Density Estimation Under Measurement Error

via Tikhonov Regularization with Smoothing Regularizer

David Kent

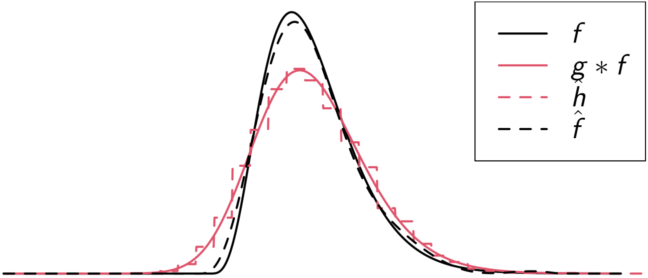
August 19, 2021

Outline

Density deconvolution

- X has unknown density f
- ε, independent of X, has known density g
- $Y = X + \varepsilon$ has density $h(y) = g * f(y) = \int f(t)g(y-t)\,dt$
  i.e. h is the convolution of f with g.
- The "deconvolution" problem is to estimate f from an i.i.d. sample $\{Y_i\}$ of random variables with density $h = g * f$.

In pictures



Applications?

- Natural: X is the measurement of interest, but instrument adds ε to observation.
- Bayesian model:

$$\theta_i \sim f(\theta)$$

$$Z_i|\theta_i \sim g_{\theta_i}(z) = g(z - \theta_i)$$

Know g, but want to choose f from the data.

Fourier Transform: Convolution $\Rightarrow$ Multiplication

Fourier transform of h: $\tilde{h}(\omega) = \int e^{-i\omega x} h(x)\, dx$

- If
$$h(y) = g * f(y) = \int f(x) g(y-x)\, dx,$$
then
$$\tilde{h}(\omega) = \tilde{g}(\omega)\tilde{f}(\omega) \text{ so that } \tilde{f}(\omega) = \tilde{h}(\omega)/\tilde{g}(\omega).$$
In this view, *deconvolution* is division by $\tilde{g}$.
- Also, $\|\tilde{h}\|^2 = 2\pi\|h\|^2$!! (Plancherel Theorem)

Deconvoluting Kernel Density Estimator

Introduced in Stefanski and Carroll [1990].

- Restrict h_n to be very smooth: Use a *kernel density estimator* with kernel K satisfying, for each λ,

$$\sup_{\omega} \left| \frac{\tilde{K}(\omega)}{\tilde{g}(\omega/\lambda)} \right| < \infty \text{ and } \int \left| \frac{\tilde{K}(\omega)}{\tilde{g}(\omega/\lambda)} \right| d\omega < \infty$$

- Form estimate $h_n^{\lambda}(y) = (n\lambda)^{-1} \sum_{j=1}^{n} K((Y_j - y)/\lambda)$, and do *exact* deconvolution.

$$f_n^{\lambda}(x) = (2\pi)^{-1} \int e^{i\omega x} \tilde{h}_n^{\lambda}(\omega)/\tilde{g}(\omega)\, d\omega.$$

Deconvoluting Kernel Density Estimator

- Alternatively: *deconvolve the kernel first:* $K^*_\lambda(x) = (2\pi)^{-1}\int e^{-ix\omega}\tilde{K}(\omega)/\tilde{g}(\omega/\lambda)\,d\omega$, then:

$$f_n^\lambda(x) = (n\lambda)^{-1}\sum_{j=1}^{n} K^*((Y_j - x)/\lambda).$$

- Pointwise consistent if $\lambda \to 0$ and $(n\lambda)^{-1}\int \tilde{K}(\omega)^2/\tilde{g}(\omega/\lambda)^2\,d\omega \to 0$.
- Attains optimal pointwise rates under assumption of Hölder-continuity of f [Fan, 1991], and optimal pointwise and L_2 rates under assumption that $\|\omega\tilde{f}(\omega)\| < M$ [Zhang, 1990].

Ill-posedness

A problem is "well-posed" if

1. For all admissible data, a solution exists,
2. For all admissible data, the solution is unique, and
3. The solution depends continuously on the data,

and "ill-posed" otherwise.

Problem, Data, and Solution

Consider $T : L_2(\mathbb{R}) \to L_2(\mathbb{R})$, $Tv = g * v$.

- **Data:** $h_n \in L_2(\mathbb{R})$ *(thinking: $h_n \approx h$)*
- **Problem:** $\arg\min_v \|Tv - h_n\|$ *(thinking: $\|Tf - h\| = 0$)*
- **Solution:** Just solve it? *(thinking: ...)*

This setting satisfies ✓ existence and ✓ uniqueness, but not ✗ continuity.

Discontinuous??

T is injective, so $T : L_2(\mathbb{R}) \to \mathcal{R}(T)$ is invertible,

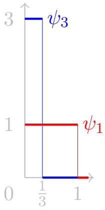
$$T^{-1}(g * v) = v$$

Continuity: $\|T^{-1}u\| \leq M\|u\|$

Let $\psi_n = n\mathbb{1}_{[0,\frac{1}{n}]}$.

