MULTIPLICATIVE CENSORING: DENSITY ESTIMATION BY A SERIES EXPANSION APPROACH

KIM E. ANDERSEN^{\dagger} AND MARTIN B. HANSEN^{\dagger}

time distributions of renewal processes. The purpose of this paper is to present an approach to this of multiplicatively censored random variables. This is a problem arising in e.g. estimation of waiting artificial data example. expansion is suggested. converges to zero for increasing sample size. density. problem using a singular value decomposition or orthonormal based series expansion of the desired Abstract. We consider the linear inverse problem of recovering the density function for a sample We establish conditions under which the mean integrated square error of the estimator Finite sample properties of the estimation procedure are studied on an An empirical method for determining the order of

of convergence; least squares cross-validation, finite sample study Key words. Multiplicative censoring; linear inverse problem; singular value decomposition; rate

www.maphysto.dk

AMS subject classifications. Primary, 62G05; Secondary, 62C20 65R20 65U05

1. Introduction. Let Y_1, \ldots, Y_n be a random sample from a density, g, supported on $(0, \infty)$. Assume that $Y_i = U_i X_i$, where U_1, \ldots, U_n is a random sample from to estimate the unknown density f from a random sample of Y_i 's. the uniform distribution on (0, 1) and X_1, \ldots, X_n is a random sample from a density, f, supported on $(0,\infty)$. In the present paper we will use a series expansion approach

manner It is straightforward to derive that density g can be expressed by f in the following

(1.1)
$$g(y) = \int_{y}^{\infty} \frac{f(x)}{x} dx$$

sense. Assume we have a random sample Y_1, \ldots, Y_n from a density g which is related to another density f by g = Kf, where K is a linear operator. Estimate the density Hence the problem is seen to be a statistical inverse problem in the following

decreasing densities. next replacement. If we let g denote the density of the sampled forward recurrence at an arbitrary time far from installation and then each process is surveyed until the processes. Here we think of n independent stationary renewal processes with waiting renewal processes, certain non-parametric deconvolution problems, and estimation of unifies several well studied statistical problems, including non-parametric inference for time distributions which have density f_0 . Assume that the processes are inspected Vardi [26] termed the model above multiplicative censoring and showed how it Consider, for example, non-parametric estimation in renewal

Ø, Denmark. [†]Department of Mathematical Sciences, Aalborg University, Fredrik Bajers Vej 7E, 9220 Aalborg

National Research Foundation. ‡ MaPhySto, Centre for Mathematical Physics and Stochastics, funded by a grant from the Danish

times. Then, one can prove (see e.g. [5, p. 63]) that g and f_0 are related by

(1.2)
$$g(y) = \frac{\int_{y}^{\infty} f_0(x) \, dx}{\int_{0}^{\infty} z f_0(z) \, dz}.$$

Substituting

$$f(x) = \frac{xf_0(x)}{\int_0^\infty zf_0(z)\,dz},$$

into (1.2) we see that g and f are related by (1.1). Once obtaining an estimate \hat{f} of f one can transform back to an estimate \hat{f}_0 of f_0 by

$$\hat{f}_0(x) = rac{\hat{f}(x)/x}{\int_0^\infty \hat{f}(z)/z \, dz}.$$

The reason for this relation is we are sampling with a length bias. Length biased sampling of forward recurrence times arises naturally in many applications. Examples can be found in [6, 20]. The application we have in mind, which have a sample size where non-parametric density estimation is sensible, is due to Støvring and Vach [24]. In OPED, Odense PharmaEpidemiologic Database, prescription redemptions are registered for the county of Funen, Denmark (approximately 450000 inhabitants). Connected to this study it is of interest to estimate the waiting time distribution of redemptions of a certain drug based on the observed time to the first redemption.

A tempting approach for estimating f would be to make an estimate \hat{g} of g by standard methods and then use $\hat{f} = K^{-1}\hat{g}$ as an estimate of f. However, this could be a hazardous procedure as in many interesting scientific applications (including the present) K^{-1} is not a bounded linear operator. Whence, even small perturbations of \hat{g} may result in large distortions of \hat{f} . Such inverse problems are often termed ill-posed. For mathematical and statistical perspectives on ill-posedness see [16] and [19], respectively.

A vast literature exists on solutions to these types of problems. The most eyecatching is off-course to assume that f belongs to a flexible parametric family of densities and then estimate the parameters by maximum likelihood or a Bayesian procedure. However, lack of knowledge and high irregularity of the underlying density f can make this procedure infeasible and it is tempting to turn to more flexible nonparametric methods. One approach is to use non-parametric maximum likelihood estimation (NPMLE). This was suggested for the present problem in [26]. For recent references see [10, 11]. An alternative to NPMLE would be to use a non-parametric Bayesian approach as suggested in [4].

