The Fast RODEO for Local Polynomial Regression - Supplemental Material

Daniel V. Samarov *

September 20, 2013

1 Supplemental Material

This document provides additional discussion on the fast RODEO. Section 2 gives an overview of methods for estimating σ in the fast RODEO and Section 3 provides a more detailed discussion of the relationship between scale-space and the RODEO.

2 Estimating σ

In Lafferty & Wasserman (2008) several approaches to estimating the standard error σ are given. Their discussion is included here for completeness. Note, in our implementation the estimation of σ is performed on *unbinned* data. This first approach is based on a generalization of Rice (1984). First, let

$$d_{il} = ||X_i - X_l||, \text{ for } i < l. \tag{1}$$

*Daniel V. Samarov is a Mathematical Statistician, Statistical Engineering Division, Information Technology Laboratory, National Institute of Standards and Technology, Gaithersburg, MD 20899, (Email: daniel.samarov@nist.gov).

For a fixed integer J , let $\mathcal{E}$ denote the set of pairs (i, l) corresponding to the J smallest values of d_{il} . With this we estimate σ as

$$\hat{\sigma}^2 = \frac{1}{2J} \sum_{i,l \in \mathcal{E}} (Y_i - Y_l)^2. \quad (2)$$

An alternative estimate which is potentially more robust can be computed as

$$\hat{\sigma} = \frac{\sqrt{\pi}}{2} \text{median}\{|Y_i - Y_l|\}_{i,l \in \mathcal{E}}. \quad (3)$$

Additional details and dicussion can be found in Section 4.5 of Lafferty & Wasserman (2008).

A novel scale-space view of σ is provided at the end of Section 3.

3 A Scale-Space view of the RODEO

The *scale-space* (Lindeberg (1994)) model of nonparametric regression and density estimation (Chaudhuri & Marron (2000), Godtliebsen et al. (2002), also see Holmström (2010) for a comprehensive review) is that there is not necessarily one optimal bandwidth, but rather many relevant bandwidths which reveal useful features about the data. The term *scale* is a reflection of this idea, i.e. that large-scale/coarse features in the data are best captured by a “zoomed out” view (associated with a larger bandwidth), and that small-scale/finer features are best captured by a “zoomed in” view (associated with a smaller bandwidth).

In the univariate case we can visualize the scale-space by fitting a sequence of smoothers (here LPR), to a given set of data across a range of bandwidths, from large to small. Placing these smooths, in order of bandwidth on the same plot forms the “scale-space surface”, an illustration of which is shown in Figure 1. Here the points are the observed data, and the surface is the collection of smooths over a range of bandwidths (shown on the $\log_2$ scale along the z-axis). The larger bandwidths capture the general trend of the data, and moving to smaller bandwidths reveals finer features.

Figure 1: An illustration of a scale surface. The points correspond to the observed data, and the surface the fitted smooths for each bandwidth (shown on the $\log_2$ scale along the z-axis).

Next we look at how a scale-space interpretation of the RODEO can provide us with useful insights on how the algorithm works. This will also help build intuition for the binned implementation of the RODEO which we discuss in Section 3. We start by taking a closer look at a region with finer, i.e. “non-flat” features shown in Figure 2. Consider the point falling closer to the edge of the surface (furthest to the right) whose coordinates are $(x^*, \hat{m}_{h_R}(x^*), h_R)$. Here $\hat{m}$ is the LPR estimate and h_R is the corresponding bandwidth. The RODEO looks to test whether updating our estimate from the this point to the one above it (with estimator $\hat{m}_{h_G}(x^*)$, $h_G < h_R$) produces a significant change in $\hat{m}$. The magnitude of this change is captured by the slope of the line connecting the two points shown on the surface, i.e.

$$\frac{\hat{m}_{h_R} - \hat{m}_{h_G}}{h_R - h_G}.$$

As we take the limit, $h_R \rightarrow h_G$ (i.e. taking smaller and smaller steps along the bandwidth axis) this gives us the derivative of $\hat{m}_h$ with respect to h evaluated at h_G , which is exactly the test statistic Z in (??).

Following the procedures discussed in Section 2, for a given point, we move across the scale-space surface along the bandwidth axis (going from larger to smaller bandwidths) until the test statistic Z is no longer significant. This process is repeated for each x^* . Figure 3 shows the final RODEO solution path for the example in Figure 1. As can be seen algorithm tends to stop at larger bandwidths in regions with larger features and at smaller bandwidth where there are finer features, as expected.

As we move into higher dimensions it becomes more challenging to visualize such a surface, however, the principles remain the same.

Remark 3.1. While not discussed in detail here, the scale-space view of the RODEO algorithm provides a way to extend and generalize the scale-space view of smoothing. The concept is based the fact that in the RODEO algorithm the standard deviation, σ controls the amount of smoothing performed. This is straightforward to see by noting that the test parameter λ_j is a function of σ . Thus for smaller values of the latter there is in an increased probability

Figure 2: Zooming in on a region of the scale space surface shown in Figure 1. The point falling nearer the edge of the surface has coordinates $(x^*, m_{h_R}(x^*), h_R)$ and the point above it $(x^*, m_{h_G}(x^*), h_G)$ with $h_R > h_G$.

Figure 3: The line along the scale-space surface shows the final RODEO solution path.

that $|Z_j| > \lambda_j$, and correspondingly the value of the bandwidth h_j will be also tend to be smaller, the reverse holds true for larger values of σ .

By considering a range of values for σ the result is quite similar to the standard scale-space view, the difference is that now our scale adapts locally as opposed to globally. We refer to this as *adaptive scale-space* model and we will explore this in more detail in future work.

References

- CHAUDHURI, P. & MARRON, J. (2000). Scale space view of curve estimation. *Annals of Statistics* 28 408–428.
- GODTLIEBSEN, F., MARRON, J. & CHAUDHURI, P. (2002). Significance in scale space for bivariate density estimation. *Journal of Computational and Graphical Statistics* 11 1–21.

HOLMSTRÖM, L. (2010). Scale space methods. *WIREs Comp Stat* 150–159.

LAFFERTY, J. & WASSERMAN, L. (2008). Rodeo: Sparse, greedy nonparametric regression. *Annals of Statistics* 36 28–63.

LINDBERG, T. (1994). Scale-space theory in computer vision.

RICE, J. (1984). Bandwidth choice for nonparametric regression. *Annals of Statistics* 12 1215–1230.