
Efficient Nonparametric Smoothness Estimation: Supplementary Information

Shashank Singh
Carnegie Mellon University
sss1@andrew.cmu.edu

Simon S. Du
Carnegie Mellon University
ssdu@cs.cmu.edu

Barnabás Póczos
Carnegie Mellon University
bapocz@cs.cmu.edu

1 Proof of Variance Bound

Theorem 2. (Variance Bound) *If $p, q \in H^{s'}$ for some $s' > s$, then*

$$\mathbb{V} \left[\hat{S}_n \right] \leq 2C_1 \frac{Z_n^{4s+D}}{n^2} + \frac{C_2}{n}, \quad (1)$$

where C_1 and C_2 are the constants (in n)

$$C_1 := \frac{2^D \Gamma(4s+1)}{\Gamma(4s+D+1)} \|p\|_{L^2} \|q\|_{L^2}$$

and $C_2 := (\|p\|_{H^s} + \|q\|_{H^s}) \|p\|_{W^{2s,4}} \|q\|_{W^{2s,4}} + \|p\|_{H^s}^4 \|q\|_{H^s}^4$.

Proof: We will use the Efron-Stein inequality [Efron and Stein, 1981] to bound the variance of \hat{S}_n . To do this, suppose we were to draw n additional IID samples $X'_1, \dots, X'_n \sim p$, and define, for all $\ell, j \in \{1, \dots, n\}$,

$$X_j^{(\ell)} = \begin{cases} X'_j & \text{if } j = \ell \\ X_j & \text{else} \end{cases}.$$

Let

$$\hat{S}_n^{(\ell)} := \frac{1}{n^2} \sum_{|z| \leq Z_n} z^{2s} \sum_{j=1}^n \sum_{k=1}^n \psi_z(X_j^{(\ell)}) \overline{\psi_z(Y_k)}$$

denote our estimate when we replace X_ℓ by X'_ℓ . Noting the symmetry of \hat{S}_n in p and q , the Efron-Stein inequality tells us that

$$\mathbb{V} \left[\hat{S}_n \right] \leq \sum_{\ell=1}^n \mathbb{E} \left[\left| \hat{S}_n - \hat{S}_n^{(\ell)} \right|^2 \right], \quad (2)$$

where the expectation above (and elsewhere in this section) is taken over all $3n$ samples $X_1, \dots, X_{2n}, X'_1, \dots, X'_{2n}, Y_1, \dots, Y_n$. Expanding the difference in (2), note that any terms with $j \neq \ell$ cancel, so that¹

$$\begin{aligned} \hat{S}_n - \hat{S}_n^{(\ell)} &= \frac{1}{n^2} \sum_{|z| \leq Z_n} z^{2s} \sum_{j=1}^n \sum_{k=1}^n \psi_z(X_j) \overline{\psi_z(Y_k)} - \psi_z(X_j^{(\ell)}) \overline{\psi_z(Y_k)} \\ &= \frac{1}{n^2} \sum_{|z| \leq Z_n} z^{2s} (\psi_z(X_\ell) - \psi_z(X'_\ell)) \sum_{k=1}^n \overline{\psi_z(Y_k)}, \end{aligned}$$

and so

$$\left| \hat{S}_n - \hat{S}_n^{(\ell)} \right|^2$$

¹It is useful here to note that $\overline{\psi_z(x)} = \psi_{-z}(x)$ and that $\psi_y \psi_z = \psi_{y+z}$.

$$= \frac{1}{n^4} \sum_{|y|, |z| \leq Z_n} (yz)^{2s} (\psi_y(X_\ell) - \psi_y(X'_\ell)) (\psi_{-z}(X_\ell) - \psi_{-z}(X'_\ell)) \left(\sum_{k=1}^n \psi_{-y}(Y_k) \right) \left(\sum_{k=1}^n \psi_z(Y_k) \right). \quad (3)$$