Handy facts:

- $\|\psi_n - \psi_{n+\ell}\| = \sqrt{\ell}$
- $\|T\psi_n - T\psi_{n+\ell}\| \to 0$ as $n \to \infty$ (Folland [1999])



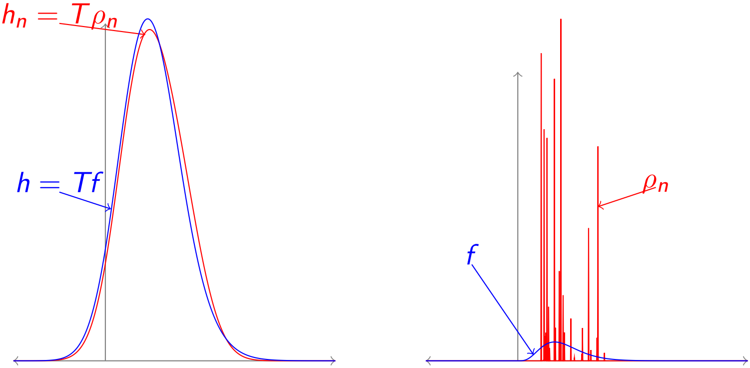
Uh-oh:

$$\|T^{-1}(T\psi_n - T\psi_{n+\ell})\| = \|\psi_n - \psi_{n+\ell}\| \overset{?}{\leq} M\|T\psi_n - T\psi_{n+\ell}\|$$

T^{-1} is not continuous!!

$\Rightarrow$ *any* extension of T^{-1} is not continuous!

The moral in a picture



Problem, Data, and Solution

Consider $T : L_2(\mathbb{R}) \to L_2(\mathbb{R})$, $Tv = g * v$.

- **Data:** $h_n \in L_2(\mathbb{R})$ (*thinking:* $h_n \approx h$)
- **Problem:** $\arg\min_v \|Tv - h_n\|$ (*thinking:* $\|Tf - h\| = 0$)
- **Solution:** Regularized solution (*thinking:* . . . 🧐)

Continuous approximations!

What do we do different?

Deconvoluting kernel estimators:

- Restrict admissible h_n
- Solve deconvolution problem exactly
- Approach exact problem as set of admissible h_n is enlarged

Addressed here: a Tikhonov regularization

- Do not restrict admissible h_n (decouple estimation of h from estimation of f)
- Solve deconvolution problem approximately
- Approach exact problem as approximation is relaxed

The Estimator: Tikhonov Regularization

$$f_n^\alpha = \arg\min_v \underbrace{\|g * v - h_n\|^2}_{\text{“data error”}} + \alpha \underbrace{\|v^{(2)}\|^2}_{\text{“smoothness penalty”}}$$

- As $\alpha \to 0$, exact (discontinuous) solution.
- For each α, $h_n \mapsto f_n^\alpha$ is continuous in h_n.

In Yang et al. [2021], we proved that a discrete approximation to this problem is weakly consistent

Outline

"Normal Equations"

If f_n^α minimizes $\|g * v - h_n\|^2 + \alpha \|v^{(2)}\|^2$ in v, then (Locker and Prenter [1980])

$$g * g * f_n^\alpha(x) + \alpha D^4 f_n^\alpha(x) = g * h_n(x)$$

Fourier transform reduces convolution and derivatives to *multiplication*:

$$\tilde{g}(\omega)^2 \tilde{f}_n^\alpha(\omega) + \alpha \omega^4 \tilde{f}_n^\alpha(\omega) = \tilde{g}(\omega) \tilde{h}_n(\omega)$$

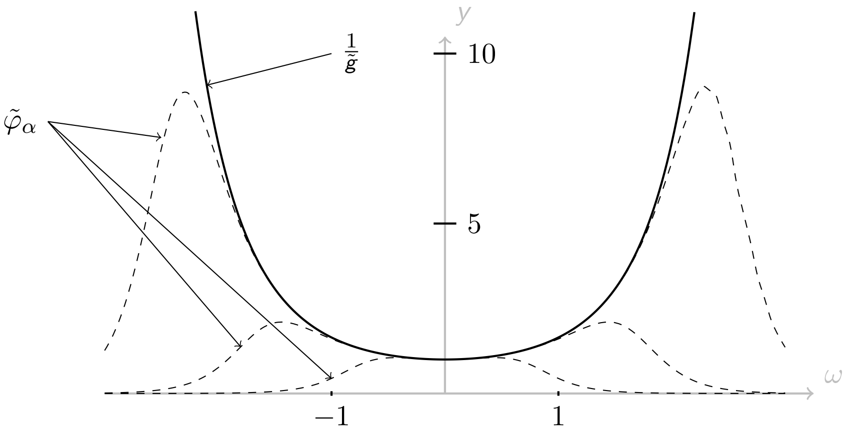
$$\tilde{f}_n^\alpha(\omega) = \frac{\tilde{g}(\omega)}{\tilde{g}(\omega)^2 + \alpha\omega^4} \tilde{h}_n(\omega) = \tilde{\varphi}_\alpha(\omega) \tilde{h}_n(\omega)$$

The Multiplier

$$\tilde{\varphi}_\alpha(\omega) = \frac{\tilde{g}(\omega)}{\tilde{g}(\omega)^2 + \alpha\omega^4}, \qquad \varphi_\alpha(x) = \frac{1}{2\pi}\int e^{i\omega x}\tilde{\varphi}_\alpha(\omega)\,d\omega.$$

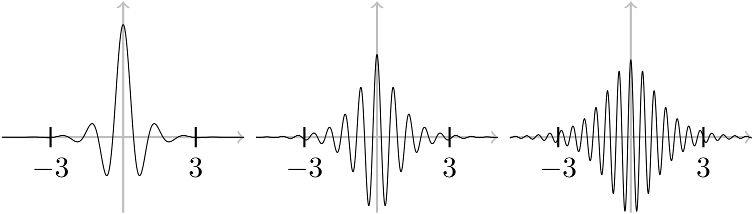
Theorem (Representing the Solution i.)

