A Appendix

A.1 Proofs

Theorem 1. *With $\hat{F}_N^{KDI}(x;h)$ as defined in Eq (6), $\lim_{h\to\infty} \hat{F}_N^{KDI}(x;h) = \hat{S}_N(x)$.*

Proof. Our proof relies upon our earlier observation that the min-max transformation $\hat{S}_N(x)$ is the c.d.f of a uniform distribution over $[X_{(1)}, X_{(N)}]$, showing that as $h \to \infty$, the distribution with c.d.f. represented by our transformation converges to this uniform distribution. First, we claim that a bounded distribution over a fixed region $[A, B]$ with density proportional to the density of a normal distribution $\mathcal{N}(\mu, \sigma^2$ over that same region, converges as $\sigma \to \infty$ to the uniform distribution over $[A, B]$. Relying heavily upon the proof of (Yuan, 2023), this bounded distribution has density

$$g(x) = \frac{\exp\left(-\frac{(x-\mu)^2}{\sigma^2}\right)}{\int_A^B \exp\left(-\frac{(x-\mu)^2}{\sigma^2}\right)}.$$

The unnormalized density is $\tilde{g}(x) = \exp\left(-\frac{((x-\mu)^2}{\sigma^2}\right)$, with $\tilde{g}(x) \leq 1$ and $\tilde{g}(x) > \exp\left(-\frac{\max((A-\mu)^2,(B-\mu)^2)}{\sigma^2}\right)$. Thus, we have

$$\frac{\exp\left(-\frac{\max((A-\mu)^2,(B-\mu)^2)}{\sigma^2}\right)}{B-A} \leq g(x) = \frac{\exp\left(-\frac{(x-\mu)^2}{\sigma^2}\right)}{\int_A^B \exp\left(-\frac{(x-\mu)^2}{\sigma^2}\right)} \leq \frac{1}{(B-A)\exp\left(-\frac{\max((A-\mu)^2,(B-\mu)^2)}{\sigma^2}\right)},$$

and therefore $\lim_{\sigma\to\infty} g(x) = \frac{1}{B-A}$.

Without loss of generality, consider a particular sample X_i, and the corresponding distribution $\mathcal{N}(X_i, h^2)$. Plugging in $\mu := X_i, \sigma := h, A := X_{(1)}, B := X_{(N)}$, we have that the bounded distribution based on $\mathcal{N}(X_i, h^2)$ converges to the uniform distribution over $[A, B]$. Because samples are assumed to be i.i.d., we have $\hat{F}_N^{\text{KDI}}(x;h)$ as the c.d.f. of a distribution whose p.d.f. is the mean of these bounded distributions. The p.d.f. therefore converges to the p.d.f of the uniform distribution, and therefore $\hat{F}_N^{\text{KDI}}(x;h)$ converges to the c.d.f. of the uniform distribution, which is $\hat{S}_N(x)$.

□

Theorem 2. *With $\hat{F}_N^{KDI}(x;h)$ as defined in Eq (6), $\lim_{h\to 0} \hat{F}_N^{KDI}(x;h) = \hat{F}_N(x)$.*

Proof. We first show that $\hat{F}_N^{\text{KDI,naive}}(x;h)$ as defined in Eq (5) converges to $\hat{F}_N(x)$. We note that the KDE distribution represented by Eq. 3 converges to the empirical distribution as $h \to 0$, and therefore the c.d.f. of the KDE converges to the c.d.f. of the empirical distribution. The c.d.f. of the KDE is $\hat{F}_N^{\text{KDI,naive}}(x;h)$, and the c.d.f. of the empirical distribution is the quantile transformation function $\hat{F}_N(x)$. Therefore, $\lim_{h\to 0} \hat{F}_N^{\text{KDI,naive}}(x;h) = \hat{F}_N(x)$.

We next address $\hat{F}_N^{\text{KDI}}(x;h)$, utilizing the above result. By construction, $\hat{F}_N^{\text{KDI}}(x;h) = \hat{F}_N(x)$ for $x < X_{(1)}$ and $X_{(N)} \leq x$. For $[X_{(1)}, X_{(N)})$, we show that $\lim_{h\to 0} \hat{F}_N^{\text{KDI}}(x;h) = \lim_{h\to 0} \hat{F}_N^{\text{KDI,naive}}(x;h)$ on this interval.

$$\lim_{h\to 0} \hat{F}_N^{\text{KDI}}(x;h) = \lim_{h\to 0} \frac{P_h(X_{(1)}, x)}{P_h(X_{(1)}, X_{(N)})} \tag{7}$$

$$= \lim_{h\to 0} \frac{P_h(-\infty, x) - P_h(-\infty, X_{(1)})}{1 - P_h(-\infty, X_{(1)}) - P_h(X_{(N)}, \infty)} \tag{8}$$

$$= \frac{\lim_{h\to 0} P_h(-\infty, x) - 0}{1 - 0 - 0} \tag{9}$$

$$= \lim_{h\to 0} \hat{F}_N^{\text{KDI,naive}}(x;h), \tag{10}$$

where the second last step uses $\lim_{h\to 0} P_h(-\infty, X_{(1)}) = \lim_{h\to 0} P_h(X_{(N)}, \infty) = 0$ and continuity at these values. □

