

where V is the volume of \mathcal{R} . Combining (3.178) and (3.179), we obtain our density estimate in the form

$$p(\mathbf{x}) = \frac{K}{NV}. \quad (3.180)$$

Note that the validity of (3.180) depends on two contradictory assumptions, namely that the region \mathcal{R} is sufficiently small that the density is approximately constant over the region and yet sufficiently large (in relation to the value of that density) that the number K of points falling inside the region is sufficient for the binomial distribution to be sharply peaked.

We can exploit the result (3.180) in two different ways. Either we can fix K and determine the value of V from the data, which gives rise to the K -nearest-neighbour technique discussed shortly, or we can fix V and determine K from the data, giving rise to the kernel approach. It can be shown that both the K -nearest-neighbour density estimator and the kernel density estimator converge to the true probability density in the limit $N \rightarrow \infty$ provided that V shrinks with N and that K grows with N , at an appropriate rate (Duda and Hart, 1973).

We begin by discussing the kernel method in detail. To start with we take the region \mathcal{R} to be a small hypercube centred on the point \mathbf{x} at which we wish to determine the probability density. To count the number K of points falling within this region, it is convenient to define the following function:

$$k(\mathbf{u}) = \begin{cases} 1, & |u_i| \leq 1/2, \\ 0, & \text{otherwise,} \end{cases} \quad i = 1, \dots, D, \quad (3.181)$$

which represents a unit cube centred on the origin. The function $k(\mathbf{u})$ is an example of a *kernel function*, and in this context, it is also called a *Parzen window*. From (3.181), the quantity $k((\mathbf{x} - \mathbf{x}_n)/h)$ will be 1 if the data point \mathbf{x}_n lies inside a cube of side h centred on \mathbf{x} , and zero otherwise. The total number of data points lying inside this cube will therefore be

$$K = \sum_{n=1}^N k\left(\frac{\mathbf{x} - \mathbf{x}_n}{h}\right). \quad (3.182)$$

Substituting this expression into (3.180) then gives the following result for the estimated density at \mathbf{x} :

$$p(\mathbf{x}) = \frac{1}{N} \sum_{n=1}^N \frac{1}{h^D} k\left(\frac{\mathbf{x} - \mathbf{x}_n}{h}\right) \quad (3.183)$$

where we have used $V = h^D$ for the volume of a hypercube of side h in D dimensions. Using the symmetry of the function $k(\mathbf{u})$, we can now reinterpret this equation, not as a single cube centred on \mathbf{x} but as the sum over N cubes centred on the N data points \mathbf{x}_n .

As it stands, the kernel density estimator (3.183) will suffer from one of the same problems that the histogram method suffered from, namely the presence of artificial discontinuities, in this case at the boundaries of the cubes. We can obtain a smoother