1. $\tilde{f}_n^\alpha(\omega) = \tilde{\varphi}_\alpha(\omega)\tilde{h}_n(\omega)$
2. $f_n^\alpha(x) = \varphi_\alpha * h_n(x)$

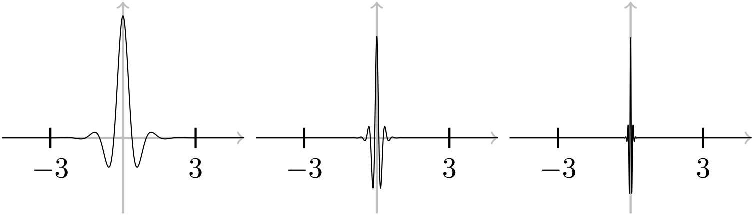


y
10
5
$\frac{1}{\tilde{g}}$
$\tilde{\varphi}_\alpha$
-1
1
ω

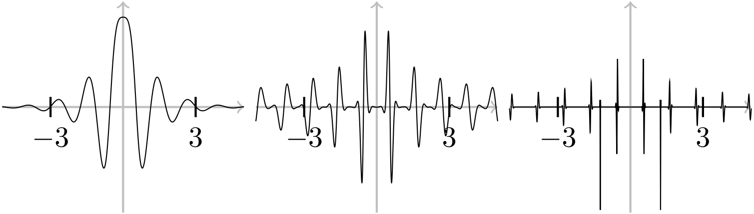
φ_α: Gaussian g



φ_α: Laplace g

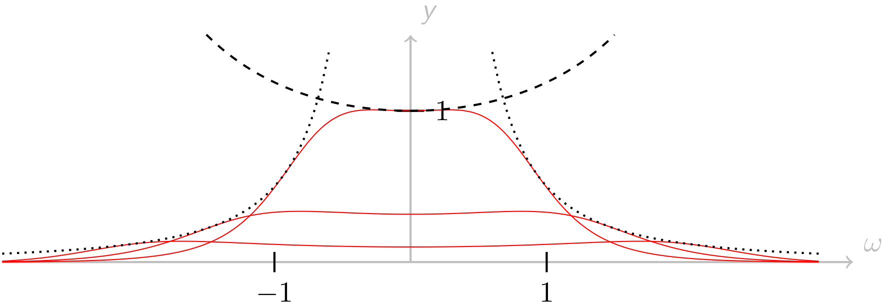


φ_α: Uniform g



Theorem

3. $\sup_\omega |\tilde{\varphi}_\alpha(\omega)| \leq C\alpha^{-\frac{1}{2}}$ *for some* $C > 0$ *depending only on* g*, as long as* $\alpha < 1$.
4. *If* $\varphi_\alpha \in L_1(\mathbb{R})$*, then* $\int \varphi_\alpha(x)\,dx = 1$*, so that* $\int f_n^\alpha(x)\,dx = 1$.



Proof of (3): Solid lines: $\sqrt{\alpha}|\tilde{\varphi}_\alpha(\omega)|$ with $\alpha = 1, 0.1, 0.01$. Dotted line: is $\frac{1}{2}\omega^{-2}$. Dashed line: $1/\tilde{g}(\omega)$.

Decomposition of Error

Define:

$$f^\alpha = \arg\min_{v} \|g * v - h\|^2 + \alpha \|v^{(2)}\|^2$$

It is useful to think:

$$f_n^\alpha - f = f_n^\alpha - f^\alpha + f^\alpha - f$$

- The first is the "random error"
- The second is the "approximation error"

Proposition (The Bias)

$$\mathbb{E}[f_n^\alpha(x) - f(x)] = \varphi_\alpha * \mathbb{E}[h_n - h](x) + (f^\alpha(x) - f(x))$$

In words: The bias of f_n^α for f is the Tikhonov-approximate deconvolved bias of h_n for h, plus the approximation error.

An Upper Bound on MISE

Let $\delta_n^2 = \mathbb{E}\|h_n - h\|^2$

Lemma

For α sufficiently small, we have the upper bound

$$\mathbb{E}\|f_n^\alpha - f\|^2 \leq C\delta_n^2/\alpha + 2\|f^\alpha - f\|^2.$$

Outline

A Consistency Theorem

Theorem (L_2 consistency)

Assume that α_n is chosen so that $\delta_n^2/\alpha_n \to 0$ and $\alpha_n \to 0$. Then

$$\lim_{n\to\infty} \mathbb{E}\|f_n^{\alpha_n} - f\|^2 = 0.$$

Rates for Smooth f

Theorem (Rates when f is very smooth)

Suppose $f^{(4)} \in L_2(\mathbb{R})$, *and* $f^{(4)} = g * g * \psi$ *for some* $\psi \in L_2(\mathbb{R})$.
If $\alpha_n = C_3 \delta^{\frac{2}{3}}$, *then*

$$\mathbb{E}\|f_n^{\alpha_n} - f\|^2 = O(\delta_n^{\frac{4}{3}}).$$

Corollary

If h_n *is KDE with optimal bandwidth, then*

$$\mathbb{E}\|f_n^{\alpha_n} - f\|^2 = O(n^{-8/15}).$$

If h_n *is a histogram with optimal bin widths, then*

$$\mathbb{E}\|f_n^{\alpha_n} - f\|^2 = O(n^{-4/9}).$$

Rates for normal g

Theorem (Rates for normal errors)

Assume g is a normal density, and assume $\|\omega^4 \tilde{f}(\omega)\| < M$. If $\alpha_n = \delta_n^2 W(\delta_n^{-\frac{2}{5}})^5$, then

$$\mathbb{E}\|f_n^\alpha - f\|^2 = O([\log \delta_n^{-1}]^{-4}) = O([\log n]^{-4}).$$

($W(\cdot)$ is the principal branch of the Lambert W function)

Compare (Zhang [1990])

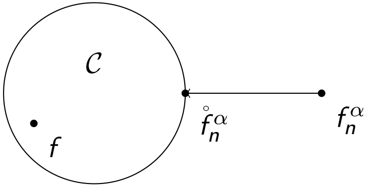
Assumes $\|\omega \tilde{f}(\omega)\| < M$ and finds lower bound

$$\mathbb{E}\|\hat{f}_n - f\|^2 = O([\log n]^{-1}).$$

Probability Density Constraints

We know: $\int f = 1$, $f \geq 0$: Produce estimate $\mathring{f}_n^\alpha$ which is a pdf as well?