Different routes are to use a kernel estimate or a series expansion of the desired density f. For a review of these ideas up to 1985 see [19]. The theory has since then developed in different directions. A direction (which is taken in the present paper) is to expand f based on a singular value decomposition (SVD) of K. In the

statistics context this was popularized by [14, 15]. For recent contributions drawing on more general spectral theory for bounded operators cf. [7, 18, 25]. The most recent direction being applying wavelets as basis functions in the reconstruction, see e.g. [1] and [8]. From non-parametric function estimation Koo and Chung [17] imported the idea of applying log-density estimation in conjunction with maximum indirect likelihood estimation (MILE) to inverse problems.

An overview of the general theory behind SVD based estimators is given in Section 2. In Section 3 an SVD of the operator K is derived and two truncated SVD estimators are formulated. We justify the estimator in Section 4 by studying asymptotic convergence properties. A data-driven method for choosing the regularization parameter is suggested in Section 5. In order to get some idea of the finite sample properties of the estimators suggested a simulation study is conducted in Section 6. Finally, a brief description of some open problems is given in Section 7.

2. General theory. In our treatment of the problem, we assume f has bounded support. By redefining the units of measurements if necessary, we assume without loss of generality that f and hence g are supported on (0, 1).

Throughout the paper let f and g be densities with respect to Lebesgue measure on (0,1) and $\mathcal{F} = L^2_{\mu}(0,1)$ and $\mathcal{G} = L^2_{\nu}(0,1)$ be Hilbert spaces of functions which are square integrable with respect to the absolutely continuous measures μ and ν , respectively.

Consider now the following general statistical inverse problem. Assume we have a random sample Y_1, \ldots, Y_n from a density g which is related to another density f by

$$(2.1) g = Kf$$

where $K : \mathcal{F} \to \mathcal{G}$ is a linear operator. Estimate the density f.

If we assume K is bounded we can consider the selfadjoint operator $L = K^*K$, where K^* is the adjoint of K. Additionally, if L is compact, we know K possesses an SVD (see [16, Appendix A.5]). That is, there exist an ordered sequence of positive numbers $\mu_1 \ge \mu_2 \ge \ldots > 0$ (called singular values) and orthonormal systems $(f_j) \subset \mathcal{F}$ and $(g_j) \subset \mathcal{G}$ such that

(2.2)
$$Kf_j = \mu_j g_j$$

(2.3)
$$K^*g_j = \mu_j f_j,$$

and

(2.4)
$$f = f_0 + \sum_{j=1}^{\infty} \mu_j^{-1}(g, g_j) f_j,$$

for some $f_0 \in \mathcal{N}(K)$, where $\mathcal{N}(K)$ denotes the null-space of K. Equation (2.1) is uniquely solvable when g satisfies the Picard conditions $\sum_{j=1}^{\infty} |(g,g_j)|/\mu_j^2 < \infty$ and $g \in \mathcal{N}(K^*)^{\perp}$ and in this case

$$f = \sum_{j=1}^{\infty} \mu_j^{-1}(g,g_j) f_j.$$

Actually the singular values can be expressed by $\mu_j = \sqrt{\lambda_j}$, where λ_j is the j'th eigenvalue of the selfadjoint compact operator $L : \mathcal{F} \to \mathcal{F}$.

We now use that given a sample Y_1, \ldots, Y_n from g we can approximate the inner product (g, g_j) by

(2.5)
$$(g,g_j) = \int_0^1 g_j(y)g(y)\nu(dy) \simeq n^{-1} \sum_{i=1}^n g_j(Y_i) \frac{d\nu(Y_i)}{dy},$$

and thereby if g satisfies the Picard conditions obtain the following SVD reconstruction formula

(2.6)
$$\hat{f}_n(x) = \sum_{j=1}^{\infty} \mu_j^{-1} \left[n^{-1} \sum_{i=1}^n g_j(Y_i) \frac{d\nu(Y_i)}{dy} \right] f_j(x).$$

Normally, the singular functions f_j and g_j oscillates more and more for decreasing values of μ_j and we would therefore expect the approximation in (2.5) to be bad for high values of j. This is amplified in the reconstruction (2.6). However, if we assume that only the coefficients (g, g_j) early in the expansion are important we can choose to down weight the contribution from μ_j^{-1} . We now introduce the following regularization family

$$\phi_{\alpha}(t) = \frac{1}{t} \mathbf{1}(t \ge \alpha), \ t > 0, \ \alpha > 0,$$

which is used to construct the following windowed SVD reconstruction formula

(2.7)
$$\hat{f}_n(x) = \sum_{j=1}^{\infty} \phi_\alpha(\mu_j) \left[n^{-1} \sum_{i=1}^n g_j(Y_i) \frac{d\nu(Y_i)}{dy} \right] f_j(x)$$

(2.8)
$$= \sum_{j:\mu_j \ge \alpha} \mu_j^{-1} \left[n^{-1} \sum_{i=1}^n g_j(Y_i) \frac{d\nu(Y_i)}{dy} \right] f_j(x).$$

Thus, ϕ_{α} is seen to cutoff big values of μ_j^{-1} and is therefore termed spectral cutoff. Hence (2.7) is the spectral cutoff solution.

