5.3 Kernel Estimation

We may generalize the idea of the histogram by replacing the indicator function I with something else:

$$\hat{f}(x) = \frac{1}{N} \sum_{n=1}^{N} k(x - x_n; w) \,, \tag{5.4}$$

where $k(x, w)$ is the *kernel function*, normalized to unity:

$$\int_{-\infty}^{\infty} k(x; w) dx = 1 \,. \tag{5.5}$$

This provides a means to avoid some of the drawbacks of histograms, for example, we can obtain a continuous density estimate. We are usually interested in kernels of the form

$$k(x - x_n; w) = \frac{1}{w} K\left(\frac{x - x_n}{w}\right) \,. \tag{5.6}$$

The kernel estimator for the pdf is then

$$\hat{f}(x) = \frac{1}{nw} \sum_{n=1}^{N} K\left(\frac{x - x_n}{w}\right) \,, \tag{5.7}$$

The role of parameter w as a "smoothing" parameter is apparent. The delta functions of the empirical distribution are spread over regions of order w.

Often, the particular form of the kernel used doesn't matter very much. This is illustrated with a comparison of several kernels (with commensurate smoothing parameters) in Figure 5.4. The Gaussian is probably the most popular, and is smooth. Optimization criteria and error estimation that may be applied to kernel methods are discussed in Sections 5.6 and 5.7.

5.3.1 Multivariate Kernel Estimation

Kernel estimation may easily be applied to multivariate situations. For example, in $D = 2$ dimensions:

$$\hat{f}(x, y) = \frac{1}{N w_x w_y} \sum_{n=1}^{N} K\left(\frac{x - x_n}{w_x}\right) K\left(\frac{y - y_n}{w_y}\right) \,. \tag{5.8}$$

This is a "product kernel" form, with the same kernel in each dimension, except for possibly different smoothing parameters. It does not have correlations. The kernels we have introduced are classified more explicitly as "fixed kernels": the smoothing parameters are independent of x and y.

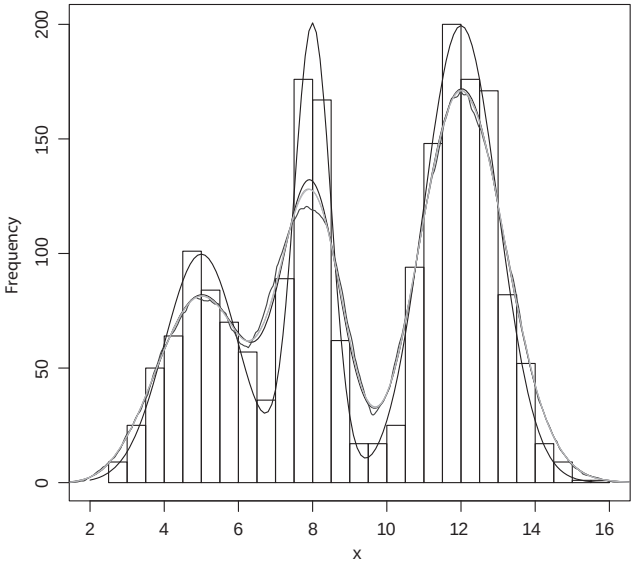

Figure 5.4 Comparison of density estimates using different kernels. The curve with the largest excursions is the sampling distribution, the next largest is the estimate with Gaussian kernel, followed by indistinguishable triangular and cosine kernel estimates, and finally a rectangular kernel estimate.

5.4 Ideogram

A simple variant on the kernel idea is to permit the kernel to depend on additional knowledge in the data. Physicists call this an *ideogram*. Most common is the *Gaussian ideogram*, in which each data point is entered as a Gaussian of area one and standard deviation appropriate to that datum. This addresses a way that the i.i.d. assumption might be broken.

The Particle Data Group uses ideograms as a means to convey information about possibly inconsistent measurements. Figure 5.5 shows an example of this.

5.5 Parametric vs. Nonparametric Density Estimation

The distinction between parametric and nonparametric is somewhat murky. A histogram is nonparametric, in the sense that no assumption about the form of the sampling distribution is made. Often an implicit assumption is made that the distribution is "smooth" on a scale smaller than the bin size. For example, we might know something about the resolution of our apparatus and adjust the bin size to be commensurate. But the estimator of the parent distribution made with a histogram is parametric – the parameters are populations (or frequencies) in each bin. The es-