Let $\mathcal{C} = \{v \in L_2(\mathbb{R}), v \text{ is a pdf}\}$.



Define the *metric projection* P_A onto a closed, convex set A by

$$P_A u = \arg\min_{v \in A} \|v - u\|.$$

P_A is *non-expansive* (Engl et al. [1996] Lemma 5.13):

$$\|P_A u - P_A v\| \leq \|u - v\|$$

Probability Density Constraints

Problem: $\mathcal{C}$ is convex, but *not closed*
Instead, we need to do this approximately:

$$\mathcal{C}_t = \{v \in \mathcal{C}, v(x) = 0 \text{ for all } x \notin [-t, t]\}$$

$\mathcal{C}_t$ is both closed and convex.

Probability Density Constraints: How?

Solve unconstrained problem:

$$f_n^\alpha = \arg\min_v \|g * v - h_n\|^2 + \|v^{(2)}\|^2$$

Then project:

$$\mathring{f}_n^\alpha = P_{\mathcal{C}_t} f_n^\alpha$$

Probability Density Constraints

Theorem

Suppose $\mathbb{E}[|X|] < \infty$ *and* $f(x) = o(1)$ *as* $|x| \to \infty$*. Then all of the previous theorems on consistency and rates hold for the solution projected to* $\mathcal{C}_t$*, as long as* $t \to \infty$ *fast enough.*

Proof.

$$\begin{aligned} \|\mathring{f}_n^\alpha - f\| &\leq \|P_{\mathcal{C}_t} f_n^\alpha - P_{\mathcal{C}_t} f\| + \|P_{\mathcal{C}_t} f - f\| \\ &\leq \|f_n^\alpha - f\| + O(t^{-1}) \end{aligned}$$

□

Outline

Approximation Space

(Caveat: I do not have theory justifying this yet)

Instead of minimizing the Tikhonov functional over

$$W^{2,2}(\mathbb{R}) = \{v \in L_2(\mathbb{R}) : v^{(1)}, v^{(2)} \in L_2(\mathbb{R})\},$$

we minimize over an "approximation space" of dimension q, comprising cubic splines with knots $\{t_i\}$:

$$\mathscr{S} := \mathscr{S}(\{t_i\}) \subset W^{2,2}(\mathbb{R}) :$$

$$f_n^\alpha = \arg\min_{v \in \mathscr{S}} \|g * v - h_n\|^2 + \alpha \|v^{(2)}\|^2.$$

Two steps

This can be computed by:

1. Solve unconstrained problem over $\mathscr{S}$: a matrix inversion.
2. Project onto $\mathcal{C}_a \cap \mathscr{S}$: a quadratic program(ish).

Fact: $\mathscr{S}$ has as a B-spline basis $\{B_i(x)\}_{i=1}^{q}$, so if $v \in \mathscr{S}$, then

$$v(x) = \sum_{i=1}^{q} \theta_i B_i(x).$$

(1): The Linear System

Letting $\mathbf{M}$ and $\mathbf{P}$ be certain inner-product matrices, and $\mathbf{b}_n$ the vector of inner products of h_n with the $g * B_i$, we can write

$$\|g * v - h_n\|^2 + \alpha\|v^{(2)}\|^2 = \boldsymbol{\theta}^T\mathbf{M}\boldsymbol{\theta} - 2\mathbf{b}_n^T\boldsymbol{\theta} + \|h_n\|_2^2 + \alpha\boldsymbol{\theta}^T\mathbf{P}\boldsymbol{\theta},$$

which is minimized by

$$\boldsymbol{\theta}_n^\alpha = (\mathbf{M} + \alpha\mathbf{P})^{-1}\mathbf{b}_n.$$

(2) The Quadratic Program

Now we need

$$\mathring{f}_n^{\alpha} = \arg\min_{v \in \mathscr{S} \cap \mathcal{C}_t} \|v - f_n^{\alpha}\|^2,$$

which has coefficients

$$\mathring{\boldsymbol{\theta}}_n^{\alpha} = \arg\min_{\boldsymbol{\theta}^T \mathbf{1} = 1, \mathbf{B}\boldsymbol{\theta} \geq 0} \boldsymbol{\theta}^T \mathbf{G} \boldsymbol{\theta} - 2\boldsymbol{\theta} \mathbf{G} \boldsymbol{\theta}_n^{\alpha}.$$

This can be solved with `quadprog` in `R`.

Outline

Choosing α in practice

The results above do not tell us how to choose α.
$\mathrm{MISE}(\alpha, f) = \mathbb{E}\|f_n^\alpha - f\|^2$ contains a linear and a quadratic term involving h_n, and involves the unknown f as well.
Not bad if h_n is a histogram and we know f
I propose an iterated parametric bootstrap:

1. Choose a provisional $\alpha^{(0)}$. Then iterate:
2. Compute $f_n^{\alpha^{(k)}}$.
3. Compute $\alpha^{(k+1)} = \arg\min_\alpha \mathrm{MISE}(\alpha, f_n^{\alpha^{(k)}})$

Choosing α in practice

In simulations, the process seems to reliably converge[1]
If it converges to $\tilde{\alpha}$, the following fact is true:

$$\tilde{\alpha} = \arg\min_{\alpha} \ \mathrm{MISE}(\alpha, f_n^{\tilde{\alpha}}),$$

[1] To prove: Banach fixed-point theorem?