Remark 2.1. In regularization theory [16, Chapter 2] a regularization family is a family of linear bounded operators $R_{\alpha} : \mathcal{G} \to \mathcal{F}, \alpha > 0$ such that $\lim_{\alpha \to 0} R_{\alpha}Kf = f$ for all $f \in \mathcal{F}$, i.e. the operator $R_{\alpha}K$ converges pointwise to the identity. If we put $R_{\alpha}g = \sum_{j=1}^{\infty} \phi_{\alpha}(\mu_{j})(g,g_{j})f_{j}$ we immediately obtain that the cutoff scheme is a regularization family. For further properties of the cutoff scheme cf. [16, Chapter 2] and [25].

Remark 2.2. There exists a number of other possibilities for regularization families, e.g. $\phi_{\alpha}(x) = 1/(\alpha + x)$, x > 0, $\alpha > 0$. This corresponds to minimizing the Tikhonov functional $J_{\alpha}(f) = ||Kf - g||_{\nu} + \alpha ||f||_{\mu}^{2}$, see for instance [16, Section 2.2]. But in the present paper we deal exclusively with the former as it is easy to deal with in asymptotics and empirical choices of the α -value. Remark 2.3. The possibility of negativeness with SVD-based estimators may be resolved by projection in $L^2_{\mu}(0,1)$, that is $\tilde{f}_n(x) = \max\{\hat{f}_n(x), 0\}$, see e.g. [9, Remark 2.3].

3. SVD of the operator.

3.1. Estimation based on Bessel functions. In this section we introduce two estimators for the unknown f based on SVD. First, we let $\mathcal{F} = L^2_{\mu}(0,1)$ denote the space of functions on (0,1) which are square integrable with respect to the dominating measure $d\mu(x) = x^{-1} dx$ and $\mathcal{G} = L^2_{\nu}(0,1)$ the space of functions on (0,1) which are square integrable with respect to Lebesgue measure, ν . Hence, the operator $K: \mathcal{F} \to \mathcal{G}$ can be expressed by

$$Kf(y) = \int_{y}^{1} f(x) d\mu(x), \qquad 0 < y < 1.$$

In order to achieve an SVD for operator K we first show K^*K is compact. First of all notice that

$$Kf(y) = \int_0^1 \mathbf{1}(y \le x) f(x) \, d\mu(x)$$

and therefore by Cauchy-Schwartz's inequality

$$\begin{split} \|Kf\|^2 &= \int_0^1 |Kf(y)|^2 \, dy \\ &\leq \int_0^1 \int_0^1 \mathbf{1}(y \le x) \, d\mu(x) \int_0^1 f(x)^2 \, d\mu(x) \, dy \\ &= -\|f\|_{\mu}^2 \int_0^1 \ln(y) \, dy \\ &= \|f\|_{\mu}^2. \end{split}$$

This shows us that $Kf \in \mathcal{G}$ for all $f \in \mathcal{F}$ and that K is a bounded linear map.

Furthermore, consider the self-adjoint operator $L = K^*K : \mathcal{F} \to \mathcal{F}$ which by Fubini's Theorem can be expressed by

$$Lf(z) = \int_0^z \int_y^1 f(x) \, d\mu(x) \, dy$$

= $\int_0^1 f(x) \int_0^1 \mathbf{1}(x \ge y) \mathbf{1}(y \le z) \, dy \, d\mu(x)$

Now, let $k(x,z) = \int_0^1 \mathbf{1}(y \le x) \mathbf{1}(y \le z) \, dy$, then by Schwartz's inequality

$$\begin{split} &\int_0^1 \int_0^1 k^2(x,z) \, d\mu(x) \, d\mu(z) \\ &\leq \int_0^1 \int_0^1 \left[\int_0^1 \mathbf{1}^2(y \le x) \, dy \right] \left[\int_0^1 \mathbf{1}^2(y \le z) \, dy \right] \, d\mu(x) \, d\mu(z) \\ &= 1. \end{split}$$

Hence, L is Hilbert-Schmidt [21, Theorem VI.23] and therefore compact [21, Theorem VI.22].

In order to determine the singular values we consider the eigenvalue problem $Lf = \lambda f$ which is equivalent to

(3.1)
$$\lambda f(x) = \int_0^x \int_y^1 f(z) \, d\mu(z) \, dy.$$

Differentiating twice, we observe that for $\lambda \neq 0$ this is equivalent to the eigenvalue problem

$$\lambda f''(x) + \frac{f(x)}{x} = 0$$

for x in (0, 1). This is seen to be a transformed version of the Bessel equation, which has the following general solution, see e.g. [2, 9.1.51]

$$f(x) = \sqrt{x} \left[A J_1\left(\frac{2}{\sqrt{\lambda}}\sqrt{x}\right) + B Y_1\left(\frac{2}{\sqrt{\lambda}}\sqrt{x}\right) \right],$$

where A and B are arbitrary constants and J_1 and Y_1 are first and second kind Bessel functions of order 1, respectively. From (3.1) we obtain that the general solution must satisfy the following boundary conditions