A.2 Implementation details for polynomial-exponential kernel

The poly$(|x|)e^{-|x|}$ class of kernels is defined as the set of kernels which can be represented as

$$K_\kappa(x) = Ce^{-|x|}\sum_{i=0}^{\kappa}\beta_i|x|^i, \tag{11}$$

where κ is the order of the polynomial and of the kernel, and C is the normalizing constant. We use kernels from the differentiable subclass of the poly$(|x|)e^{-|x|}$ kernels. Namely, for order κ, we choose coefficients β_i as $\beta_{i\in\{0,\dots,K\}} \propto 1/i!$, so that the kernel takes the form

$$K_\kappa(x) = \frac{1}{2(\kappa+1)}\sum_{i=0}^{\kappa}\frac{|x|^i}{i!}e^{-|x|}, \quad \kappa = 1, 2, \dots \tag{12}$$

To make the polynomial-exponential kernel approximate the Gaussian kernel as closely as possible, we need to rescale the Gaussian kernel bandwidth. For a given Gaussian kernel bandwidth $h_{\mathcal{N}} = \alpha\sigma_X$, the polynomial-exponential kernel bandwidth is calculated as follows:

$$h_{\text{polyexp}} = \left(\frac{2\sqrt{\pi}\sum_{i=0}^{\kappa}\sum_{j=0}^{\kappa}\frac{\beta_i\beta_j}{2^{i+j}}(i+j)!}{\left(2\sum_{i=0}^{\kappa}\beta_i(i+2)!\right)^2}\right)^{0.2} h_{\mathcal{N}}. \tag{13}$$

This rescaling factor is derived from the asymptotic mean integrated squared error (AMISE) optimal bandwidth for density estimation. For an arbitrary kernel K, its bandwidth h is AMISE optimal if

$$h \propto \left(\frac{\|K_h\|^2}{(\sigma^2_{K_h})^2}\right)^{0.2}.$$

Obtaining the expressions on the right-hand side for both the Gaussian kernel and the polynomial-exponential kernel, then setting them equal to each other, gives the above rescaling factor. The needed expressions for the polynomial-exponential kernel are given in (Hofmeyr, 2019).

A.3 Additional experiments on bandwidth vs sample size

We repeat the experiments from Figure 4(A) with CA Housing (with $N = 20460$), but with subsampling to determine whether the optimal bandwidth depends on sample size. We follow the same experimental setup as before, but in each of the 100 simulations, we use an independent subsample (without replacement) of 1000, 2000, 5000, and 10000 samples. Results are shown in Figure 12. We see that the optimal bandwidth stays constant rather than vanishing as the sample size increases.

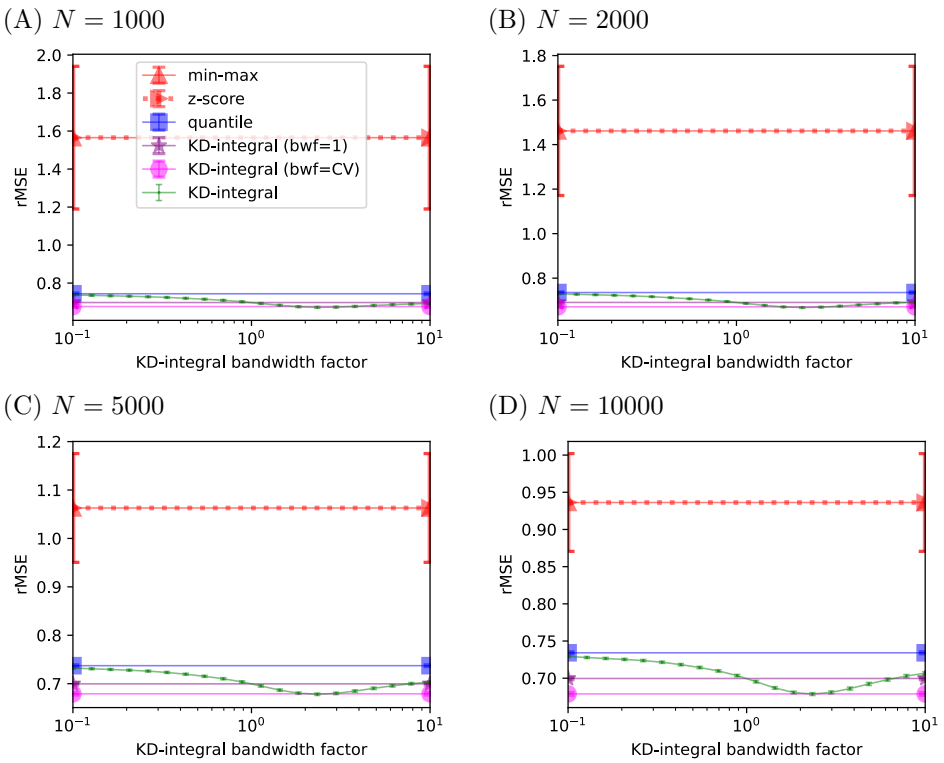

Figure 12: Root mean-squared error (rMSE) on subsampled CA housing datasets, with reduced sizes (A) $N = 1000$, (B) $N = 2000$, (C) $N = 5000$, and (D) $N = 10000$.