Outline

Three Methods

Estimate f by:

- "Tikhonov" method detailed today with bootstrap selection of α_n,
- "QP deconvolution" (QP decon/QPD) method which inspired today's topic and its "SURE" method of choosing α_n, [Yang et al., 2021]
- "Deconvoluting Kernel Density" (DKE) method, with "plug-in" method of choosing bandwidth. [Delaigle and Gijbels, 2004b]

Execution Times, $n = 1,000$

Table: Execution Time (s)

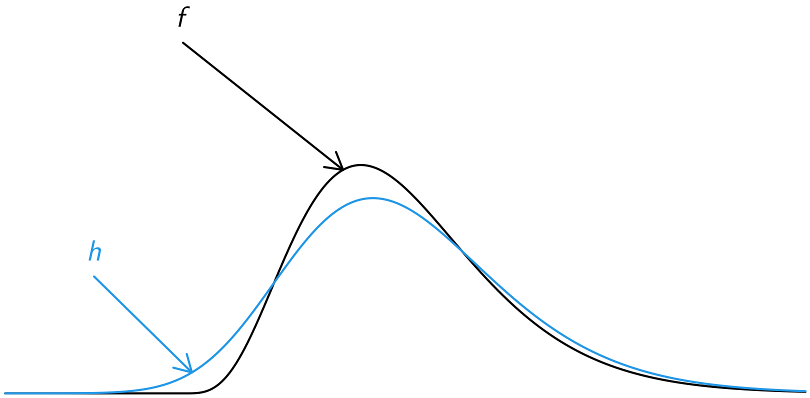
	Mean	Median	Max
DKE	1.337	1.326	1.896
Tikhonov	0.113	0.106	0.238
QPD	0.049	0.047	0.096

Experiment Setting

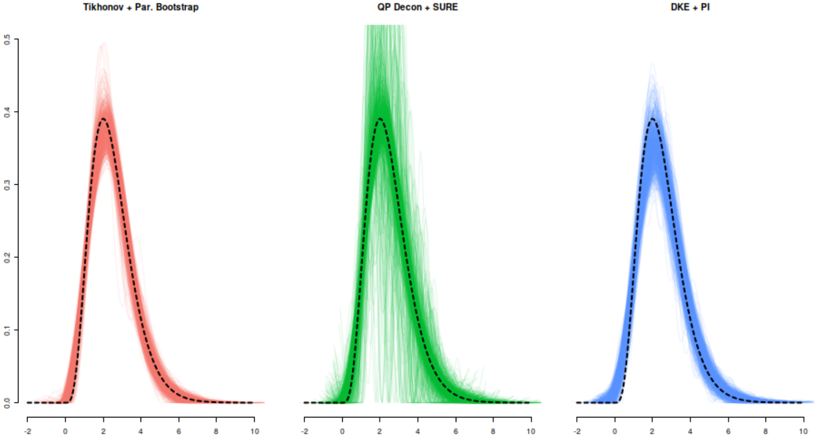
In all experiments, $\varepsilon \sim \mathrm{N}(0, \sigma^2)$, with $\mathrm{Var}(\varepsilon)/\mathrm{Var}(X) \approx 0.25$

1. $n = 1,000$; $X \sim \mathrm{Gamma}(\alpha = 5, \beta = 2)$
2. $n = 250$; $X \sim \mathrm{N}(0, 1)$
3. $n = 250$; $X \sim \chi^2(3)$
4. $n = 250$; $X \sim \frac{1}{2}\mathrm{N}(-3, 1) + \frac{1}{2}\mathrm{N}(2, 1)$
5. $n = 2,500$; $X \sim \frac{1}{2}\mathrm{N}(-3, 1) + \frac{1}{2}\mathrm{N}(2, 1)$

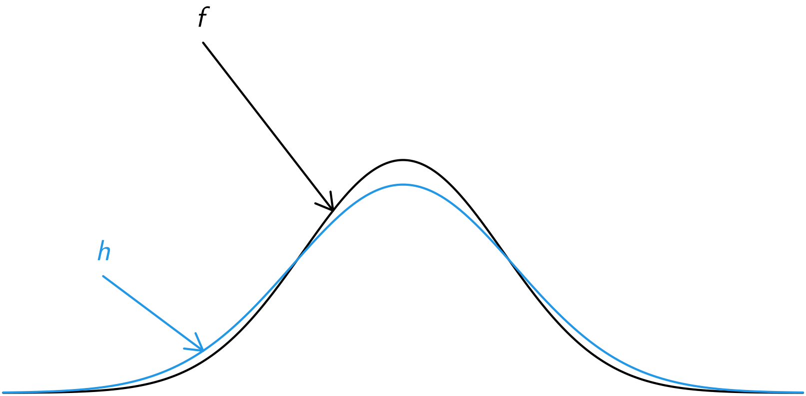
1: $X \sim \mathrm{Gamma}(\alpha = 5, \beta = 2)$, $\varepsilon \sim \mathrm{N}(0, 0.6^2)$, $n = 10^3$