(3.2)
$$\lim_{x \to 0} f(x) = 0$$

 and

(3.3)
$$\lim_{x \to 1} f'(x) = 0$$

The asymptotic properties $J_1(x) \sim x/2$ and $Y_1(x) \sim -1/(2\pi x)$ ([2, 9.1.7 and 9.1.9]) together with boundary condition (3.2) implies that B = 0. Notice, boundary condition (3.3) implies that

(3.4)
$$J_1\left(\frac{2}{\sqrt{\lambda}}\right) + \frac{2}{\sqrt{\lambda}}J_1'\left(\frac{2}{\sqrt{\lambda}}\right) = 0$$

which by [2, Equation 9.1.30] is seen to be equivalent to

(3.5)
$$\frac{2}{\sqrt{\lambda}}J_0\left(\frac{2}{\sqrt{\lambda}}\right) = 0.$$

As J_0 is known to have a countable number of real zeroes [2, Section 9.5], we let $\lambda_1 \geq \lambda_2 \geq \ldots > 0$ be the ordered set of solutions to (3.4) or (3.5).

We determine A by noting that we want an orthonormal basis and that (3.4) together with [2, Equation 11.4.5] implies

$$\|f_{j}\|_{\mu}^{2} = \int_{0}^{1} A^{2} x J_{1}^{2} \left(\frac{2}{\sqrt{\lambda_{j}}} \sqrt{x}\right) \frac{dx}{x} = 2A^{2} \int_{0}^{1} x J_{1}^{2} \left(\frac{2}{\sqrt{\lambda_{j}}} x\right) dx$$
$$= A^{2} J_{1}^{2} \left(\frac{2}{\sqrt{\lambda_{j}}}\right) = 1.$$

Hence

$$f_j(x) = \left| J_1\left(\frac{2}{\sqrt{\lambda_j}}\right) \right|^{-1} \sqrt{x} J_1\left(\frac{2}{\sqrt{\lambda_j}}\sqrt{x}\right),$$

where λ_j satisfy (3.5). From (2.2) we get the following expression for g_j with singular values given by $\mu_j = \sqrt{\lambda_j}$

$$g_j(y) = \frac{1}{\mu_j} K f_j(y)$$

= $\frac{1}{\sqrt{\lambda_j}} \int_y^1 f_j(x) d\mu(x)$
= $\left| J_1\left(\frac{2}{\sqrt{\lambda_j}}\right) \right|^{-1} \int_{2\sqrt{y}/\sqrt{\lambda_j}}^{2/\sqrt{\lambda_j}} J_1(x) dx$
= $\left| J_1\left(\frac{2}{\sqrt{\lambda_j}}\right) \right|^{-1} \left[\int_0^{2/\sqrt{\lambda_j}} J_1(x) dx - \int_0^{2\sqrt{y}/\sqrt{\lambda_j}} J_1(x) dx \right]$

which by [2, Equation 11.1.6] and (3.5) yields

$$g_j(y) = \left| J_1\left(\frac{2}{\sqrt{\lambda_j}}\right) \right|^{-1} J_0\left(\frac{2}{\sqrt{\lambda_j}}\sqrt{y}\right).$$

If we assume that only the coefficients (f, f_j) early in the expansion are important we get as estimator of g the following windowed SVD reconstruction formula

$$\hat{f}_n^{\alpha}(x) = \sum_{j=1}^{\infty} \phi_{\alpha}(\mu_j) n^{-1} \sum_{i=1}^n g_j(Y_i) f_j(x)$$
$$= \sum_{j:\lambda_j^{-\frac{1}{2}} \ge \alpha} \lambda_j^{-\frac{1}{2}} J_1\left(\frac{2}{\sqrt{\lambda_j}}\right)^{-2} \sqrt{x} J_1\left(\frac{2}{\sqrt{\lambda_j}}\sqrt{x}\right) n^{-1} \sum_{i=1}^n J_0\left(\frac{2}{\sqrt{\lambda_j}}\sqrt{Y_i}\right).$$

3.2. Estimation based on sine functions. Another possibility for reconstruction would be to substitute h(x) = f(x)/x and then consider the inverse problem of reconstructing h where we have a random sample Y_1, \ldots, Y_n from a density $g(y) = \int_y^1 h(x) dx$ or g = Kh, say. Here $K : \mathcal{H} \to \mathcal{G}$ is an operator between the spaces of functions on (0, 1) which are square integrable with respect to Lebesgue measure, ν .

