Econ 589 Assignment 1 — due October 6 via Canvas

You can do this assignment by yourself or with up to two others. Please only submit one assignment for the entire group. Some of the problems can be solved easily by looking things up on the web or by using software, except for question 4 of course.

1. Get yourself a Github account and request access to Grumps.jl. Find a real dataset to estimate a mixed logit discrete choice model with both individual consumer and product share data. Formulate a model and estimate it. Report any problems you encounter.

2. (a) Show that for any μ and any $\sigma > 0$,

$$\frac{1}{\sigma}\int \phi\Big(\frac{x-\mu}{\sigma}\Big)\phi(x)\,\mathrm{d}x = \frac{1}{\sqrt{\sigma^2+1}}\phi\Big(\frac{\mu}{\sqrt{\sigma^2+1}}\Big).$$

   (b) Suppose you used the Rosenblatt-Parzen estimator with a standard normal kernel and bandwidth $h > 0$. Suppose that the true density function is a standard normal. Let $U(h)$ be the values of x (expressed in terms of h) for which the estimator is upward biased? Determine $U(h)$. What is $\lim_{h \downarrow 0} U(h)$? Can you explain why that would be? (This is not specific to this setting; think of a general theme / lesson to be learned here.)

   (c) Work out the formula for the 'twicing kernel' $k(u) = 2\phi(u) - \int \phi(v-u)\phi(v)\,\mathrm{d}v$ and show that it is a fourth order kernel.

3. Suppose that a density f is twice continuously differentiable at x and that you use the Rosenblatt Parzen estimator with a nonnegative kernel k defined on $[-1, 1]$.

   (a) Determine the mean square error of $\hat{f}(x)$ as a function of $f(x), f''(x), h, n, k$ if bias terms beyond h^2 are ignored.

   (b) Determine the optimal bandwidth h as a function of $f(x), f''(x), n, k$?

   (c) Determine the optimal kernel, which miraculously does not depend on anything else.....

   (d) Now consider the estimator

$$\hat{f}^*(x) = \max\Big(0, \frac{1}{nh}\sum_{i=1}^{n} k^*\Big(\frac{x-\boldsymbol{x}_i}{h}\Big)\Big)$$

   for $k^*(u) = 3(3-5u^2)\mathbb{1}(|u| \leq 1)/8$. Show that $\hat{f}^*$ can achieve a better convergence rate than $\hat{f}$ and also produces nonnegative estimates.

4. For each of $d = 1, 2, 4, 8, 16$ and $n = 100, 1000, 10000$, report the bias, standard deviation, root mean square error, and median absolute error of $\hat{f}(0)$ using a cross-validated bandwidth with data drawn from a d-variate standard normal, using a standard normal kernel, a cross-validated bandwidth, and 100 replications. Also report the same numbers when normalized by $f(0)$, e.g. for the bias report the average of $\{\hat{f}(0) - f(0)\}/f(0)$.