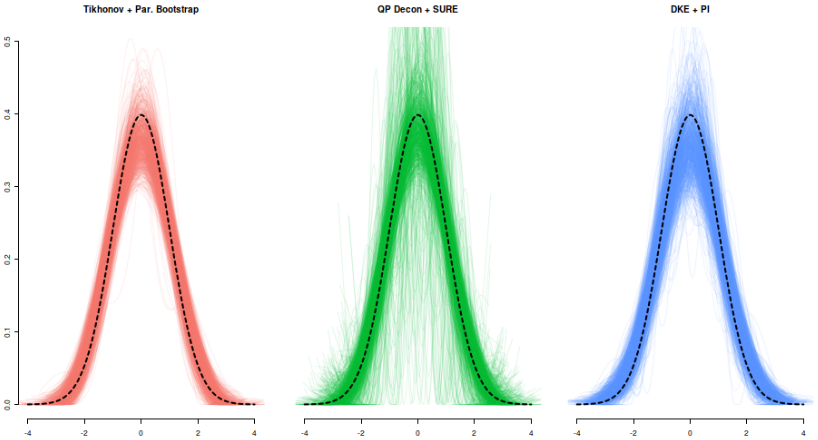
1: $X \sim \mathrm{Gamma}(\alpha = 5, \beta = 2)$, $\varepsilon \sim \mathrm{N}(0, 0.6^2)$, $n = 10^3$



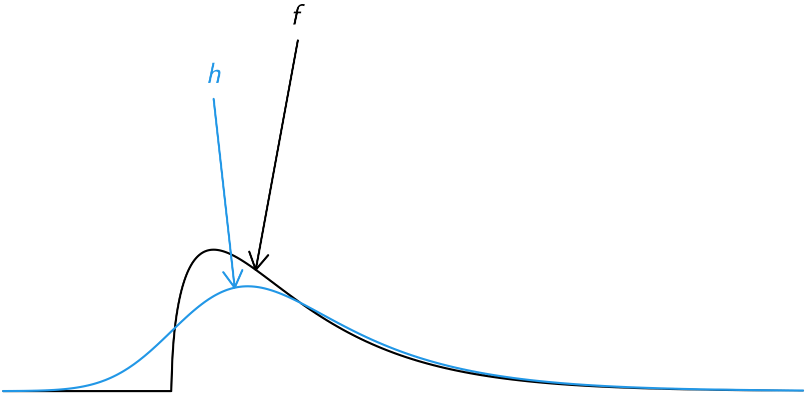
2: $X \sim \mathrm{N}(0,1)$, $\varepsilon \sim \mathrm{N}(0,0.25)$, $n = 250$



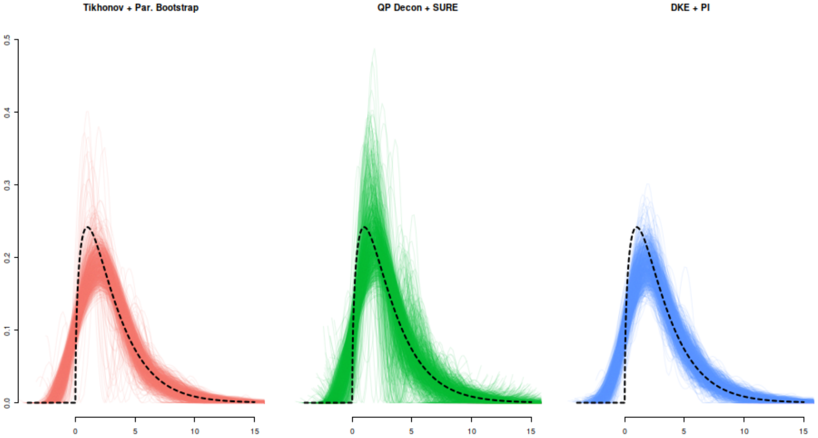
2: $X \sim \mathrm{N}(0,1)$, $\varepsilon \sim \mathrm{N}(0, 0.25)$, $n = 250$



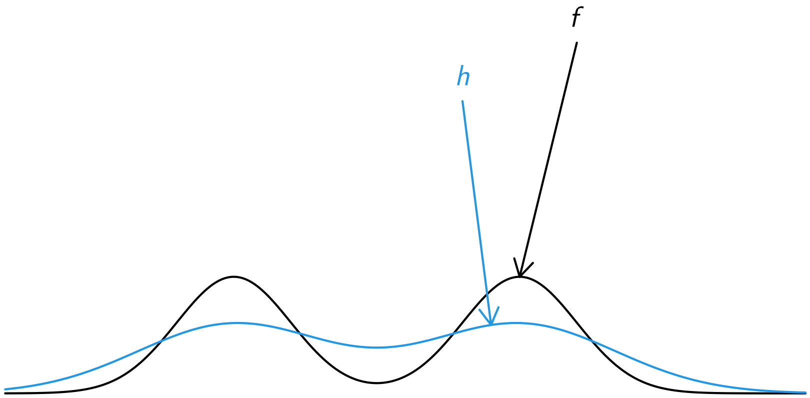
3: $X \sim \chi^2(3)$, $\varepsilon \sim \mathrm{N}(0, \sqrt{6})$, $n = 250$



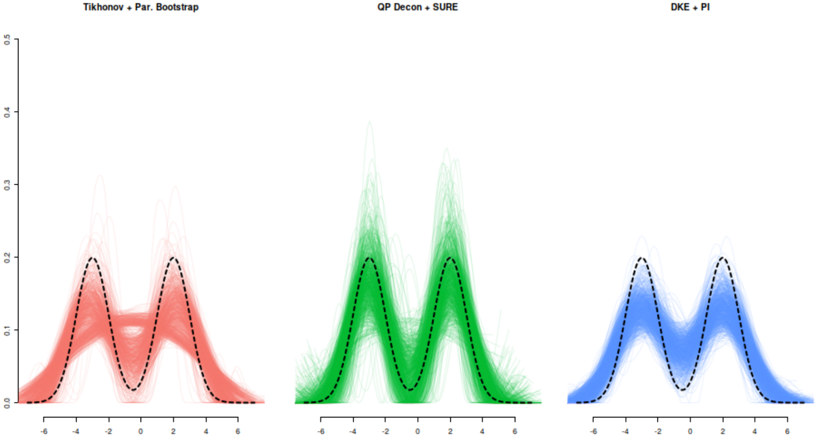
3: $X \sim \chi^2(3)$, $\varepsilon \sim \mathrm{N}(0, \sqrt{6})$, $n = 250$



