Stat 992: Lecture 03
Gaussian kernel smoothing.

Moo K. Chung `mchung@stat.wisc.edu`

December 26, 2003

1. *Anisotropic kernel.* Let $t = (t_1, \cdots, t_n)' \in \mathbb{R}^n$. The n-dimensional isotropic Gaussian kernel is given by the products of 1-dimensional Gaussian kernel:

$$K(t) = \prod_{i=1}^{n} \frac{1}{\sqrt{2\pi}} e^{-t_i^2/2}.$$

The isotropic kernel under linear transform $t \to Ht$ changes the shape of the kernel to anisotropic kernel

$$K_H(t) = K(H^{-1}t)/\det(H).$$

$\det H^{-1}$ is the Jacobian of the transformation that normalize the density. It is the distribution of n-dimensional multivariate normal with the covariance matrix HH', i.e. $N(0, HH')$. H is also called the *bandwidth matrix* in the context of kernel smoothing and it measures the amount of smoothing. It can be shown that K_H satisfy the definition of the Dirac-delta function as the eigenvalues λ_i of H go to zero, i.e.

$$\lim_{\lambda_i \to 0} K_H(t) = \delta(t).$$

The limit of the sequence of the Gaussian kernels gives the Dirac-delta function and this is how we implement the Dirac-delta function in computer programs. From now on we let $H \to 0$ if all $\lambda_i \to 0$ and $H \to \infty$ if all $\lambda_i \to \infty$.

2. *Kernel smoothing estimator.* Suppose we have a spatial model

$$Y(t) = \mu(t) + \epsilon(t) \tag{1}$$

where e is a mean zero random field and μ is an unknown signal. In image analysis, observations are so dense that we can take them to be continues functional data. Then the kernel smoothing estimator is given by

$$\hat{\mu}(t) = K_H * Y(t) = \int K_H(t-s)Y(s)\,ds \tag{2}$$

As H goes to zero, we are smoothing less. To see this note that

$$\lim_{\sigma \to 0} \hat{\mu}(t) = \int \delta(t-s)Y(s)\,ds = Y(t)$$

From equations (1) and (2),

$$\mathbb{E}\hat{\mu}(t) = K_H * \mu(t) \to \mu(t) \text{ as } \sigma \to 0$$

From the property of the Dirac-delta function as $H \to 0$, the right hand side converges to the true but unknown signal μ. So we can see that our kernel estimator becomes more unbiased as $H \to 0$.

Problem 4. Investigate the property of the estimator when $H \to \infty$. There are a couple of ways to prove what happens when $H \to \infty$. It will be shown later that the estimator is a solution to a heat equation and the condition $H \to \infty$ is equivalent to the steady state when we diffuse heat for infinite time.

Other properties of kernel smoothing estimator can be derived. Assuming $|\mu| \leq \infty$,

$$\mathbb{E}\hat{\mu}(t) \leq \int K_H(t) \sup \mu(t)\,dt \leq \sup \mu(t).$$

Similarly we can bound from below so that

$$\inf \mu(t) \leq \mathbb{E}\hat{\mu}(t) \leq \sup \mu(t). \tag{3}$$

Problem 5. Inequality (3) implies that a smoothed signal will be smaller than the maximum and larger than the minimum of the signal. Can you come up with shaper bounds than (3)? Hint: assume μ has a finite support and work on the neighborhood of t.

Other interesting property is

$$\begin{aligned} \int K_H * Y(t)\,dt &= \int_t \int_s K_H(t-s)Y(s)\,ds\,dt \\ &= \int Y(s)\,ds. \end{aligned}$$

So $\mathbb{E}\int \hat{\mu}(t)\,dt = \int \mu(t)\,dt$. If $\mu(t)$ is a probability such as the transition probability in Brownian motion, it implies that the total probability is conserved even after smoothing.

3. *Covariance of the kernel estimator.* Let R_Y be the covariance function of field Y. It is trivially $R_Y(t,s) = R_\epsilon(t,s) = \mathbb{E}[\epsilon(t)\epsilon(s)]$, the covariance function of ϵ. The covariance function of the kernel smoother is

$$R(t,s) = \int\int K_H(t-t')K_H(t-s')R_Y(t',s')\,dt'\,ds'.$$

Problem 6. Is the variance of signal smaller after smoothing? Come up with interesting inequality involving variances. Note that $\mathbf{Var}\hat{\mu}(t) = R(t,t)$.