Efficient Nonparametric Smoothness Estimation

Shashank Singh, Simon S. Du, and Barnabás Póczos

Machine Learning Department, Carnegie Mellon University

Carnegie Mellon University

Introduction

- The difficulty of a statistical problem is often determined by the **complexity of the data source**.
- In nonparametric statistics, complexity is often measured by the **smoothness of a density or regression function**.

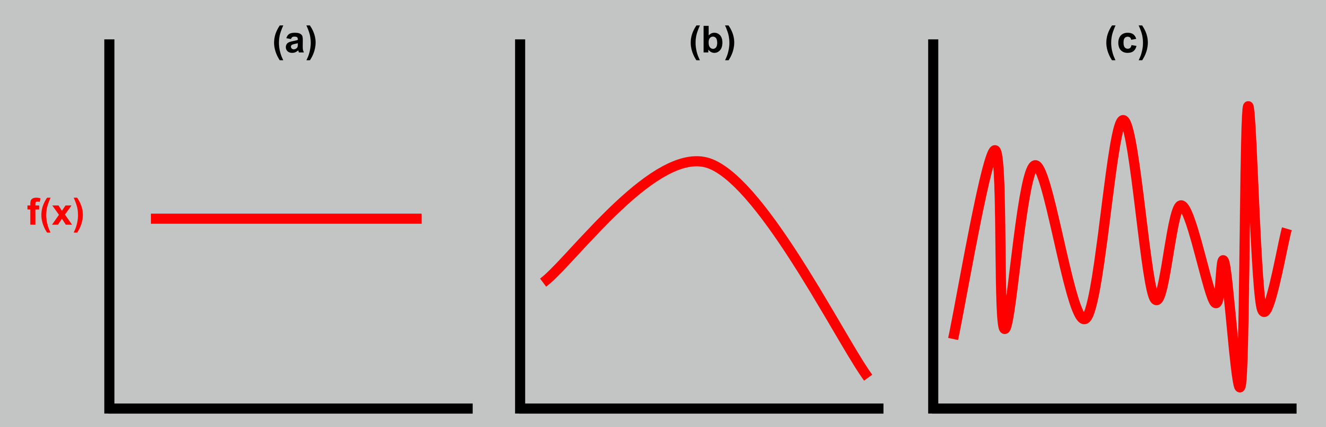


Figure: Functions of decreasing smoothness (increasing complexity) are harder to estimate. (a) constant ($\|f\|_{\mathcal{H}^1} = 0$) (b) parabola ($\|f\|_{\mathcal{H}^3} = 0$) (c) complex ($\|f\|_{\mathcal{H}^s}$ large, $\forall s > 0$)

- Smoothness of function f can be quantified in several ways, such as:
  - Sobolev norms, e.g. $\|f\|_{\mathcal{H}^1}^2 = \|f'\|_{\mathcal{L}_2}^2 = \int (f'(x))^2 \, dx$.
  - Hölder norms, e.g. $\|f\|_{\mathcal{C}^1} = \|f'\|_{\mathcal{L}^\infty} = \operatorname{ess\,sup}_x |f'(x)|$.
  - Various RKHS norms
- **Question: Can we estimate complexity from data?**
- **Our Answer: Yes (much easier than estimating f!)**

Sobolev Norms

- **Sobolev norms are $\mathcal{L}_2$-norms of derivatives.** For example, the s^{th}-order Sobolev norm of an s-times differentiable $f : [-\pi, \pi] \to \mathbb{R}$ is

$$\|f\|_{\mathcal{H}^s}^2 = \|f^{(s)}\|_2^2 = \int_{-\pi}^{\pi} \left(f^{(s)}(x)\right)^2 dx.$$

- **Notation:** Let $\varphi_z(x) = e^{izx}$ be the z^{th} Fourier basis element and

$$\tilde{f}(z) := \langle f, \varphi_z \rangle_{\mathcal{L}_2} = \int_{-\pi}^{\pi} f(x)\varphi_z(x) \, dx$$

be the z^{th} Fourier coefficient of F.

- By Parseval's identity and the Fourier transform of a derivative, the **Sobolev norm can be written in terms of Fourier coefficients**:

$$\int_{-\pi}^{\pi} \left(f^{(s)}(x)\right)^2 dx = \sum_{z \in \mathbb{Z}} \left|\widetilde{f^{(s)}}(z)\right|^2 = \sum_{z \in \mathbb{Z}} z^{2s} \left|\tilde{f}(z)\right|^2.$$

(Intuition: The smoother f is, the faster its Fourier coefficients decay.)