Since X_ℓ and X'_ℓ are IID,

$$\begin{aligned} \mathbb{E}[(\psi_y(X_\ell) - \psi_y(X'_\ell))(\psi_{-z}(X_\ell) - \psi_{-z}(X'_\ell))] &= 2 \left(\mathbb{E}_{X \sim p} [\psi_{y-z}(X)] - \mathbb{E}_{X \sim p} [\psi_y(X)] \mathbb{E}_{X \sim p} [\psi_{-z}(X)] \right) \\ &= 2(\tilde{p}(y-z) - \tilde{p}(y)\tilde{p}(-z)), \end{aligned}$$

and, since Y_1, \dots, Y_n are IID,

$$\begin{aligned} \mathbb{E} \left[\left(\sum_{k=1}^n \psi_{-y}(Y_k) \right) \left(\sum_{k=1}^n \psi_z(Y_k) \right) \right] &= n \mathbb{E}_{Y \sim q} [\psi_{z-y}(Y)] + n(n-1) \mathbb{E}_{Y \sim q} [\psi_{-y}(Y)] \mathbb{E}_{Y \sim q} [\psi_z(Y)] \\ &= n\tilde{q}(z-y) + n(n-1)\tilde{q}(-y)\tilde{q}(z). \end{aligned}$$

In view of these two equalities, taking the expectation of (3) and using the fact that X_ℓ and X'_ℓ are independent of X_{n+1}, \dots, X_{2n} , (3) reduces:

$$\begin{aligned} \mathbb{E} \left[\left| \hat{S}_n - \hat{S}_n^{(\ell)} \right|^2 \right] &= \frac{2}{n^3} \sum_{|y|, |z| \leq Z_n} (yz)^{2s} (\tilde{p}(y-z) - \tilde{p}(y)\tilde{p}(-z)) (\tilde{q}(z-y) + (n-1)\tilde{q}(-y)\tilde{q}(z)) \\ &= \frac{2}{n^3} \sum_{|y|, |z| \leq Z_n} (yz)^{2s} (\tilde{p}(y-z)\tilde{q}(z-y) - \tilde{p}(y)\tilde{p}(-z)\tilde{q}(z-y) \\ &\quad + (n-1)\tilde{p}(y-z)\tilde{q}(-y)\tilde{q}(z) - (n-1)\tilde{p}(y)\tilde{p}(-z)\tilde{q}(-y)\tilde{q}(z)). \end{aligned} \quad (4)$$

We now need to bound following terms in magnitude:

$$\sum_{|y|, |z| \leq Z_n} (yz)^{2s} \tilde{p}(y-z)\tilde{q}(z-y), \quad (5)$$

$$\sum_{|y|, |z| \leq Z_n} (yz)^{2s} \tilde{p}(y-z)\tilde{q}(-y)\tilde{q}(z), \quad (6)$$

$$\text{and} \quad \sum_{|y|, |z| \leq Z_n} (yz)^{2s} \tilde{p}(y)\tilde{p}(-z)\tilde{q}(-y)\tilde{q}(z) \quad (7)$$

(the second term in (4) is bounded identically to the third term).

To bound (5), we perform a change of variables, replacing y by $k = y - z$:

$$\sum_{|y|, |z| \leq Z_n} (yz)^{2s} \tilde{p}(y-z)\tilde{q}(z-y) = \sum_{|k| \leq 2Z_n} \tilde{p}(k)\tilde{q}(-k) \sum_{j=1}^D \sum_{z_j = \max\{-Z_n, k_j - Z_n\}}^{\min\{Z_n, k_j + Z_n\}} ((k-z)z)^{2s} \quad (8)$$

$$\leq \frac{2^D \Gamma(4s+1)}{\Gamma(4s+D+1)} Z_n^{4s+D} \sum_{|k| \leq 2Z_n} \tilde{p}(k)\tilde{q}(-k) \quad (9)$$

$$\leq C_1 Z_n^{4s+D}, \quad (10)$$

where C_1 is the constant (in n and Z_n)

$$C_1 := \frac{2^D \Gamma(4s+1)}{\Gamma(4s+D+1)} \|p\|_2 \|q\|_2. \quad (11)$$

(8) and (9) follow from observing that

$$\sum_{j=1}^D \sum_{z_j = \max\{-Z_n, k_j - Z_n\}}^{\min\{Z_n, k_j + Z_n\}} ((k_j - z_j)z_j)^{2s} = (f * f)(k_j),$$

where $f(z) := z^{2s} 1_{\{|z| \leq Z_n\}}$, $\forall z \in \mathbb{Z}^D$ and $*$ denotes convolution (over \mathbb{Z}^D). This convolution is clearly maximized when $k = 0$, in which case

$$(f * f)(k) = \sum_{|z| \leq Z_n} z^{4s} \leq \left(\int_{B_\infty(0, Z_n)} z^{4s} dz \right) = \frac{2^D \Gamma(4s + 1)}{\Gamma(4s + D + 1)} Z_n^{4s+D},$$

where we upper bounded the series by an integral over

$$B_\infty(0, Z_n) := \{z \in \mathbb{R}^D : \|z\|_\infty = \max\{|z_1|, \dots, |z_D|\} \leq Z_n\}.$$

(10) then follows via Cauchy-Schwarz.

