Chapter 5

Nonparametric regression and distribution estimation

Nonparametric methods in statistics refer to analysis methods which try to avoid assuming certain parametric distribution in the model. Usually, the assumption to be avoided is the normal distribution. As contrary to the name nonparametric (*epäparametrinen*), these methods usually have a large number of parameters.

Nonparametric methods are used in all the fields in data-analysis, for example there is a variety of nonparametric tests available. However, here we mention only two nonparametric methods — spline regression and kernel density estimation.

5.1 Spline regression and other smoothing techniques

Sometimes the functional form or dependence between explanatory variable(s) and dependent variable is not interesting in such, only some kind of smooth description of the behavior. In these cases either direct smoothing of the data or regression smoothing is searched for.

There are many different data smoothing techniques, from which moving average or moving median are the most simple ones. In these, the values of y_i are replaced by average (or median) over a smoothing window that holds k observations around the i'th observation. An example of such smoothings are shown in Fig. 5.1 with window size of 10. Other, more advanced methods include e.g. LOESS or LOWESS smoothing.

One more interesting smoothing or nonparametric regression technique is the spline regression. This method should not be mixed with spline interpolation where all the variability of the observations is reproduced. In spline regression, a small number of so-called cubic B-splines that are local third-order polynomials are used as a

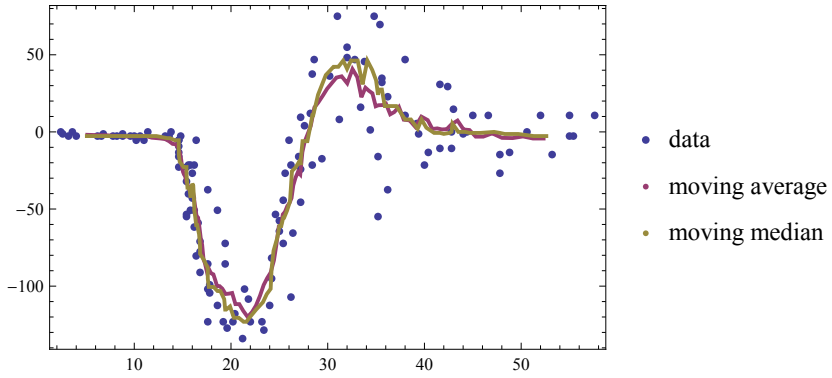

Figure 5.1: Moving average and moving median smoothing to the data.

basis for linear regression. When the spline basis $\mathrm{B}_j(x)$ is formed, the sum of these, $\sum_j \beta_j \mathrm{B}_j(x)$ is fitted to the data in least-square sense.

The spline basis functions are distributed to the range of explanatory variables x_i evenly, or preferably to the quantiles of the data. We will not go into details with B-spline basis derivation, there are suitable material in e.g. Wikipedia or in Numerical Recipies. A spline regression for the data in previous moving average/median example is shown in Fig. 5.2, together with the cubic spline basis that is distributed along x to 7 quantiles of the data plus the end-points, 0%, 12.5%, 25%, 37.5%, 50%, 62.5%, 75%, 87.5%, 100%.

For technical reasons, the spline basis if formed with knots where the end-points are repeated four times in the knot list, so with k quantiles there are $k + 2 \times 4$ knots in the basis. With those knots, total of $k + 4$ splines are available.

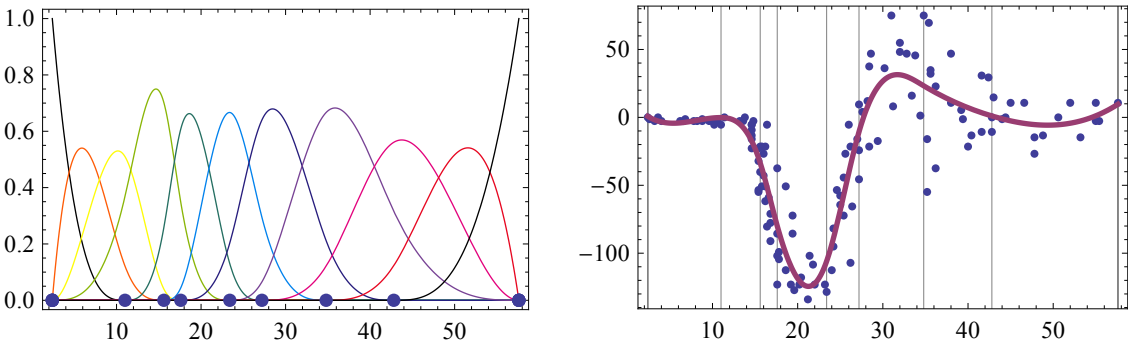

Figure 5.2: Spline basis for 7 quantiles and end-points of the data (left) and fitted regression spline of the basis functions (right).