It is easy to verify along the above sketched lines that K is compact and the SVD becomes (see e.g. [16, Example A.52])

$$\mu_j = \frac{2}{(2j-1)\pi}$$
$$h_j(x) = \sqrt{2} \sin\left(\frac{2j-1}{2}\pi x\right)$$
$$g_j(y) = \sqrt{2} \cos\left(\frac{2j-1}{2}\pi y\right)$$

for j = 1, 2, ... A reconstruction formula for h is then given by

$$\hat{h}_{n}^{\alpha}(x) = \sum_{j=1}^{\infty} \phi_{\alpha}(\mu_{j}) \left[n^{-1} \sum_{i=1}^{n} g_{j}(Y_{i}) h_{j}(x) \right],$$

and for f(x) = xh(x) by

$$\tilde{f}_n^{\alpha}(x) = x\hat{h}_n^{\alpha}(x).$$

It seems reasonable to let α depend on the sample size n, as the more data we get the more coefficients we should be able to estimate. Matters related to this question are discussed in the following section.

Remark 3.1. The procedure sketched above can also be given a interpretation as estimation of density derivatives. Equation (1.1) can namely be written as g'(y) = f(y)/y. Now, expand g in the cosine basis $g_j(y) = \sqrt{2}\cos((2j-1)\pi y/2), j = 1, 2, ...$ and estimate f in the following way

$$\tilde{f}_{n}^{\alpha}(x) = x \frac{d\hat{g}_{n}^{\alpha}(x)}{dx} = x \sum_{j=1}^{\infty} \phi_{\alpha}(\mu_{j}) \left[n^{-1} \sum_{i=1}^{n} g_{j}(Y_{i}) f_{j}(x) \right].$$

Then we arrive at the same estimator.

4. Asymptotic properties of the estimators. In the present section we shall derive an upper bound on the mean integrated square error (MISE) which provides a pseudo-consistency result for the estimators suggested in Section 3.

A common way of measuring the quality of an estimator f based on observations Y_1, \ldots, Y_n is the mean integrated square error (MISE) given by

MISE
$$(\hat{f}, f) = \mathbf{E} \int (\hat{f} - f)^2 d\mu(x).$$

By standard calculations

$$\mathrm{MISE}(\hat{f}, f) = \int (\mathbf{E}\hat{f} - f)^2 d\mu(x) + \int \mathrm{Var}(\hat{f}) d\mu(x),$$

which yields the MISE as a sum of the integrated square bias and the integrated variance. An assessment of the quality of \hat{f} should not depend on a particular unknown f, but is more naturally obtained by restricting f to belong to a certain class of functions, $\tilde{\mathcal{F}}$, say. The maximum risk, defined by

$$R(\hat{f}) = \sup_{f \in \widetilde{\mathcal{F}}} \text{MISE}(\hat{f}, f)$$

gives an indication of how well an estimator performs.

If we for a general one-dimensional problem, g = Kf, assume the singular values of the SVD satisfies $\mu_j \sim j^{-\beta}$ as $j \to \infty$, for a $\beta > 1$, and $\mathcal{F}_{\alpha,C}$ is a subspace of \mathcal{F} satisfying

$$\mathcal{F}_{\alpha,C} = \left\{ f = \sum_{j=1}^{\infty} b_j f_j \in \mathcal{F} : \sum_{j=1}^{\infty} |b_j|^2 (1+j)^{2\alpha} \le C \right\},$$

where $\alpha > 1$ and $0 < C < \infty$. Then it is possible by standard calculations to get estimates for the windowed SVD-estimator introduced in Section 3.1. First an expression for the integrated square bias

$$\int_0^1 (\mathbf{E} \ \hat{f}_n^{\alpha} - f)^2 \, d\mu(x) = \sum_{j:\mu_j < \alpha} \mu_j^{-2} |(g, g_j)|^2$$

and secondly an expression for the integrated variance

$$\int_{0}^{1} \operatorname{Var}(\hat{f}_{n}^{\alpha}) \, d\mu(x) = n^{-1} \sum_{j:\mu_{j} \ge \alpha} \mu_{j}^{-2} \operatorname{Var}(g_{j}(Y_{1})).$$

Hence a first expression for the MISE is given by

(4.1)
$$\operatorname{MISE}(\hat{f}_n^{\alpha}, f) = \sum_{j:\mu_j < \alpha} \mu_j^{-2} |(g, g_j)|^2 + n^{-1} \sum_{j:\mu_j \ge \alpha} \mu_j^{-2} \operatorname{Var}(g_j(Y_1)).$$

Now, let $L^1_{\mu,M}(0,1)$ denote the space of functions on (0,1) which have integrals with respect to $d\mu$ which are bounded from above by a strictly positive number M. If we additionally assume $f \in L^1_{\mu,M}(0,1)$ it is possible to uniformly bound the variance of $g_j(Y_1)$ by M.

Finally let $\alpha(n) = n^{1/(2\alpha+2\beta+1)}$ and assume $f \in \mathcal{F}_{\alpha,C} \cap L^1_{\mu,M}(0,1)$, then there exists a constant c > 0 depending on the SVD only such that

$$\mathrm{MISE}(\hat{f}_n^{\alpha(n)}, f) \le cMn^{-2\alpha/(2\alpha+2\beta+1)}$$

Therefore, the MISE for the chosen $\alpha(n)$ tends to zero as the number of observations tends to infinity. Moreover we have an upper bound on the rate of convergence. Hence we have the following result.