Bounding (6) for general s is more involved and requires rigorously defining more elaborate notions from the theory distributions, but the basic idea is as follows:

$$\begin{aligned} \sum_{|y|, |z| \leq Z_n} (yz)^{2s} \tilde{p}(y-z) \tilde{q}(-y) \tilde{q}(z) &= \sum_{|y| \leq Z_n} y^{2s} \tilde{q}(-y) \sum_{|z| \leq Z_n} z^{2s} \tilde{p}(y-z) \tilde{q}(z) \\ &= \sum_{|y|, |z| \leq Z_n} y^{2s} \tilde{q}(-y) \left(\widetilde{p_n^{(s)} q_n^{(s)}} \right)(y) \\ &\leq \sqrt{\sum_{|y| \leq Z_n} y^{2s} |\tilde{q}(y)|^2 \sum_{|y| \leq Z_n} y^{2s} \left(\widetilde{p_n^{(s)} q_n^{(s)}}(y) \right)^2} \\ &= \|q\|_{H^s} \|p_n^{(s)} q_n^{(s)}\|_{H^s} \leq \|q\|_{H^s} \|p_n\|_{W^{2s,4}} \|q_n\|_{W^{2s,4}}. \quad (12) \end{aligned}$$

Here, $p_n^{(s)}$ and $q_n^{(s)}$ denote s -order fractional derivatives of p_n and q_n , respectively, and $W^{2s,4}$ is a Sobolev space (with associated pseudonorm $\|\cdot\|_{W^{2s,4}}$), which can be informally thought of as $W^{2s,4} := \{p \in L^2 : (p^{(s)})^2 \in H^s\}$. The equality between the first and second lines follows from Proposition 6, and both inequalities are simply applications of Cauchy-Schwarz. For sake of intuition, it can be noted that the above steps are relatively elementary when $s = 0$. Now, it suffices to note that, by the Rellich-Kondrachov embedding theorem [Rellich, 1930, Evans, 2010], $W^{2s,4} \subseteq H^{s'}$, and hence $\|p_n\|_{W^{2s,4}} \leq \|p\|_{W^{2s,4}} < \infty$, as long as $s' \geq 2s + \frac{D}{4}$.

Bounding (7) is a simple application of Cauchy-Schwarz:

$$\begin{aligned} \sum_{|y|, |z| \leq Z_n} (yz)^{2s} \tilde{p}(y) \tilde{p}(-z) \tilde{q}(-y) \tilde{q}(z) &= \left(\sum_{|y| \leq Z_n} y^{2s} \tilde{p}(y) \tilde{q}(-y) \right) \left(\sum_{|z| \leq Z_n} z^{2s} \tilde{p}(-z) \tilde{q}(z) \right) \\ &\leq \left(\sum_{|z| \leq Z_n} z^{2s} |\tilde{p}(z)|^2 \right)^2 \left(\sum_{|z| \leq Z_n} z^{2s} |\tilde{q}(z)|^2 \right)^2 \\ &= \|p\|_{H^s}^4 \|q\|_{H^s}^4 \quad (13) \end{aligned}$$

Plugging (10), (12), and (13) into (4) gives

$$\mathbb{E} \left[\left| \hat{S}_n - \hat{S}_n^{(\ell)} \right|^2 \right] \leq 2C_1 \frac{Z_n^{4s+D}}{n^3} + \frac{C_2}{n^2},$$

where C_2 denotes the constant (in n and Z_n)

$$C_2 := (\|p\|_{H^s} + \|q\|_{H^s}) \|p\|_{W^{2s,4}} \|q\|_{W^{2s,4}} + \|p\|_{H^s}^4 \|q\|_{H^s}^4. \quad (14)$$

