STAT 526 TERM 2 2002/3 Homework 4
Due Monday March 31

Consider estimating $\hat{\mu}'$ based on regression data (x_i, Y_i) $Y_i = \mu(x_i) + \epsilon_i$, ϵ_i iid mean 0 variance σ^2. You'll use a local quadratic estimate.

-1- Use the theorem presented in class to calculate the asymptotic bias and variance of $\hat{\mu}'(x)$, for x an interior point, K a symmetric kernel. You should get

$$Bias \sim \frac{h^2}{6}\mu'''(x)\frac{\nu_4(K)}{\nu_2(K)}$$

$$var \sim \frac{\sigma^2}{nh^3 f(x)}\frac{\nu_2(K^2)}{\nu_2^2(K)}$$

-2- Using -1- and ignoring the boundary (which you can, in this case), find the asymptotic integrated mean squared error. Suppose that the weight function $w \equiv 1$ and the x_i's are uniformly distributed on $[0, 1]$.

-3- Based on your answer and assumptions in -2-, what is the (asymptotically) optimal bandwidth h?

-4- Data analysis: consider the data from homework 3, for Casey and Hudson. For each whale, estimate the rate of nursing (eg the derivative of the regression function). As your calculations above assume that the x_i's are in $[0, 1]$, you should transform your time scales (eg 1:228 $\rightarrow$ [1/229:228/229]). *Estimate of h:* Use an estimate of the optimal h calculated in -3- using a first generation rule of thumb. Specifically, fit the regression function by a piecewise polynomial (you can choose the degree) with $N =$ around 10-15 pieces. If you'd like to try another first generation plug-in type method, that's OK - just describe explicitly what you've done. You can use Rice's estimate of the variance. *Estimate of the derivative:* Use locpoly (available in R - library(KernSmooth) and in Splus - not Splus 6 - on hajek). Locpoly uses a standard Gaussian kernel. In your calculations, please use the fact that the moment generating function for Z a standard normal random variable, is $\psi(t) = E(\exp(itZ)) = \exp(t^2/2)$.