4: $X \sim \frac{1}{2}\,\mathrm{N}(-3,1) + \frac{1}{2}\,\mathrm{N}(2,1)$, $\varepsilon \sim \mathrm{N}(0, \frac{1}{2}\sqrt{7.25})$, $n = 250$



4: $X \sim \frac{1}{2}\,\mathrm{N}(-3,1) + \frac{1}{2}\,\mathrm{N}(2,1)$, $\varepsilon \sim \mathrm{N}(0, \frac{1}{2}\sqrt{7.25})$, $n = 250$



5: $X \sim \frac{1}{2}\,\mathrm{N}(-3,1) + \frac{1}{2}\,\mathrm{N}(2,1)$, $\varepsilon \sim \mathrm{N}(0, \frac{1}{2}\sqrt{7.25})$, $n = 2,500$

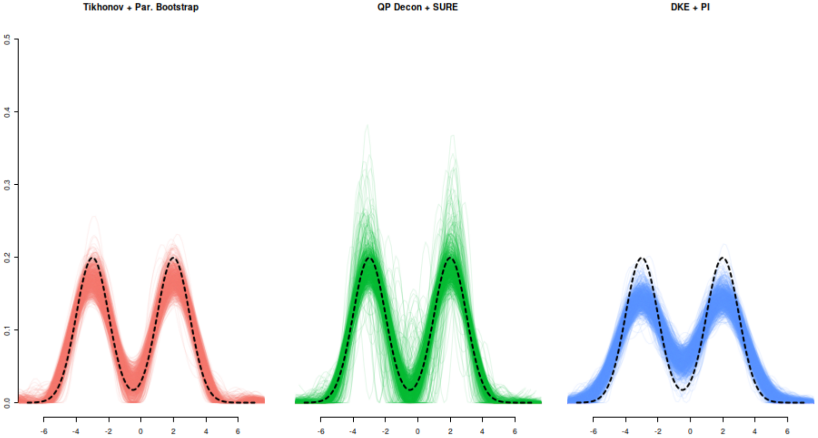
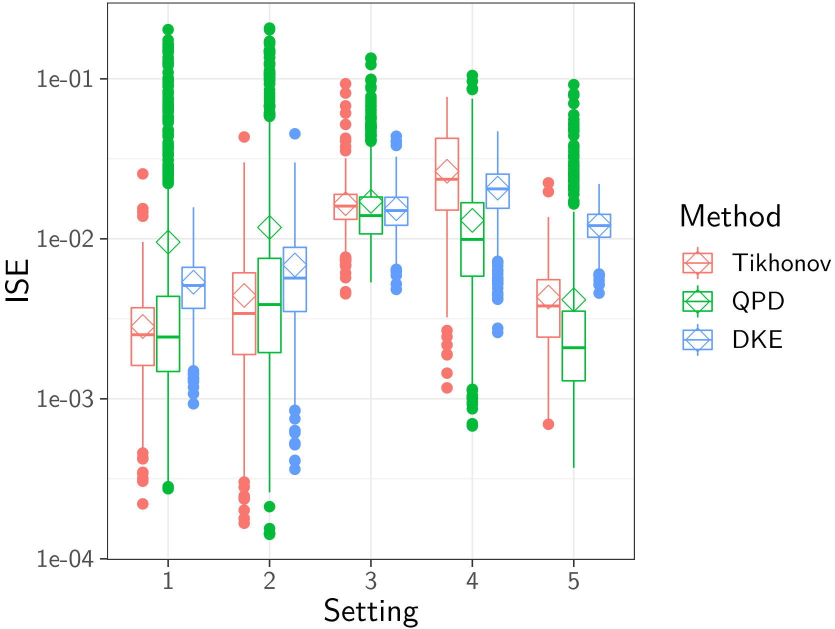


Table: Mean L_2 error

Method	Setting				
					
	$n = 1\,000$	$n = 250$	$n = 250$	$n = 250$	$n = 2\,500$
Tikhonov	0.0028	0.0044	0.0166	0.0265	0.0043
QPD	0.0095	0.0118	0.0172	0.0130	0.0042
DKE	0.0054	0.0069	0.0154	0.0208	0.0123


ISE
1e-01
1e-02
1e-03
1e-04
1
2
3
4
5
Setting
Method
Tikhonov
QPD
DKE

Outline

Paper #1

For a first paper:

- ✓ $L_2(\mathbb{R})$ Consistency,
- ✓ "Good" method for PDF-constrained result,
- ✓ Rates under strong assumption on f,
- Relax assumption on f to match [Zhang, 1990] bounds
- Rates under weak assumptions on f for several choices of g: ✓ Normal, ✗ Laplace, ✗ Cauchy, ✗ Uniform.
- Offer "easy" choice of α_n: with estimates of δ_n^2 and $\|\omega \tilde{f}(\omega)\|$, plug in constants and minimize upper bound on $\mathbb{E}\|f_n^\alpha - f\|^2$.

Time estimate: 2 weeks to finish the red results; one month to write the paper?