Plugging this into the Efron-Stein inequality (2) gives, by symmetry of \hat{S}_n in X_1, \dots, X_n ,

$$\mathbb{V} \left[\hat{S}_n \right] \leq 2C_1 \frac{Z_n^{4s+D}}{n^2} + \frac{C_2}{n}.$$

■

2 Proofs of Asymptotic Distributions

Theorem 4. (Asymptotic Normality) Suppose that, for some $s' > 2s + \frac{D}{4}$, $p, q \in H^{s'}$, and suppose $Z_n n^{\frac{1}{4(s-s')}} \rightarrow \infty$ and $Z_n n^{-\frac{1}{4s+D}} \rightarrow 0$ as $n \rightarrow \infty$. Then, \hat{S}_n is asymptotically normal with mean $\langle p, q \rangle$. In particular, for $j \in \{1, \dots, n\}$, define the following quantities:

$$W_j := \begin{bmatrix} Z_n^s e^{iZ_n X_j} \\ \vdots \\ e^{iX_j} \\ e^{iX_j} \\ \vdots \\ Z_n^s e^{-iZ_n X_j} \end{bmatrix}, \quad V_j := \begin{bmatrix} Z_n^s e^{iZ_n Y_j} \\ \vdots \\ e^{iY_j} \\ e^{iY_j} \\ \vdots \\ Z_n^s e^{-iZ_n Y_j} \end{bmatrix}, \quad \bar{W} := \frac{1}{n} \sum_{j=1}^n W_j, \quad \bar{V} := \frac{1}{n} \sum_{j=1}^n V_j \in \mathbb{R}^{2Z_n},$$

$$\Sigma_W := \frac{1}{n} \sum_{j=1}^n (W_j - \bar{W})(W_j - \bar{W})^T, \quad \text{and} \quad \Sigma_V := \frac{1}{n} \sum_{j=1}^n (V_j - \bar{V})(V_j - \bar{V})^T \in \mathbb{R}^{2Z_n \times 2Z_n}.$$

Then, for

$$\hat{\sigma}_n^2 := \begin{bmatrix} \bar{V} \\ \bar{W} \end{bmatrix}^T \begin{bmatrix} \Sigma_W & 0 \\ 0 & \Sigma_V \end{bmatrix} \begin{bmatrix} \bar{V} \\ \bar{W} \end{bmatrix},$$

we have

$$\sqrt{n} \left(\frac{\hat{S}_n - \langle p, q \rangle_{H^s}}{\hat{\sigma}_n} \right) \xrightarrow{D} \mathcal{N}(0, 1).$$

Proof: Let

$$\begin{aligned} \tilde{p}_{Z_n} &:= \begin{bmatrix} \tilde{p}(-Z_n) \\ \tilde{p}(-Z_n + 1) \\ \vdots \\ \tilde{p}(Z_n - 1) \\ \tilde{p}(Z_n) \end{bmatrix}, \quad \hat{p}_{Z_n} := \begin{bmatrix} \hat{p}(-Z_n) \\ \hat{p}(-Z_n + 1) \\ \vdots \\ \hat{p}(Z_n - 1) \\ \hat{p}(Z_n) \end{bmatrix}, \\ \tilde{q}_{Z_n} &:= \begin{bmatrix} \tilde{q}(-Z_n) \\ \tilde{q}(-Z_n + 1) \\ \vdots \\ \tilde{q}(Z_n - 1) \\ \tilde{q}(Z_n) \end{bmatrix}, \quad \text{and} \quad \hat{q}_{Z_n} := \begin{bmatrix} \hat{q}(-Z_n) \\ \hat{q}(-Z_n + 1) \\ \vdots \\ \hat{q}(Z_n - 1) \\ \hat{q}(Z_n) \end{bmatrix}. \end{aligned}$$

Also let

$$\sigma_n^2 := (\nabla h(\tilde{p}_{Z_n}, \tilde{q}_{Z_n}))' \begin{bmatrix} \Sigma_p & 0 \\ 0 & \Sigma_q \end{bmatrix} (\nabla h(\tilde{p}_{Z_n}, \tilde{q}_{Z_n})),$$

where, $h : \mathbb{R}^{2Z_n+1} \times \mathbb{R}^{2Z_n+1} \rightarrow \mathbb{R}$ is defined by $h(x, y) = \sum_{z=-Z_n}^{Z_n} z^{2s} x_z y_{-z}$. By the bias bound and the assumption $Z_n^{4(s-s')} n \rightarrow \infty$, it suffices to show

$$\sqrt{n} \left(\frac{\hat{S}_n - \mathbb{E}[\hat{S}_n]}{\sigma_n} \right) \xrightarrow{D} \mathcal{N}(0, 1) \quad \text{as } n \rightarrow \infty. \quad (15)$$

