

1. The optimal bandwidth for a kernel density estimate, h_{norm} , is given by

$$h_{norm} = \left(\frac{\|K\|_2^2}{\|f''\|_2^2 \mu_2^2(K)} \right)^{-1/5} n^{-1/5}$$

Notice that this choice is dependent on the underlying distribution f (the very thing we are trying to estimate). One method to choose the optimal bandwidth is assume f is approximately normal with standard deviation σ^2 . First, we collect some results that will be helpful in our derivation. If f is the kernel of $N(0, \sigma^2)$ and ϕ is the kernel of $N(0, 1)$

$$f''(x) = \frac{x^2 - \sigma^2}{\sqrt{2\pi}\sigma^5} e^{-x^2/2\sigma^2} \quad \phi''(x) = \frac{x^2 - 1}{\sqrt{2\pi}} e^{-x^2/2}$$

$$\|f''\|_2^2 = \int (f''(x))^2 dx = \int \left(\frac{x^2 - \sigma^2}{\sqrt{2\pi}\sigma^5} e^{-x^2/2\sigma^2} \right)^2 dx$$

Now completing a substitution $z = x/\sigma$, $\sigma dz = dx$ we have

$$= \int \left(\frac{z^2 - 1}{\sqrt{2\pi}\sigma^3} e^{-z^2/2} \right)^2 \sigma dz = \frac{\sigma}{\sigma^6} \int \left(\frac{z^2 - 1}{\sqrt{2\pi}} e^{-z^2/2} \right)^2 dz = \sigma^{-5} \|\phi''\|_2^2$$

Now to evaluate $\|\phi''\|_2^2$ consider the following

$$\begin{aligned} \sigma^{-5} \int (\phi(z''))^2 dz &= \frac{1}{2\pi\sigma^5} \int (z^2 - 1)^2 e^{-z^2} dz \\ &= \frac{1}{2\pi\sigma^5} \int (z^4 - 2z^2 + 1) e^{-z^2} dz \\ &= \frac{\sqrt{1/2}}{\sqrt{2\pi}\sigma^5} \frac{1}{\sqrt{2\pi}\sqrt{1/2}} \left(\int z^4 e^{-z^2} dz - 2 \int z^2 e^{-z^2} dz + \int e^{-z^2} dz \right) \end{aligned}$$

We recognize each integrand and the kernel of a normal random variable with mean zero variance $1/2$, $N(0, 1/2)$. Seeing the distribution is mean zero, we see the three terms in the sum are the Kurtosis (which we find via the MGF), Variance, and density respectively. Thus,

$$= \frac{1}{2\sqrt{\pi}\sigma^5} \left(\frac{3}{4} - 2 \cdot \frac{1}{2} + 1 \right) = \frac{3\sigma^{-5}}{8\sqrt{\pi}}$$

Note further for the standard normal kernel K , we have the following

$$\|K\|_2^2 = \frac{1}{2\pi} \int e^{-x^2} dx = \frac{\sqrt{1/2}}{\sqrt{2\pi}} \frac{1}{\sqrt{2\pi}\sqrt{1/2}} \int e^{-x^2} dx = \frac{1}{2\sqrt{\pi}}$$

where the third equality is due to the fact that the integral is simply the kernel of a Gaussian random variable with mean zero and variance $1/2$. Now lastly, since K is mean zero, we see that $\mu_2(K) = 1$, the variance of K . Moreover, $\mu_2^2(K) = 1$. Using all of these equations, we have

$$h_{norm} = \left(\frac{1/2\sqrt{\pi}}{\frac{3}{8}\sigma^{-5}\sqrt{\pi}} \right)^{-1/5} n^{-1/5} = (4/3)^{-1/5} \sigma^{-1/5} \approx 1.06\hat{\sigma}n^{-1/5}$$

2. First consider the following equality.

$$\mathbb{E}(\|\widehat{f_h''}\|_2^2) = \mathbb{E}\left[\int \left(\frac{1}{nh^3} \sum_{i=1}^n K''\left(\frac{x_i - x}{h}\right)\right)^2 dx\right] = \frac{1}{n^2 h^6} \int \mathbb{E}\left[\left(\sum_{i=1}^n K''\left(\frac{x_i - x}{h}\right)\right)^2\right] dx$$

Now rewriting this sum in terms of variances, we have

$$\begin{aligned} &= \frac{1}{n^2 h^6} \int \text{Var}\left(\sum_{i=1}^n K''\left(\frac{X_i - x}{h}\right)\right) dx + \frac{1}{n^2 h^6} \int \left[\mathbb{E}\left(\sum_{i=1}^n K''\left(\frac{X_i - x}{h}\right)\right)\right]^2 dx \\ &\stackrel{iid}{=} \frac{1}{nh^6} \int \text{Var}\left(K''\left(\frac{X - x}{h}\right)\right) dx + \frac{1}{h^6} \int \left[\mathbb{E}\left(K''\left(\frac{X - x}{h}\right)\right)\right]^2 dx \\ &= \frac{1}{nh^6} \int \int \left[K''\left(\frac{w - x}{h}\right) - \mu(K'')\right]^2 f(w) dw dx + \frac{1}{nh^6} \int \left[\int K''\left(\frac{w - x}{h}\right) f(w) dw\right]^2 dx \\ &= \frac{1}{nh^5} \int \int (K''(z))^2 f(x + zh) dz dx + \frac{1}{nh^4} \int \left[\int K''(z) f(x + zh) dz\right]^2 dx \\ &= \frac{1}{nh^5} \int \int (K''(z))^2 \{f(x) + O(h)\} dz dx + \frac{1}{nh^4} \int \left[\int K''(z) f(x + zh) dz\right]^2 dx \end{aligned}$$