Paper #2

For a second paper:

- ✓ Write R package to compute these estimates,
- ✗ Include method for fast computation of DKE estimates?
- ✗ Theory justifying all the numerical approximations involved,
- ✗ Theory for bootstrapping choice of α_n and for inference.
  - Note: There is literature about bootstrap bandwidth selection [Delaigle and Gijbels, 2004a] for DKE, and iterated bootstrap has been used for bandwidth selection [Faraway and Jhun, 1990].
- ✗ Compute finite-n MISE for some choice of h_n, compare Wand [1998]

Beyond

Other things on my mind:

- Can the solution be computed exactly out of some finite-dimensional space using RKHS magic? Adapt Sections 2.1/2.2 of Berlinet and Thomas-Agnan [2004]? Cubic splines are natural for the smoothing spline problem, but have no special relationship to this problem
- Similar regularization methods (e.g. smoothing splines) have a Bayesian interpretation. Can anything similar be said here? If not, is there a "nearby" problem with such an interpretation?
- How do functionals of f_n^α behave?

Beyond

- How does the theory change with a weighted smoothness penalty $\|\alpha(x)v^{(2)}(x)\|^2$ replacing $\alpha\|v^{(2)}(x)\|^2$ in the objective? (The regularizer is no longer self-adjoint)
- How does the theory change with a more general convolution operator, e.g. with scale depending on X?
- What if we have g_n estimated which converges somehow to g with n?
- What if the scale $\sigma \to 0$ with n? Maybe inform rate of resource expense on lowering σ.
  - Sampling effort (in the lab or in the computer)
  - Design of higher-precision instruments

Bibliography I

Alain Berlinet and Christine Thomas-Agnan. *Reproducing Kernel Hilbert Spaces in Probability and Statistics*. Springer US, Boston, MA, 2004. ISBN 978-1-4613-4792-7 978-1-4419-9096-9. doi: 10.1007/978-1-4419-9096-9. URL http://link.springer.com/10.1007/978-1-4419-9096-9.

A. Delaigle and I. Gijbels. Bootstrap bandwidth selection in kernel density estimation from a contaminated sample. *Annals of the Institute of Statistical Mathematics*, 56(1):19–47, March 2004a. ISSN 0020-3157, 1572-9052. doi: 10.1007/BF02530523. URL http://link.springer.com/10.1007/BF02530523.

A. Delaigle and I. Gijbels. Practical bandwidth selection in deconvolution kernel density estimation. *Computational Statistics & Data Analysis*, 45(2):249–267, March 2004b. ISSN 01679473. doi: 10.1016/S0167-9473(02)00329-8. URL https://linkinghub.elsevier.com/retrieve/pii/S0167947302003298.

Heinz W Engl, Martin Hanke, and Andreas Neubauer. *Regularization of inverse problems*. Kluwer Academic Publishers, Dordrecht, 1996. ISBN 978-0-7923-4157-4. http://newcatalog.library.cornell.edu/catalog/2908683.

Jianqing Fan. On the Optimal Rates of Convergence for Nonparametric Deconvolution Problems. *The Annals of Statistics*, 19(3):1257–1272, September 1991. ISSN 0090-5364, 2168-8966. doi: 10.1214/aos/1176348248. URL http://projecteuclid.org/euclid.aos/1176348248.

Julian J. Faraway and Myoungshic Jhun. Bootstrap Choice of Bandwidth for Density Estimation. *Journal of the American Statistical Association*, 85(412):1119–1122, December 1990. ISSN 0162-1459. doi: 10.1080/01621459.1990.10474983. URL http://amstat.tandfonline.com/doi/abs/10.1080/01621459.1990.10474983.

G. B Folland. *Real analysis: modern techniques and their applications*. Wiley, New York, 2nd ed. edition, 1999. ISBN 978-0-471-31716-6. http://newcatalog.library.cornell.edu/catalog/3444791.

John Locker and P.M Prenter. Regularization with differential operators. I. General theory. *Journal of Mathematical Analysis and Applications*, 74(2):504–529, April 1980. ISSN 0022247X. doi: 10.1016/0022-247X(80)90145-6. URL https://linkinghub.elsevier.com/retrieve/pii/0022247X80901456.

Bibliography II

Leonard Stefanski and Raymond J. Carroll. Deconvoluting Kernel Density Estimators. *Statistics*, 21(2):169–184, January 1990. ISSN 0233-1888. doi: 10.1080/02331889008802238.

M. P. Wand. Finite sample performance of deconvolving density estimators. *Statistics & Probability Letters*, 37(2): 131–139, 1998. URL https://ideas.repec.org/a/eee/stapro/v37y1998i2p131-139.html. Publisher: Elsevier.

Ran Yang, David Kent, Daniel W. Apley, Jeremy Staum, and David Ruppert. Bias-corrected Estimation of the Density of a Conditional Expectation in Nested Simulation Problems. *ACM Transactions on Modeling and Computer Simulation*, 31(4):22:1–22:36, July 2021. ISSN 1049-3301. doi: 10.1145/3462201. URL https://doi.org/10.1145/3462201.

Cun-Hui Zhang. Fourier Methods for Estimating Mixing Densities and Distributions. *The Annals of Statistics*, 18 (2):806–831, 1990. ISSN 0090-5364. URL https://www.jstor.org/stable/2242135.

Acknowledgements

Big Thanks:

- David Ruppert for introducing me to this problem and listening to me whine about how I couldn't figure it out
- Tom & Kengo for support
- Elise for reliably going "wow, yeah... the wiggly lines look extra good this time..."
- Helen, Bailey, and Raymond for being soft and laying around a lot