Shape-Adaptive Kernel Density Estimation

L.E.N. Baakman

November 14, 2016

1 Introduction

Density estimation tries to find the density $f(\mathbf{x})$ in d-dimensional Euclidean space underlying N points $\mathbf{x}_1 \dots \mathbf{x}_N$, that have been selected independently from $f(\mathbf{x})$. One often used method to solve this problem is the Parzen approach [4], which gives the following estimate of the density function:

$$\hat{f}(\mathbf{x}) = \frac{1}{N}\sum_{j=1}^{N}\frac{1}{\sigma^d}K\left(\frac{\mathbf{x}-\mathbf{x}_j}{\sigma}\right). \tag{1}$$

Thus the estimated density is the mean of bumps placed at each observation. The shape of these bumps is determined by the shape of the kernel function $K(\cdot)$, their width is controlled by the bandwidth σ [5]. The Parzen approach requires that $K(\mathbf{x}) \geq 0$, with

$$\int K(\mathbf{x}) = 1, \tag{2}$$

it is often the case that the kernel statisfies some additional conditions.

One downside of the Parzen method is that it cannot respond appropriately to variations in the magnitude of the density function, i.e. the peakedness of the kernel is not data-responsive. Consequently in regions of low $f(\mathbf{x})$ that contain only one sample point, the estimate will have a peak at $\mathbf{x}$ and be too low over the rest of the region. In areas where the density is high, the sample points are more densely packed together, and the Parzen estimate will tend to spread out the high density region [1]. Breiman, Meisel, and Purcell introduced an variant of the Parzen estimator which makes the sharpness of the kernel responsive to the local data:

$$\hat{f}(\mathbf{x}) = \frac{1}{N}\sum_{j=1}^{N}(\alpha d_{j,k})^{-d}K_{\mathscr{G}}\left(\frac{\mathbf{x}-\mathbf{x}_j}{\alpha d_{j,k}}\right), \tag{3}$$

where $K_{\mathscr{G}}(\cdot)$ represents a Gausian kernel, α is a multiplicative constant and $d_{j,k}$ the distance between $\mathbf{x}_j$ and the k nearest neighbour of $\mathbf{x}_j$. Comparing Equation (1) with (3) we find that the bandwidth σ, has been replaced with $\alpha d_{j,k}$. In low density regions $d_{j,k}$ will be large, and the kernel will be spread out, in high density regions the converse occurs. Breiman, Meisel, and Purcell use a minimization algorithm on a goodness of fit statistic to find suitable values for k and α.

One disadvantage of the Breiman estimator is that it is computationally expensive, both due to its minimization procedure and the infinite support kernel. Wilkinson and Meijer proposed the modified Breiman estimator:

$$\hat{f}(\mathbf{x}) = \frac{1}{N}\sum_{j=1}^{N}(\sigma\lambda_j)^{-d}K_{\mathscr{E}}\left(\frac{\mathbf{x}-\mathbf{x}_j}{\sigma\lambda_j}\right). \tag{4}$$

This estimator differs on two points from the Breiman estimator. Firstly contrary to Breiman, Meisel, and Purcell who use a Gaussian kernel, Ferdosi et al. use an Epanechnikov kernel [2]. Secondly they replaced the expensive minimization procudure with a pilot density estimate:

$$\hat{f}_{\text{pilot}}(\mathbf{x}) = \frac{1}{N}\sum_{j=1}^{N}\sigma^{-d}K\left(\frac{\mathbf{x}-\mathbf{x}_j}{\sigma}\right). \tag{5}$$

By computing the pilot density first on grid points and using multi-linear interpolation to determine the pilot density for each point the computational complexity is further decreased. Based on these pilot densities the local bandwidths are computed:

$$\lambda_i = \left(\frac{\hat{f}_{\text{pilot}}(\mathbf{x}_i)}{g}\right)^{-\beta}. \tag{6}$$

Here g is the geometric mean of the pilot densities and β is the sensitivity parameter.