Since \hat{p}_{Z_n} and \hat{q}_{Z_n} are empirical means of bounded random vectors with means \tilde{p}_{Z_n} and \tilde{q}_{Z_n} , respectively, by the central limit theorem, as $n \rightarrow \infty$,

$$\sqrt{n}(\hat{p}_{Z_n} - \tilde{p}_{Z_n}) \xrightarrow{D} \mathcal{N}(0, \Sigma_p) \quad \text{and} \quad \sqrt{n}(\hat{q}_{Z_n} - \tilde{q}_{Z_n}) \xrightarrow{D} \mathcal{N}(0, \Sigma_q),$$

where

$$(\Sigma_p)_{w,z} := \text{Cov}_{X \sim p}(\psi_w(X), \psi_z(X)) \quad \text{and} \quad (\Sigma_q)_{w,z} := \text{Cov}_{X \sim q}(\psi_w(X), \psi_z(X)).$$

(15) follows by the delta method. ■

Theorem 5. (Asymptotic Null Distribution) Suppose that, for some $s' > 2s + \frac{D}{4}$, $p, q \in H^{s'}$, and suppose $Z_n n^{\frac{1}{4(s-s')}} \rightarrow \infty$ and $Z_n n^{-\frac{1}{4s+D}} \rightarrow 0$ as $n \rightarrow \infty$. For $j \in \{1, \dots, n\}$, define

$$W_j := \begin{bmatrix} Z_n^s (e^{iZ_n X_j} - e^{iZ_n Y_j}) \\ \vdots \\ e^{iX_j} - e^{iY_j} \\ e^{-iX_j} - e^{-iY_j} \\ \vdots \\ Z_n^s (e^{-iZ_n X_j} - e^{-iZ_n Y_j}) \end{bmatrix} \in \mathbb{R}^{2Z_n}.$$

Let

$$\bar{W} := \frac{1}{n} \sum_{j=1}^n W_j \quad \text{and} \quad \Sigma := \frac{1}{n} \sum_{j=1}^n (W_j - \bar{W})(W_j - \bar{W})^T$$

denote the empirical mean and covariance of W , and define $T := n\bar{W}^T \Sigma^{-1} \bar{W}$. Then, if $p = q$, then

$$Q_{\chi^2(2Z_n)}(T) \xrightarrow{D} \text{Uniform}([0, 1]) \quad \text{as} \quad n \rightarrow \infty,$$

where $Q_{\chi^2(2Z_n)} : [0, \infty) \rightarrow [0, 1]$ denotes the quantile function (inverse CDF) of the χ^2 distribution $\chi^2(2Z_n)$ with $2Z_n$ degrees of freedom.

Proof: Since, as shown in the proof of the previous theorem, the distance estimate is a sum of squared asymptotically normal, zero-mean random variables, this is a standard result in multivariate statistics. See, for example, Theorem 5.2.3 of Anderson [2003]. \blacksquare

3 Generalizations: Weak and Fractional Derivatives

As mentioned in the main text, our estimator and analysis can be generalized nicely to non-integer s using an appropriate notion of fractional derivative.

For non-negative integers s , let $\delta^{(s)}$ denote the measure underlying of the s -order derivative operator at 0; that is, $\delta^{(s)}$ is the distribution such that

$$\int_{\mathbb{R}} f(x) \delta^{(s)}(x) dx = f^{(s)}(0),$$

for all test functions $f \in H^s$. Then, for all $z \in \mathbb{R}$, the Fourier transform of $\delta^{(s)}$ is

$$\tilde{\delta}(z) = \int_{\mathbb{R}} e^{-izx} \delta^{(s)}(x) dx = (-iz)^s.$$

Thus, we can naturally generalize the derivative operator $\delta^{(s)}$ to general $s \in [0, \infty)$ as the inverse Fourier transform of the function $z \mapsto (-iz)^s$. Generalization to differentiation at an arbitrary $y \in \mathbb{R}$ follows from translation properties of the Fourier transform, and, in multiple dimensions, for $s \in \mathbb{R}^D$, we can consider the inverse Fourier transform of $z \in \mathbb{R}^D \mapsto \prod_{j=1}^D (iz_j)^{s_j}$.