PROPOSITION 4.1. Let the setup be as in Section 3.1. Let $\alpha(n) = n^{1/(2\alpha+2\beta+1)}$ and assume that $f \in \mathcal{F}_{\alpha,C} \cap L^1_M(0,1)$. Then

$$\mathbf{E} \| \hat{f}_n^{\alpha(n)} - f \|_{\mu}^2 \to 0, \qquad n \to \infty.$$

Now, turning to the SVD-estimator introduced in Section 3.2 and proceeding as above (but without the L^1 -assumption, as the basis functions are now bounded by 1), we get

$$\mathbf{E} \|\tilde{f}_{n}^{\alpha(n)} - f\|_{\mu}^{2} \le \mathbf{E} \|\hat{h}_{n}^{\alpha(n)} - h\|_{\nu}^{2} = O(n^{-2\alpha/(2\alpha+2\beta+1)}),$$

and hence

PROPOSITION 4.2. Let the setup be as in Section 3.2. Let $\alpha(n) = n^{1/(2\alpha+2\beta+1)}$ and assume that $h \in \mathcal{H}_{\alpha,C}$, where

$$\mathcal{H}_{\alpha,C} = \left\{ h = \sum_{j=1}^{\infty} b_j h_j \in \mathcal{H} : \sum_{j=1}^{\infty} |b_j|^2 (1+j)^{2\alpha} \le C \right\}$$

Then

$$\mathbf{E} \| \tilde{f}_n^{\alpha(n)} - f \|_{\mu}^2 \to 0, \qquad n \to \infty.$$

Remark 4.1. The assumption $f \in \mathcal{F}_{\alpha,\mathcal{C}}$ is normally interpreted as a smoothness condition on the function we attempt to reconstruct. This can intuitively be seen as α determines the decay of the high frequency terms. However, for $\mathcal{H}_{\alpha,\mathcal{C}}$ in the setup of Section 3.2 it can be nicely formalized. We have that h_j^2 are trigonometric polynomials on (0,1), $\mathcal{H}_{\alpha} = \{h = \sum_{j=1}^{\infty} b_j h_j \in \mathcal{H} : |b_j|^2 (1+j)^{2\alpha} < \infty\}$ corresponds to the Sobolev space of order α . Therefore $h \in \mathcal{H}_{\alpha}$ if and only if h is a periodic function that has square integrable α -derivatives on the interval. For details see [3].

5. Empirical choice of cutoff scheme. As seen in the previous section, minimization of the MISE as a function of α requires information about the density f, which is unknown in practical situations. We therefore turn to a data-driven method for choosing the smoothing parameters as suggested by [7]. In the present section we also use the assumption that the variance of $g_j(Y_1)$ is bounded, which is indeed the case if we assume $f \in L^1_\mu(0, 1)$.

It is straightforward from (4.1) to show that the $MISE(\hat{f}_n^{\alpha}, f)$ has the same minimum as the functional

$$M_{n}(\alpha) = \sum_{j=1}^{\infty} \mu_{j}^{-2} \left[\frac{1}{n} \phi_{\alpha}^{2}(\mu_{j}) \mu_{j}^{2} \mathbf{E} g_{j}^{2}(Y_{1}) + \left\{ \frac{n-1}{n} \phi_{\alpha}(\mu_{j}) \mu_{j} - 2 \right\} \phi_{\alpha}(\mu_{j}) \mu_{j} \mathbf{E} g_{j}(Y_{1}))^{2} \right].$$

This still implicitly depend on f and we choose to approximate $M_n(\alpha)$ by plugging in the following unbiased and consistent estimates of $\mathbf{E}g_i^2(Y_1)$ and $(\mathbf{E}g_j(Y_1))^2$

$$\mathbf{E}g_j^2(Y_1) \simeq \overline{g_j^2} = n^{-1} \sum_{i=1}^n g_j^2(Y_i)$$

and

$$(\mathbf{E}g_j(Y_1))^2 \simeq n^{-1} \sum_{i=1}^n \overline{g}_j^{(i)} g_j(Y_i),$$

where

$$\overline{g}_{j}^{(i)} = \frac{1}{n-1} \sum_{k=1, k \neq i}^{n} g_{j}(Y_{k}),$$

(see Remark 5.1 below). This gives the following unbiased and consistent estimate of $M_n(\alpha)$ (see Remark 5.1 below) using the spectral cut-off

$$\widehat{M}_n(\alpha) = \sum_{\mu_j \ge \alpha} \mu_j^{-2} \left[n^{-1} \overline{g_j^2} + \left(\frac{n-1}{n} - 2 \right) n^{-1} \sum_{i=1}^n \overline{g}_j^{(i)} g_j(Y_i) \right].$$

Remark 5.1. To show the stated consistencies one should use the law of large numbers and the equivalents for U-statistics, see e.g. [13, Chapter 7.27]. Drawing on the connection to U-statistics it is also possible to prove that $\operatorname{Var}(\widehat{M}_n(\alpha)) = O(n^{-1})$, see [12, Theorem 5.2].