Although the widths of the kernels used in both the Breiman estimator and the modified Breiman estimator respond to the data, the shapes of the kernels are dependent of the kernel not the data. To further increase the response of the estimator to the data we propose shape-adaptive kernels, kernels of which both the width and the shape are steered by the data.

A disadvantage of these shape-adaptive kernels is that in regions where the density of sample points

is low there are not enough data points to compute the shape of the kernel reliably. Consequently we propose to let the amount in which the shape of the kernel is influenced by the local data depend on their density.

This paper is organized as follows. Section 2 discusses the proposed shape-adaptive kernels and somee implementation details.

Aanvullen

2 Method

Contrary to Ferdosi et al. we will use a Gaussian kernels, since those can be simply reshaped by using a different covariance matrix. However we do use the method introduced by Ferdosi et al. to estimate the window width. In short the procedure works as follows. First a window size is computed for each dimension of the data:

$$\sigma_l = \frac{P_{80}(l) - P_{20}(l)}{\log N}, l = 1, \ldots, d \tag{7}$$

where $P_x(l)$ denotes the xth percentile of the data point in dimension l. To avoid over smoothing the final pilot window σ is chosen as

$$\sigma = \min\{\sigma_1, \ldots, \sigma_d\}. \tag{8}$$

In the literature there is some discussion on the value of the sensitivity parameter β. Breiman, Meisel, and Purcell propose a value of $1/d$, whereas Silverman prefers a value of $1/2$ regardless of the dimensionality of the data. We have emperically determined that .

Empirisch vaststellen

2.1 Kernel Shape

We use the k-nearest neighours algorithm (k-NN) to find $C_{k,\mathbf{x}}$, the set with k neighbours of data point $\mathbf{x}$.

k-NN ensures, assuming a reasonable value of k, that we have enough points to base the shape of the kernel on. Note that we include the pattern itself in the set of its k neighbours. We have followed Silverman's recommendation of choosing $k = \sqrt{N}$.

The basic shape of the kernel used for $\mathbf{x}$ is determined by the covariance matrix Σ of $C_{k,\mathbf{x}}$, which results in a kernel which has the same shape as the data points in $C_{k,\mathbf{x}}$. However kernels computed in this way differ in area from each other, to ensure that the density estimation for each point takes an equal area of data points into account we have to scale the covariance matrix. To do this we compute a scaling factor for the covariance matrix as

$$S = \frac{\sigma^2}{\sqrt[d]{\prod_{l=1}^{d} \sqrt{\lambda_l}}} \tag{9}$$

Anders formuleren, in feite is de Gaussian oneindig.

where λ_l for $l = 1, \ldots, d$ represent the eigenvalues of the covariance matrix of $C_{k,\mathbf{x}}$. Note that since we compute the eigenvalues of this matrix it may not be singular and k should thus be greater than $d+1$. The scaling factor S ensures that the area of all kernels at a fixed isoline are equal. We have taken the area at that isoline of the kernel with as covariance matrix the matrix with σ on the diagonal as our guideline.

2.2 Steering the Kernel Shape

References

[1] L. Breiman, W. Meisel, and E. Purcell. "Variable Kernel Estimates of Multivariate Densities". In: *Technometrics* 19.2 (1977), pp. 135–144.

[2] V.A. Epanechnikov. "Non-Parametric Estimation of a Multivariate Probability Density". In: *Theory of Probability & Its Applications* 14.1 (1969), pp. 153–158.

[3] B.J. Ferdosi et al. "Comparison of Density Estimation Methods for Astronomical Datasets". In: *Astronomy & Astrophysics* 531 (2011).

[4] E. Parzen. "On Estimation of a Probability Density Function and Mode". In: *The Annals of Mathematical Statistics* 33.3 (1962), pp. 1065–1076.

[5] B.W. Silverman. *Density Estimation for Statistics and Data Analysis*. Monographs on Statistics and Applied Probablity. Springer-Science+Business Media, B.V., 1986.

[6] M.H.F. Wilkinson and B.C. Meijer. "DATAPLOT: A Graphical Display Package for Bacterial Morphometry and Fluorimetry Data". In: *Computer Methods and Programs in Biomedicine* 47.1 (1995), pp. 35–49.