With this definition in place, we can prove the following the Convolution Theorem, which equates a particular weighted convolution of Fourier transforms and a product of particular fractional derivatives. Note that we will only need this result in the case that f is a trigonometric polynomial (i.e., \tilde{f} has finite support), because we apply it only to p_n and q_n . Hence, the sum below has only finitely many non-zero terms and commutes freely with integrals.

Proposition 6. Suppose $p, q \in L^2$ are trigonometric polynomials. Then, $\forall s \in [0, \infty)$, and $y \in \mathbb{Z}^D$,

$$\sum_{z \in \mathbb{Z}^D} z^{2s} \tilde{p}(y - z) \tilde{q}(z) = \widetilde{(p^{(s)} q^{(s)})}(y).$$

Proof: By linearity of the integral,

$$\begin{aligned}
\sum_{z \in \mathbb{Z}^D} z^{2s} \tilde{p}(y-z) \tilde{q}(z) &= \sum_{z \in \mathbb{Z}^D} z^{2s} \int_{\mathbb{R}^D} p(x_1) e^{-i\langle y-z, x_1 \rangle} dx_1 \int_{\mathbb{R}^D} q(x_2) e^{-i\langle z, x_2 \rangle} dx_2 \\
&= \int_{\mathbb{R}^D} \int_{\mathbb{R}^D} p(x_1) q(x_2) e^{-i\langle y, x_1 \rangle} \sum_{z \in \mathbb{Z}^D} z^{2s} e^{i\langle z, x_1 - x_2 \rangle} dx_1 dx_2 \\
&= \int_{\mathbb{R}^D} \int_{\mathbb{R}^D} p(x_1) q(x_2) e^{-i\langle y, x_1 \rangle} \delta^{(s)}(x_1 - x_2) dx_1 dx_2 \\
&= \int_{\mathbb{R}^D} p^{(s)}(x) q^{(s)}(x) e^{-i\langle y, x \rangle} dx = (\widetilde{p^{(s)} q^{(s)}})(y).
\end{aligned}$$

■

4 Additional Experimental Results

Figure 1 presents results of one additional experiment showing the effect of increasing s on the convergence rate. In all four cases, we estimate squared distance of the density $p(x) = 1 + \cos(2\pi x)$ from the uniform density (both on the interval $[0, 1]$), but we vary s , such that the distance is computed according to different metrics. As s increases, the bias of the estimator increases, due to greater weight on higher frequencies of the density that are omitted by the truncated estimator.

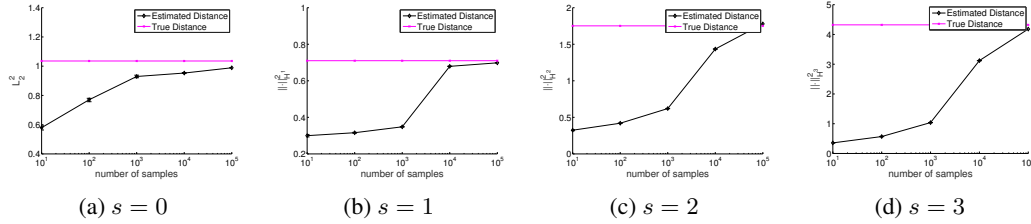


Figure 1: Estimated and true s -order Sobolev distances between the density $p(x) = 1 + \cos(2\pi x)$ and the uniform density, for $s \in \{0, 1, 2, 3\}$.

References

- TW Anderson. An introduction to multivariate statistical analysis. Wiley, 2003.
- Bradley Efron and Charles Stein. The jackknife estimate of variance. *The Annals of Statistics*, pages 586–596, 1981.
- Lawrence C Evans. *Partial differential equations*. American Mathematical Society, 2010.
- Franz Rellich. Ein satz über mittlere konvergenz. *Nachrichten von der Gesellschaft der Wissenschaften zu Göttingen, Mathematisch-Physikalische Klasse*, 1930:30–35, 1930.