Remark 5.2. The method described above was also termed least-squares cross-validation in [23], as it has an intimate relation with cross-validation. For a detailed description of the connection see [22].

6. Finite sample study. The finite sample performance of the windowed SVD estimator presented in Section 3.1 is explored using simulated data based on a bimodal density. For comparison we also present the reconstructions based on sine functions as described in Section 3.2.

Reconstruction based on Bessel functions. We contemplate the performance of the SVD estimator based on X_i 's generated from a symmetrical bimodal density of the form f = 0.5Beta(8,3) + 0.5Beta(3,8). Here Beta(a,b) is the density function of the beta distribution with parameters a and b. We simulated 10 random samples with sample size n = 1000. The essential results are given in Figure 6.1. The quality of the estimated density is measured by the risk function \widehat{M}_n , which is shown in Figure 6.1(a) for all 10 samples together with the averaged risk function. As expected the estimated risk function varies, however, as the risk functions not always attain a well defined minimum we let the first local minimum set the order of reconstruction. Figure 6.1(b) shows the reconstructed density functions for the first



FIG. 6.1. Reconstructing a mixture of beta distributions by Bessel functions. The sample size is n = 1000. (a) An overlap plot of the risk function for all samples (thin lines) compared with the averaged risk function (thick line); and (b) an overlap plot of reconstructions of the density function for the first 5 samples with the first local minimum taken as the order of reconstruction (thin lines) and the true density function (thick line).



FIG. 6.2. Reconstructing a mixture of beta distributions by Bessel functions. The sample size is n = 1000. (a) The estimated risk function for the first sample (thin line) together with the averaged risk function (thick line); and (b) reconstructions of the first sample for different order of reconstruction compared with the true density (thick line).



FIG. 6.3. Reconstructing a mixture of beta distributions by Bessel functions. The sample size is n = 1000. (a) The mean of the estimated density functions for the 10 samples (dashed line) compared with the true density (solid line); and (b) an overlap plot of reconstructions of the density function for the 10 samples with N = 6 taken as the over all order of reconstruction (thin lines) and the true density (thick line).

5 samples. The rather large variability in the reconstructions is inherent to a large between sample variation and to a varying number of included eigenfunction.

The reconstructions shown in Figure 6.1 are rather disappointing. However, as suggested by Rudemo [22] it may be advisable to search through a set of local minima to find the optimal order of reconstruction. Investigating the risk function for the first sample we find three local minima, see Figure 6.2. It is apparent from Figure 6.2 that the reconstruction corresponding to N = 6 is the best, though this is not a global minimum. Note also, the fluctuation of the estimated density increases as we increase the reconstruction parameter.

Consulting the averaged risk function in Figure 6.2(a) we find a well defined local minimum at N = 6, which we therefore take as a new overall order of reconstruction for every sample. The performance of the reconstruction is shown in Figure 6.3. Figure 6.3(a) shows the mean of the estimated density functions for the 10 samples whereas Figure 6.3(b) shows an increased variability inherent to the increased reconstruction order. It is obvious from Figures 6.1 and 6.3, that bias decreases and variance increases as N increases.

Reconstruction based on sine functions. The performance of the SVD estimator is compared to reconstruction based on a basis of sine functions. The same 10 samples used above were exploited in the sine setting described in Section 3.2. The basic results are given in Figure 6.4. All estimated risk functions attain a well defined local minimum at N = 4, see Figure 6.4(a), which is taken at the reconstruction parameter. The corresponding reconstructions is illustrated in Figure 6.4(b) together with the true density function. It is obvious from the plot that the amplitude of the



FIG. 6.4. Reconstructing a beta distribution by sine functions. The sample size is n = 1000. (a) An overlap plot of the risk function; (b) the reconstructions (thin lines) for N = 4 compared with the true density (thick line); (c) the estimated risk function for the first sample (thin line) together with the averaged risk function (thick line); and (d) the mean density function (thin line) compared with the true density function (thick line).

reconstructions is too large. Consulting Figure 6.4(a) and (c) we find in general, that the estimated risk functions only attain one local minimum. The estimated mean density function is shown in Figure 6.4(d) together with the true density. Comparing the reconstructions presented above, we may conclude that the estimated densities in the sine setting seems to be more symmetrical than in the Bessel setting. This may be due to the symmetrical structure of the eigenfunctions, which is not the case in the Bessel setting. Nevertheless, the averaged reconstruction in Figure 6.3(d) seems to capture the modes of the true density better. 7. Final Remarks. Although, we have constructed a windowed SVD reconstruction formula for the (bounded) density of multiplicatively censored random variables several questions remain open.

