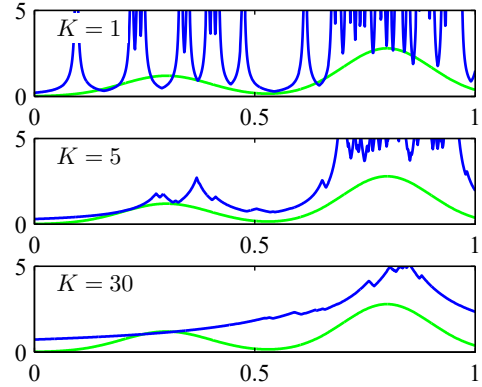


Figure 3.15 Illustration of K -nearest-neighbour density estimation using the same data set as in Figures 3.14 and 3.13. We see that the parameter K governs the degree of smoothing, so that a small value of K leads to a very noisy density model (top panel), whereas a large value (bottom panel) smooths out the bimodal nature of the true distribution (shown by the green curve) from which the data set was generated.



3.5.3 Nearest-neighbours

One of the difficulties with the kernel approach to density estimation is that the parameter h governing the kernel width is fixed for all kernels. In regions of high data density, a large value of h may lead to over-smoothing and a washing out of structure that might otherwise be extracted from the data. However, reducing h may lead to noisy estimates elsewhere in the data space where the density is smaller. Thus, the optimal choice for h may be dependent on the location within the data space. This issue is addressed by nearest-neighbour methods for density estimation.

We therefore return to our general result (3.180) for local density estimation, and instead of fixing V and determining the value of K from the data, we consider a fixed value of K and use the data to find an appropriate value for V . To do this, we consider a small sphere centred on the point \mathbf{x} at which we wish to estimate the density $p(\mathbf{x})$, and we allow the radius of the sphere to grow until it contains precisely K data points. The estimate of the density $p(\mathbf{x})$ is then given by (3.180) with V set to the volume of the resulting sphere. This technique is known as K nearest neighbours and is illustrated in Figure 3.15 for various choices of the parameter K using the same data set as used in Figures 3.13 and 3.14. We see that the value of K now governs the degree of smoothing and that again there is an optimum choice for K that is neither too large nor too small. Note that the model produced by K nearest neighbours is not a true density model because the integral over all space diverges.

We close this chapter by showing how the K -nearest-neighbour technique for density estimation can be extended to the problem of classification. To do this, we apply the K -nearest-neighbour density estimation technique to each class separately and then make use of Bayes' theorem. Let us suppose that we have a data set comprising N_k points in class \mathcal{C}_k with N points in total, so that $\sum_k N_k = N$. If we wish to classify a new point \mathbf{x} , we draw a sphere centred on \mathbf{x} containing precisely K points irrespective of their class. Suppose this sphere has volume V and contains K_k points from class \mathcal{C}_k . Then (3.180) provides an estimate of the density associated

Exercise 3.38