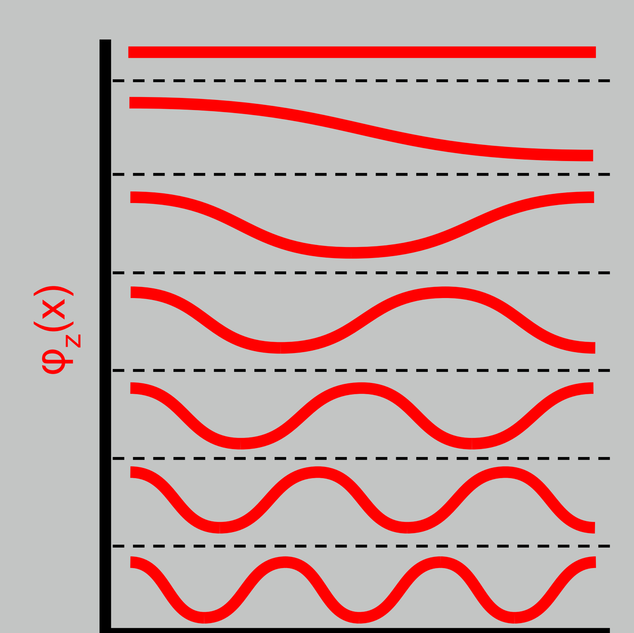


Figure: First 7 Fourier basis elements.

Estimating Sobolev Norms

- For a probability density function p on $[-\pi, \pi]$, **Fourier coefficients are expectations**, and hence easy to estimate from IID data $X_1, \ldots, X_n \sim p$:

$$\tilde{p}(z) = \int_{-\pi}^{\pi} p(x)\varphi_z(x) \, dx = \mathop{\mathbb{E}}_{X \sim p} [\varphi_z(X)] \approx \frac{1}{n} \sum_{j=1}^{n} \varphi_z(X_j) =: \hat{p}(z).$$

- To estimate $\|p\|_{\mathcal{H}^s}$, **truncate $z \le Z_n$ and plug in $\hat{p}(z)$ for $\tilde{p}(z)$**:

$$\hat{S}_n(s) = \sum_{|z| \le Z_n} z^{2s} |\hat{p}(z)|^2.$$

Overview of Results

- Our main results are the following:
  - $\hat{S}_n(s)$ converges at the **minimax-optimal rate**.
  - We derive the **asymptotic distribution** of $\hat{S}_n(s)$.
  - $\hat{S}_n(s)$ can be **computed in $O(n \log n)$ time** via FFT
- All our results **extend to more general cases**:
  - p has multidimensional support $\mathbb{R}^D$
  - p has unbounded support (with minor caveats)
  - Sobolev inner products $\langle p, q \rangle_{\mathcal{H}^s}$ and metrics $\|p - q\|_{\mathcal{H}^s}$
  - non-integer $s \in \mathbb{R}$
  - estimating $\|f\|_{\mathcal{H}^s}$ for a regression function f
    - can scale to higher dimensions assuming an additive model

Results: Convergence Rates

- Assume, for some $t > s$, $p \in \mathcal{H}^t$. Then, for some constant C, we prove:
  - **bias bound:** $\left| \mathbb{E}\left[\hat{S}_n(s)\right] - \|p\|_{H^s}^2 \right| \le C Z_n^{2(s-t)}$.
  - **variance bound:** $\mathbb{V}\left[\hat{S}_n(s)\right] \le C \frac{Z_n^{4s+D}}{n} + \frac{C}{n}$.
- These imply a **mean squared error bound**:

$$\mathbb{E}\left[\left(\hat{S}_n(s) - \|p\|_{H^s}^2\right)^2\right] \le C \left(Z_n^{4(s-t)} + \frac{Z_n^{4s+D}}{n} + n^{-1} \right),$$

for some constant $C > 0$ independent of n.

- Minimizing over Z_n gives $Z_n \asymp n^{\frac{2}{4t+D}}$, and hence

$$\mathbb{E}\left[\left(\hat{S}_n(s) - \|p\|_{H^s}^2\right)^2\right] \asymp n^{\max\left\{\frac{8(s-t)}{4t+D}, -1\right\}},$$

which is precisely the **minimax optimal rate**. [1]

- When $t \ge 2s + D/4$, setting $Z_n \asymp n^{\frac{1}{4s+D}}$ gives the parametric rate MSE $\asymp n^{-1}$ **adaptively (without knowing t)**.

Results: Asymptotic Distributions

- Assume $t > 2s + D/4$, and set $Z_n \asymp n^{\frac{1}{4s+D}}$. Then, we prove:
  - $\hat{S}_n(s)$ has an χ^2 **asymptotic distribution** with non-centrality parameter $\|p\|_{\mathcal{H}^s}$.
  - Specifically, define $W \in \mathbb{R}^{n \times Z_n}$ by $W_{j,z} := z^s \varphi_z(X_j)$. If $\hat{\mu} \in \mathbb{R}^{Z_n}$ and $\hat{\Sigma} \in \mathbb{R}^{Z_n \times Z_n}$ are the empirical mean and covariance of W, then

$$Q_{\chi^2(Z_n, \|p\|_{\mathcal{H}^s})}\left(n\hat{\mu}^T \hat{\Sigma}^{-1} \hat{\mu}\right) \xrightarrow{D} \text{Uniform}([0,1]),$$

where $Q_{\chi^2(d,\lambda)}$ denotes the quantile function of the χ^2 distribution with d degrees of freedom and non-centrality parameter λ.