Now for the second part of this equation, we can integrate by parts twice to achieve the desired result. We integrated in such a manor in the previous HW assignment and found

$$\frac{1}{h^2} \int K''(z) f(x - zh) dz = \int K(z) f''(x - zh) dz$$

Using this result, we have

$$\begin{aligned} &= \frac{1}{nh^5} \int \int (K''(z))^2 \{f(x) + O(h)\} dz dx + \frac{1}{h^4} \int \left[h^2 \int K(z) f''(x - zh) dz\right]^2 dx \\ &= \frac{\|K''\|_2^2}{nh^5} \int f(x) dx + O(h) + \int \left[\int K(z) \{f''(x) + O(h)\} dz\right]^2 dx \\ &= \frac{\|K''\|_2^2}{nh^5} + O(h) + \int (f''(x))^2 \left[\int K(z) dz\right]^2 + O(h^2) \\ &= \|f''\|_2^2 + \frac{\|K''\|_2^2}{nh^5} + O(h^2) \end{aligned}$$

3.

$$\begin{aligned} \mathbb{E}(\widehat{f}(0)) &= \mathbb{E}\left[\frac{1}{nh} \sum_{i=1}^n K\left(\frac{x_i - 0}{h}\right)\right] \\ &\stackrel{iid}{=} \frac{1}{h} \mathbb{E}\left[K\left(\frac{X}{h}\right)\right] \\ &= \frac{1}{h} \int_0^1 K(x/h) f(x) dx \\ &\stackrel{z-sub}{=} \int_0^{1/h} K(z) f(z) dz \\ &\stackrel{taylor}{=} \int_0^{1/h} K(z) \{f(0) + O(h)\} dz \end{aligned}$$

Recall that as $n \rightarrow \infty$, $h \rightarrow 0$ albeit at a slower rate. Thus we see that

$$\lim_{h \rightarrow 0} \mathbb{E}(\hat{f}(0)) = \lim_{h \rightarrow \infty} f(0) \int_0^{1/h} K(z) dz + O(H) = f(0) \int_0^\infty K(z) dz = \frac{f(0)}{2}$$

Thus we see that even asymptotically, \hat{f} is biased at the boundaries. This implies that even as we collect more data, the boundary estimates will be biased. In addition, under the kernel density procedure, the mass that is not allocated to the boundary positions will be allocated to other estimates near the boundary. This, means that biased estimates at the boundary will negatively impact estimates throughout the range of the density.

4. In both papers, the concept of developing an “oversmoothed” bandwidth selection procedure introduced by Terrel and Scott (1985) and Terrell (1990) was introduced. The authors attempt to find an upper bound on how smooth the density estimate while simultaneously accounting for asymptotic mean integrated square error (AMISE). They find this oversmoothed bandwidth by finding \hat{f}'' that minimizes $\|\hat{f}''\|_2^2$. For a given kernel (with variance σ^2) \hat{f}'' maximizes the equation

$$h_{AMISE} = \left[\frac{\|K\|_2^2}{\mu(K)^2 \|\hat{f}''\|_2^2} \right]^{1/5} n^{-1/5}$$

Heuristically, this method is accounting for a “worst case” underlying density f . $\|\hat{f}''\|_2^2$ measures the roughness of the density in question. By maximizing our estimate of $\|\hat{f}''\|_2^2$ we account for the most wiggly density to set an upper bound for the behavior of f .

Terrell (1990) go on to show that \hat{f}'' that minimizes $\|\hat{f}''\|_2^2$ corresponds to the parameter family $\beta(4, 4)$. Moreover, using a normal kernel, they show that the optimal “oversmoothed” bandwidth is given by $h_{os} = 1.144 S n^{-1/5}$ where S is the sample standard deviation. Here, we see that this estimate is in fact larger than most estimates (such as the one found in problem 1). This method is specifically very useful as a starting point for choosing an optimal h . Many data-dependent cross validation schemes require searching over a candidate set \mathcal{H} . By finding a type of upper bound for optimal parameter selection, practitioners can more effectively choose their candidate set \mathcal{H} . This has great implications in both computational efficiency (smaller \mathcal{H} implies shorter search) as well as restricts the problem to a simpler class of density estimates.

Another interesting topic covered in the Sheather survey is one of Local Likelihood Density estimates. This method was specifically introduced to overcome the boundary problem we discussed in class. We solved the problem by combining two kernels. Here, they fit local models in a neighborhood of x of the form

$$\psi(u) = \exp \left[\sum_{i=1}^n \theta_j (u - x)^j \right]$$

where the θ_j are fit via maximum likelihood. By changing the number of parameters fit (p increases) we can effectively remove the boundary bias for $p > 1$. For $p = 1$ we still see there is asymptotic bias (as in problem 3) but remove the bias thereafter.

On the surface this approach seems preferable to nonparametric boundary kernels - the rate of convergence is better, there is no boundary bias, and the model itself is simpler. However, in the case that there are few boundary points, Hall and Toa (2002) note that the

nonparametric approach is preferable to the log-polynomial model in terms of Integrated Square Bias (ISB). This again shows the flexibility of nonparametric methods. Given that we do not know the underlying distribution, and thus do not understand its boundary behavior, the nonparametric model can be preferable to the parametric model in some cases. Moreover, as p increases, we need to solve more complex systems of equations. This implies that even the nonparametric method may have comparable computational performance and the parametric model.