5.2 Kernel estimation

Kernel estimation (*ydinestimointi*) is a nonparametric method for estimating (continuous) distribution (pdf) of the data. The methods works for both one-dimensional or multidimensional data. The result of kernel estimation is not a parametrized close-formed distribution, but a numerical function that can be used to compute values of the distribution estimate.

The idea of kernel estimation is quite simple. Every observation x_i in the data is replaced by a kernel function $\mathrm{K}_i(x; x_i, h)$, and the total kernel estimate is the scaled sum of kernels:

$$\mathrm{K}(x; \boldsymbol{x}, h) = \frac{1}{n} \sum_{i}^{n} \mathrm{K}_i(x; x_i, h), \tag{5.1}$$

where $\boldsymbol{x}$ is the data vector, x the value where the distribution is evaluated, and h is the smoothing parameter (*siloitusparametri*).

The choice of the kernel function should not be too critical, any non-negative function that is symmetric around its maximum and integrates to one should do. One suitable choice is to use the pdf of normal distribution, with expected value $\mu = x_i$ and variance $\sigma^2 = h^2$. So, kernel is

$$\mathrm{K}_i(x; x_i, h) = \frac{1}{\sqrt{2\pi}h} \exp\left(-\frac{(x - x_i)^2}{2h^2}\right). \tag{5.2}$$

More important than the actual shape of the kernel should be the choice of the smoothing parameter h. There are different advices, one of such is the method of Silverman:

$$h = s \left(\frac{4}{p+2}\right)^{\frac{1}{p+4}} n^{-\frac{1}{p+4}}, \tag{5.3}$$

where p is the dimension of the data. With one-dimensional case the s is simply the standard deviation of the data. An example of kernel estimation of the density function for three observations is shown in Fig. 5.3.

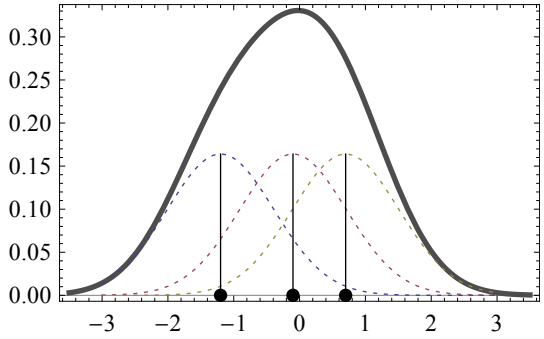

Figure 5.3: Three observations, normal pdf kernels and the kernel density estimate of the pdf.

Kernel estimation suits quite well for multidimensional cases, too. For these, a multidimensional normal distribution pdf can be used as the kernel with covariance matrix $h^2\mathbf{I}_p$ or even with $h^2\mathbf{C}$ where $\mathbf{C}$ is the correlation matrix estimated from the data. For smoothing parameter h the s in Eq. (5.3) should be computed from the diagonal elements of the covariance matrix $\mathbf{S}$ of the data:

$$s = \sqrt{\frac{1}{p}\sum_{i}^{p} S_{ii}}. \tag{5.4}$$

Example for two-dimensional kernel estimate is shown in Fig. 5.4.

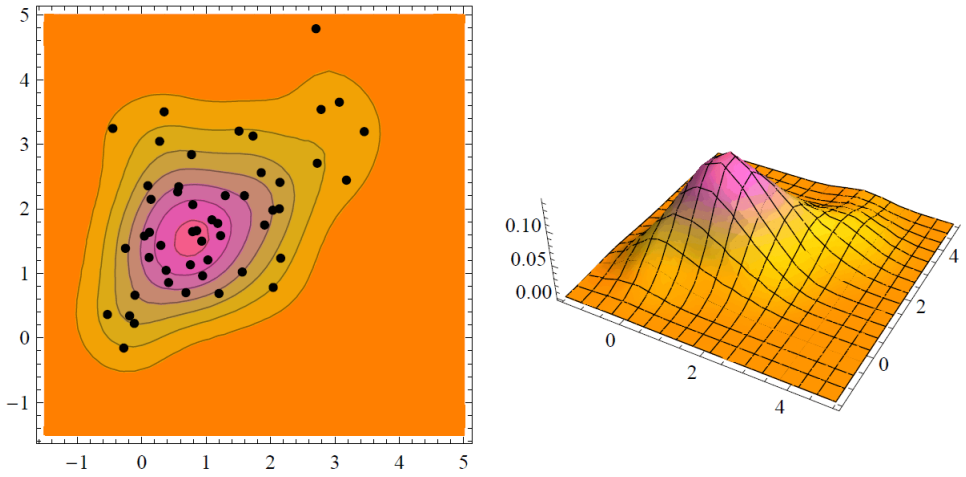

Figure 5.4: Two-dimensional observations and kernel estimate for the pdf. On left, a contour plot of the estimate with the data, on right, 3-D surface plot of the kernel estimate.