

requirements
large number
feature space
dimensionality
nonparametric

4.4 $k_n \cdot N$

One of the important requirements is the choice of the window width V_n . We can take $V_n = 1$, which is a reasonable choice for the bimodal density. If V_n is too small, the estimate will be empty, as the density is zero over the cell. If V_n is too large, the estimate will be inappropriate.

One potential problem is the choice of the function of k_n . For a fixed n , we can choose a window width V_n such that the cell about x contains approximately n samples. For a fixed V_n , we can choose k_n such that the specified function of k_n is approximately constant. If the density is smooth, this will give a good resolution. But if the density is not smooth, it will result in a poor resolution. In this case, we take $k_n = n^{1/3}$.

We want $k_n \cdot N$ to be large. As n increases, k_n/n will be smaller. This means that the volume of the cell containing x will increase. The size of the cell will also increase. It is clear from the figure that as n increases, the estimate becomes more accurate. We can supply a probability statement that $p_n(x)$ converges to $p(x)$ as $n \rightarrow \infty$. This is because $\lim_{n \rightarrow \infty} k_n/n = 0$. The probability that $p_n(x) > p(x) + \epsilon$ is zero. This is because $p_n(x) \approx p(x)$ for large n . The volume V_n is arbitrary and does not affect the estimate.

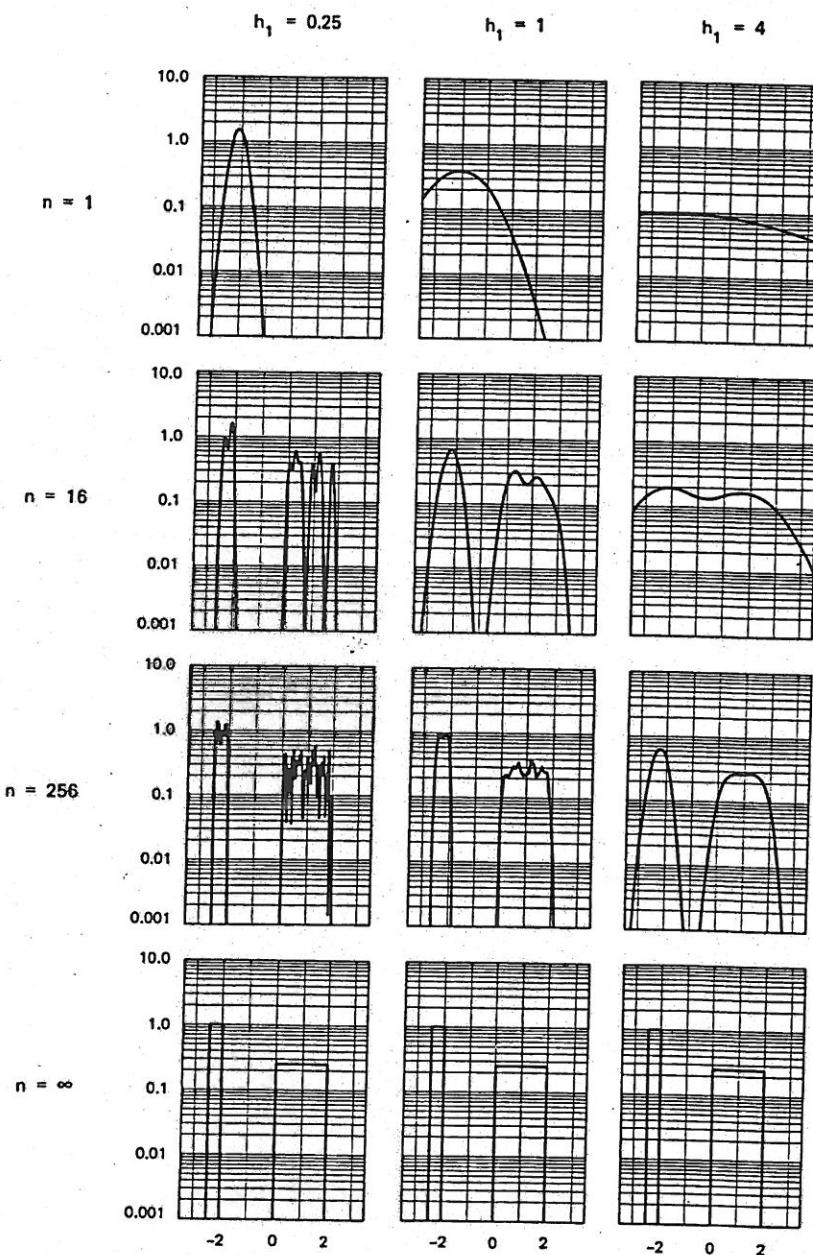


FIGURE 4.2. Parzen-window estimates of a bimodal density.