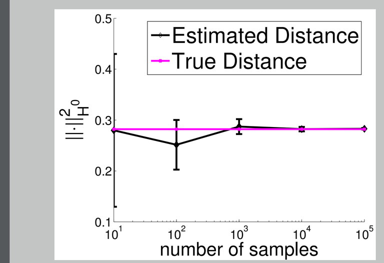
- Sobolev metrics also have χ^2 asymptotic distributions
- Sobolev inner products have normal asymptotic distributions

Consequences and Applications

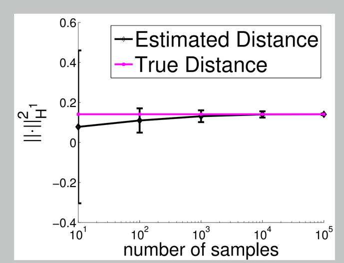
- Estimate key "theoretical" quantities in nonparametric error bounds.
- Test the null hypothesis that f satisfies a Sobolev condition.
- Provide a **fast nonparametric two-sample test**. Suggests how parameters should scale in recent work on two-sample testing. [2]

Experimental Results

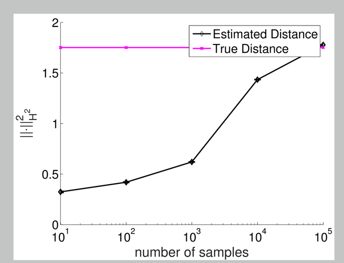
- Estimate Sobolev quantities for synthetic data with known ground truth
- Sobolev norm estimation ($\|p\|_{\mathcal{H}^s}$ for different p and s):



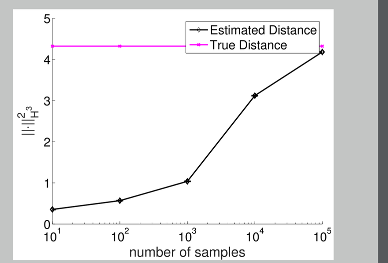
$\|\mathcal{N}(0,1)\|_{\mathcal{H}^0}$.



$\|\mathcal{N}(0,1)\|_{\mathcal{H}^1}$.

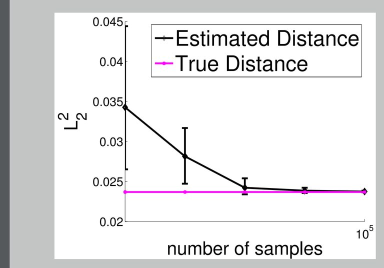


$\|1 + \cos(2\pi x)\|_{\mathcal{H}^2}$

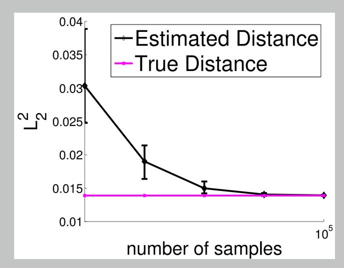


$\|1 + \cos(2\pi x)\|_{\mathcal{H}^3}$

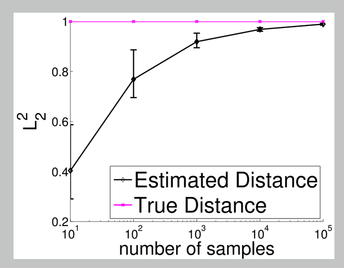
- Sobolev distance estimation ($\|p - q\|_{\mathcal{H}^0}$ for different p and q):



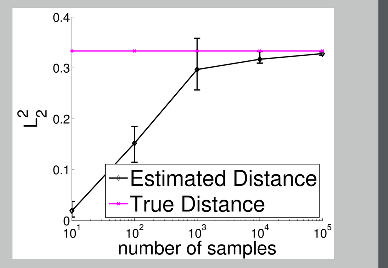
3D Gaussians with diff. means.



3D Gaussians with diff. variance.



Uniforms with diff. support.



Uniform and triangular distributions.

References

- Peter J Bickel and Yaacov Ritov. Estimating integrated squared density derivatives: sharp best order of convergence estimates. *Sankhyā: The Indian Journal of Statistics, Series A*, pages 381–393, 1988.
- Kacper P Chwialkowski, Aaditya Ramdas, Dino Sejdinovic, and Arthur Gretton. Fast two-sample testing with analytic representations of probability measures. In *NIPS*, pages 1972–1980, 2015.