One question is how to find an asymptotic lower bound on the MISE in order to make a complete asymptotic analysis. Comparing with the comprehensive literature on minimax estimation we notice that the upper bound obtained in the present paper corresponds to the rates obtained in similar problems (see e.g. [14, 18, 17]). These authors show additionally that the lower bounds equals the upper bound. It might therefore be possible along their lines to prove that the rate is a minimax bound.

Another question relates to the problem of densities with unbounded support. In this case one looses the compactness of the operator K, and has to face more general spectral theorems.

Acknowledgement. For helpful discussions we thank our colleagues in the "inverse problems study group" at Aalborg University.

REFERENCES

- F. ABRAMOVICH AND B. W. SILVERMAN, Wavelet decomposition approaches to statistical inverse problems, Biometrika, 85 (1998), pp. 115–129.
- [2] M. ABRAMOWITZ AND I. A. STEGUN, Handbook of Mathematical Functions, National Bureau of Stadards, Whasington, DC, 1964.
- [3] R. A. ADAMS, Sobolev Spaces, Academic Press, New York, 1978.
- [4] L. J. BRUNNER AND A. Y. LO, Bayes methods for a symmetric unimodal density and its mode, Ann. Statist., 17 (1989), pp. 1550–1566.
- [5] D. R. Cox, *Renewal Theory*, Meuthen, London, 1962.
- [6] —, Some sampling problems in technology, in New Developments in Survey Sampling, N. L. Johnson and H. Smith, eds., New York, 1969, Wiley, pp. 506–527.
- [7] A. K. DEY, B. A. MAIR, AND F. H. RUYMGAART, Cross-validation for parameter selection in inverse estimation, Scand. J. Statist., 23 (1996), pp. 609–620.
- [8] D. L. DONOHO, Nonlinear solution of linear inverse problems by wavelet-vaguelette decomposition, Appl. Comp. Harm. Anal., 2 (1995), pp. 101–126.
- S. EFROMOVICH, Density estimation for the case of supersmooth measurement error, J. Amer. Statist. Assoc., 92 (1997), pp. 526-535.
- [10] P. GROENEBOOM, Lectures on inverse problems, in Lectures on Probability Theory, Ecole d'Eté de Probabilités de Saint-Flour XXIV-1994, P. Bernard, ed., Springer Verlag, Berlin, 1996.
- [11] ——, Nonparametric estimation for inverse problems algorithms and asymptotics, Tech. Report 344, Department of Statistics, University of Whasington, 1998.
- [12] W. HOEFFDING, A class of statistics with asymptotically normal distribution, Ann. Math. Statist., 19 (1948), pp. 293–325.
- [13] J. HOFFMANN-JØRGENSEN, Probability With a View Toward Statistics, vol. I, Chapman and Hall, New York, 2 ed., 1994.
- [14] I. M. JOHNSTONE AND B. W. SILVERMAN, Speed of estimation in positron emission tomography and related inverse problems, Ann. Statist., 18 (1990), pp. 251–280.
- [15] ——, Discretization effects in statistical inverse problems, J. Complexity, 7 (1991), pp. 1–34.
- [16] A. KIRSCH, An Introduction to the Mathematical Theory of Inverse Problems, vol. 102 of Applied Mathematical Sciences, Springer Verlag, New York, 1996.
- [17] J.-Y. KOO AND H.-Y. CHUNG, Log-density estimation in linear inverse problems, Ann. Statist., 26 (1998), pp. 335-362.
- [18] B. MAIR AND F. H. RUYMGAART, Statistical inverse estimation in Hilbert Scales, Siam J. Appl. Math., 56 (1996), pp. 1424–1444.

- [19] F. O'SULLIVAN, A statistical perspective on ill-posed inverse problems (with discussion), Statist. Sci., 1 (1986), pp. 502–527.
- [20] G. P. PATIL AND C. R. RAO, The weighted distributions: A survey of their applications, in Applications of Statistics, P. R. Krishnaiah, ed., Amsterdam, 1977, North Holland, pp. 383– 405.
- [21] M. REED AND B. SIMON, Functional Analysis, Academic Press, London, 1980.
- [22] M. RUDEMO, Empirical choice of histogram and kernel density estimators, Scand. J. Statist., 9 (1982), pp. 65-78.
- [23] B. W. SILVERMAN, Density Estimation for Statistics and Data Analysis, vol. 26 of Monographs on Statistics and Applied Probability, Chapman and Hall, New York, 1986.
- [24] H. STØVRING AND W. VACH, Parametric estimation of prevalence, incidence and treatment duration based on the waiting time distribution, in Symposium i Anvendt Statistik, A. Holm and P. C. Linde, eds., København, 1999, AKF and SFI, pp. 106–113.
- [25] A. VAN ROIJ AND F. RUYMGAART, Asymptotic minimax rates for abstract linear estimators, J. Statist. Pl. Inf., 53 (1996), pp. 389–402.
- [26] Y. VARDI, Multiplicative censoring, renewal processes, deconvolution and decreasing density: Nonparametric estimation, Biometrika, 76 (1989), pp. 751-761.