NONPARAMETRIC SMOOTHING APTS

Lab 1 R. J. Samworth/Sep 2011

Comments and corrections to r.samworth@statslab.cam.ac.uk

For reference, if ϕ_σ is the $N(0,\sigma^2)$ density and $\phi = \phi_1$, then

$$\phi_\sigma(x-\mu)\phi_{\sigma'}(x-\mu') = \phi_{\sigma\sigma'/(\sigma^2+\sigma'^2)^{1/2}}(x-\mu^*)\phi_{(\sigma^2+\sigma'^2)^{1/2}}(\mu-\mu'),$$

where $\mu^* = (\sigma'^2\mu + \sigma^2\mu')/(\sigma^2+\sigma'^2)$. Moreover, $\int_{-\infty}^{\infty}\phi(x)^2\,dx = (2\pi^{1/2})^{-1}$.

1. Design a simulation study to compare the performance of the normal scale (`bw.nrd`), least-squares cross-validation (`bw.ucv`), biased cross-validation (`bw.bcv`) and solve-the-equation (`bw.SJ`) bandwidth selectors for kernel density estimation. You should try two underlying densities: (a) $f = \phi$; (b) $f(x) = 0.6\phi(x+1.5) + 0.4\phi(x-1.5)$, sample sizes $n = 100$ and $n = 500$, and a standard Gaussian kernel. Compare empirical estimates of the MISE based on 500 repetitions in each case.

Hint: In these cases, it is possible to compute the ISE analytically for each data set, but if you don't want to do this, you can use the `integrate` *function. If you have time after 2., try also the following cases: (c) f is the $\Gamma(3/2,1)$ density; (d) f is the Cauchy density.*

2. When $K = \phi$ and $f = \phi_\sigma$, an exact expression for the MISE of the kernel density estimator $\hat{f}_h$ is

$$\mathrm{MISE}(\hat{f}_h) = \frac{1}{2\pi^{1/2}}\left\{\frac{1}{nh} + \left(1-\frac{1}{n}\right)\frac{1}{(h^2+\sigma^2)^{1/2}} - \frac{2^{3/2}}{(h^2+2\sigma^2)^{1/2}} + \frac{1}{\sigma}\right\}.$$

Now suppose that $h = h_n$ satisfies $h \to 0$ as $n \to \infty$ and $nh \to \infty$ as $n \to \infty$. Derive an appropriate asymptotic expansion of the MISE above, and deduce that the asymptotically optimal bandwidth with respect to the MISE criterion is given by

$$h_{\mathrm{AMISE}} = \left(\frac{4}{3n}\right)^{1/5}\sigma.$$

Check that the same expression is obtained from the general formula for the asymptotically optimal bandwidth for a second-order kernel.

If you have time, verify the expression for $\mathrm{MISE}(\hat{f}_h)$ given above.