

EXERCISES (NON-PARAMETRIC ESTIMATION)

ETHAN LIGON

- (1) Let $f(x)$ be a density, and let $\hat{f}(x) = \frac{1}{nh} \sum_i k\left(\frac{\mathbf{x}_i - x}{h}\right)$ be an estimator of $f(x)$. These aren't generally equal, of course. But show that if k is a valid positive kernel then \hat{f} is also a density. (Hint: use fact that k is a density and use rules for transforming random variables.)
- (2) The density for the standard Cauchy distribution is $f(x) = \frac{1}{\pi(1+x^2)}$. Apparently this could be used as a kernel. Of the desirable properties we've discussed (non-negative, boundedness, symmetry, differentiability) which are possessed by the Cauchy kernel?
- (3) In lecture we derived an expression for the bias of $\hat{f}(x)$ in terms of the true density f . Use an analogous argument to obtain an expression for the variance in terms of f .
- (4) You use data from the Indian NSS to produce a figure describing the distribution of non-durable expenditures across households, measured in INR, and using a kernel estimator with bandwidth h . A referee asks you to re-estimate the distribution after converting the expenditure data so that the units are in USD, using the 2014 PPP rate of 18.4 INR/USD. What new bandwidth should you use?
- (5) There's a sense in which the Epanechnikov kernel,

$$k(u) = \begin{cases} \frac{3}{4\sqrt{5}} \left(1 - \frac{u^2}{5}\right) & \text{if } |u| < \sqrt{5} \\ 0 & \text{otherwise} \end{cases}$$

is optimal (it minimizes the asymptotic integrated mean square error of the estimator). Using the same code we developed in lecture as a base, implement the Epanechnikov kernel, and compare its performance in terms of the Oracle calculations of bias and variance with the Gaussian kernel.

(6) Consider the non-parametric regression model

$$y = m(X) + u$$

$$\mathbb{E}(u|X) = 0$$

$$\mathbb{E}(u^2|X) = \sigma^2(X).$$

Suppose that (unknown to the econometrician) the ‘true’ $m(x) = \alpha + \beta x$. The econometrician estimates the non-parametric model using kernel regression. What is the bias of the estimator? How does this depend on the sign of β ?