbol{\theta}}, \tilde{f}_o) \le \chi^2(\tilde{f}_{\boldsymbol{\theta}}, \tilde{f}_o) := \int_{\mathbb{R}_+} (\tilde{f}_{\boldsymbol{\theta}}(x) - \tilde{f}_o(x))^2 / \tilde{f}_o(x) dx$ it is sufficient to bound the χ-squared divergence. We notice that $\tilde{f}_{\boldsymbol{\theta}} - \tilde{f}_o$ has support in $[0,2]$ since $f_{\boldsymbol{\theta}} - f_o$ has support in $[1,2]$ and g has support in $[0,1]$ In fact for $y > 2$ holds $\tilde{f}_{\boldsymbol{\theta}}(y) - \tilde{f}_o(y) = \int_y^{\infty} (f_{\boldsymbol{\theta}} - f_o)(x) x^{-1} g(y/x) dx = 0$. Denote further $\Psi_{K,\gamma} := \sum_{k=0}^{K-1} \theta_{k+1} \psi_{k,K,\gamma} = \mathcal{S}^{\gamma}[\sum_{k=0}^{K-1} \theta_{k+1} \psi_{k,K}] =: \mathcal{S}^{\gamma}[\Psi_K]$. Now by using the compact support property and a single substitution we get

$$\chi^2(\tilde{f}_{\boldsymbol{\theta}}, \tilde{f}_o) \le \tilde{f}_o(2)^{-1} \|\tilde{f}_{\boldsymbol{\theta}} - \tilde{f}_o\|^2 = \tilde{f}_o(2)^{-1} \delta^2 K^{-2s-2\gamma} \|\widetilde{\Psi}_{K,\gamma}\|^2.$$

Let us now consider $\|\widetilde{\Psi}_{K,\gamma}\|^2$. In the first step we see by application of the Parseval that $\|\widetilde{\Psi}_{K,\gamma}\|^2 = \frac{1}{2\pi} \int_{-\infty}^{\infty} |\mathcal{M}_{1/2}[\widetilde{\Psi}_{K,y}](t)|^2 dt$. Now for $t \in \mathbb{R}$, we see by using the convolution theorem for Mellin transforms that $\mathcal{M}_{1/2}[\widetilde{\Psi}_{K,\gamma}](t) = \mathcal{M}_{1/2}[g](t) \cdot \mathcal{M}_{1/2}[\mathcal{S}^{\gamma}[\Psi_K]](t)$. Again $\mathcal{M}_{1/2}[\mathcal{S}^{\gamma}[\Psi_K]](t) = (1/2 + it)^{\gamma} \mathcal{M}_{1/2}[\Psi_K](t)$. Together with assumption **[G1']** we get

$$\|\widetilde{\Psi}_{K,\gamma}\|^2 \le \frac{C_1(g)}{2\pi} \int_{-\infty}^{\infty} |\mathcal{M}_{1/2}[\Psi_K](t)|^2 dt = C_1(g) \|\Psi_K\|^2 \le C_1(g) \|\psi\|^2.$$

Since $M \geq 2^K$ we have thus $\mathrm{KL}(\widetilde{f}_{\boldsymbol{\theta}^{(j)}}, \widetilde{f}_{\boldsymbol{\theta}^{(0)}}) \leq \frac{C_1(g)\|\psi\|^2}{\widetilde{f}_o(2)\log(2)}\delta^2 \log(M)K^{-2s-2\gamma-1}.$ $\quad\square$

References

[1] ANDERSEN, K. E. and HANSEN, M. B. (2001). Multiplicative censoring: density estimation by a series expansion approach. *J. Statist. Plann. Inference* **98** 137–155. MR1860231

[2] ASGHARIAN, M. and WOLFSON, D. B. (2005). Asymptotic behavior of the unconditional NPMLE of the length-biased survivor function from right censored prevalent cohort data. *Ann. Statist.* **33** 2109–2131. MR2211081

[3] BARUCQ, H., MATTESI, V. and TORDEUX, S. (2015). The Mellin Transform. **RR-8743** 2.

[4] BELOMESTNY, D., COMTE, F. and GENON-CATALOT, V. (2016). Nonparametric Laguerre estimation in the multiplicative censoring model. *Electron. J. Stat.* **10** 3114–3152. MR3571964

[5] BELOMESTNY, D. and GOLDENSHLUGER, A. (2020). Nonparametric density estimation from observations with multiplicative measurement errors. *Ann. Inst. Henri Poincaré Probab. Stat.* **56** 36–67. MR4058981

[6] BRENNER MIGUEL, S. and JOHANNES, J. (2020). Data-driven aggregation in non-parametric density estimation on the real line. arXiv: 2001.10910

[7] BRUNEL, E., COMTE, F. and GENON-CATALOT, V. (2016). Nonparametric density and survival function estimation in the multiplicative censoring model. *TEST* **25** 570–590. MR3531843

[8] COHN, D. L. (2013). *Measure theory*, second ed. *Birkhäuser Advanced Texts: Basler Lehrbücher. [Birkhäuser Advanced Texts: Basel Textbooks]*. Birkhäuser/Springer, New York. MR3098996

[9] COMTE, F. and DION, C. (2016). Nonparametric estimation in a multiplicative censoring model with symmetric noise. *J. Nonparametr. Stat.* **28** 768–801. MR3555457

[10] ENGL, H. W., HANKE, M. and NEUBAUER, A. (1996). *Regularization of inverse problems. Mathematics and its Applications* **375**. Kluwer Academic Publishers Group, Dordrecht. MR1408680

[11] FAN, J. (1991). On the optimal rates of convergence for nonparametric deconvolution problems. *Ann. Statist.* **19** 1257–1272. MR1126324

[12] MEISTER, A. (2009). *Deconvolution problems in nonparametric statistics. Lecture Notes in Statistics* **193**. Springer-Verlag, Berlin. MR2768576

[13] PARIS, R. B. and KAMINSKI, D. (2001). *Asymptotics and Mellin-Barnes integrals. Encyclopedia of Mathematics and its Applications* **85**. Cambridge University Press, Cambridge. MR1854469

[14] TSYBAKOV, A. B. (2009). *Introduction to nonparametric estimation. Springer Series in Statistics*. Springer, New York Revised and extended from the 2004 French original, Translated by Vladimir Zaiats. MR2724359

[15] VAN ES, B., KLAASSEN, C. A. J. and OUDSHOORN, K. (2000). Survival analysis under cross-sectional sampling: length bias and multiplicative

censoring. **91** 295–312. Prague Workshop on Perspectives in Modern Statistical Inference: Parametrics, Semi-parametrics, Non-parametrics (1998). MR1814785

[16] VARDI, Y. (1989). Multiplicative censoring, renewal processes, deconvolution and decreasing density: nonparametric estimation. *Biometrika* **76** 751–761. MR1041420

[17] VARDI, Y. and ZHANG, C.-H. (1992). Large sample study of empirical distributions in a random-multiplicative censoring model. *Ann. Statist.* **20** 1022–1039. MR1165604