EXAM C QUESTIONS OF THE WEEK

S. Broverman, 2008

Week of March 3/08

A uniform kernel with bandwidth b is used to estimate $f(x)$, the pdf of X, based on a random sample of size n from the distribution of X. The kernel density estimator is $\widehat{f}_n(x)$.

A new sample value becomes available, and the kernel density estimator is reformulated based on $n+1$ sample values. The new kernel density estimator is $\widehat{f}_{n+1}(x)$.

Find the minimum possible value of $\widehat{f}_{n+1}(x) - \widehat{f}_n(x)$ over all values of x.

The solution can be found below.

Week of March 3/08 - Solution

The empirical distribution for the sample of size n assigns a probability of $\frac{1}{n}$ to each of the original sample values, and $\frac{1}{n+1}$ to each sample value in the random sample of size $n-1$.

The new estimator is $\widehat{f}_{n+1}(x) = \frac{n}{n+1} \cdot \widehat{f}_n(x) + \frac{1}{n+1} \cdot k_{x_{n+1}}(x)$.

Then

$\widehat{f}_{n+1}(x) - \widehat{f}_n(x) = \frac{n}{n+1} \cdot \widehat{f}_n(x) + \frac{1}{n+1} \cdot k_{x_{n+1}}(x) - \widehat{f}_n(x)$

$= \frac{1}{n+1} \cdot [k_{x_{n+1}}(x) - \widehat{f}_n(x)]$.

The maximum possible value of $\widehat{f}_n(x)$ is $\frac{1}{2b}$ and the minimum possible value of $k_{x_{n+1}}(x)$ is 0.

The minimum value of the difference will occur if $k_{x_{n+1}}(x) = 0$ and $\widehat{f}_n(x) = \frac{1}{2b}$,

and the difference is $\frac{1}{n+1} \cdot [-\frac{1}{2b}] = -\frac{1}{2b(n+1)}$ in that case.