STAT 704 Sections 11.4-11.5. IRLS and Bootstrap

John Grego

Department of Statistics, University of South Carolina

Stat 704: Data Analysis I

LOWESS

LOWESS (LOcally WEighted Scatterplot Smoothing) is a highly prescriptive scatterplot smoothing method developed by Cleveland, 1979. Other scatterplot smoothing methods provide more flexibility in weighting functions, smoothing criteria, etc.

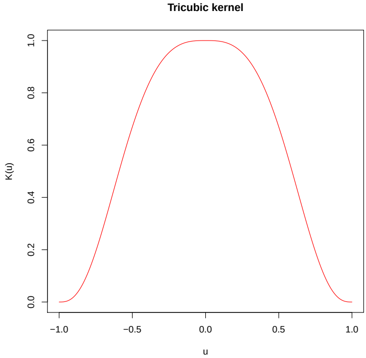
Features of LOWESS include:

- local regression
- weighted regression
- robustness to outlying observations.

We will study LOWESS assuming a single predictor variable.

The first step in LOWESS is a locally weighted regression with the weight function based on the tricubic kernel:

$$K(u) = \begin{cases} \left\{1 - (|u|)^3\right\}^3 & |u| \leq 1 \\ 0 & |u| > 1 \end{cases}$$



At each x_{i^*}, $i^* = 1, \ldots, n$, construct a local linear or quadratic regression based on $K(\cdot)$ using weights

$$w_k(x_{i^*}) = K\left(\frac{|x_k - x_{i*}|}{\Delta_q}\right), \; k = 1, \ldots, n$$

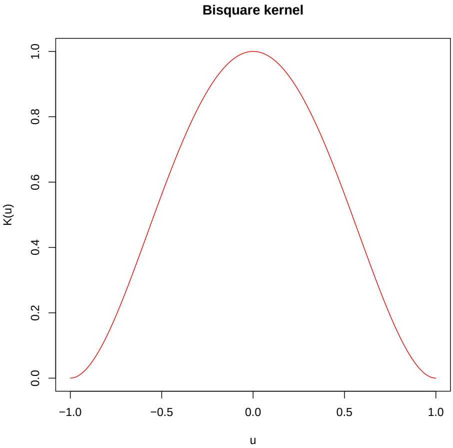
where Δ_q is the q^{th} order statistic of $\{|x_k - x_{i*}|\}_k$.

For each weighted least squares, we focus on the predicted value at x_{i^*}: $\hat{y}_{i^*}$–rather than the weighted least squares line itself.

Residuals $\{e_{i*}\}$ are calculated, and the next weighted least squares regression includes robust adjustments for outliers.

Define the bisquare kernel:

$$B(u) = \begin{cases} \left\{1 - u^2\right\}^2 & |u| \leq 1 \\ 0 & |u| > 1 \end{cases}$$



Set $s = \text{med}\,|e_k|$ and define robustness weights:

$$\delta_k = B\left(\frac{e_k}{6s}\right)$$

Use weights $\delta_k \times w_k(i^*)$ in a second series of local weighted least squares regressions.

Repeat the steps until the process converges.

Example

From the water quality data, we will study fecal coliform levels for Station C-076 (Cedar Creek).

- Plots show the presence of outliers, even after a scale transformation, as well as some local behavior that suggests a need for robust scatterplot smoothing.
- PROC LOESS in SAS conducts local regression by default, with robust iterative weighting of outliers introduced by the ITERATIONS= option.
- PROC LOESS uses methods similar to LOWESS, though with many more options for smoothing criteria available.

Section 11.6

The text provides an introduction to the bootstrap without much context. We will adopt a similar approach; details are more suitable for, e.g., STAT 740. Intuitively, the empirical distribution function (below) can be used as an estimate of the distribution function F of the independent identically distributed error terms ϵ_i.

$$F_n(x) = \frac{1}{n} \sum_{i=1}^{n} I(x_n \leq x)$$

As such, sampling from F_n can be used to model sampling from F.

- Sampling from F_n should be *with* replacement to mimic repeated sampling from F.
- Functionals in F have their counterparts in F_n. E.g., $\mu(F_n) = \bar{x}$.
- These analogies lead to methods for deriving sampling distributions, and hence ready-made estimates of standard errors and confidence bounds, in the absence of closed-form results.
- Simpler will not always prove better!

The text distinguishes two types of bootstrap regressions

- When the predictor variables are fixed and errors have constant variance, bootstrap $\{e_i\}$.
- When the predictor variables are random, bootstrap $\{(\mathbf{x}_i, y_i)\}$

For the former case, the bootstrap sample will be $\{e_i^*\}$. We compute $Y_i^* = \hat{Y}_i + e_i^*$, $i = 1, \ldots, n$, then regress $\{Y_i^*\}$ on $\{\mathbf{X}_i\}$, typically to obtain B bootstrap slope estimates, $b_{1(b)}^*$, $b = 1, \ldots, B$.

We can then compute $s^*\{b_1^*\}$, the standard deviation of $\left\{b_{1(b)}^*\right\}$, as an estimate of the standard error of b_1.

We can also use the bootstrap sample to compute confidence intervals; the number of bootstrap samples, B, tends to be large for this, particularly for empirical methods.

There are numerous approaches to bootstrap confidence intervals; the book introduces one of the most interesting, the *reflection method*. The *percentile method* simply uses the $\alpha/2$ and $1-\alpha/2$ sample percentiles from $\left\{b^*_{1(b)}\right\}$ to construct a $100(1-\alpha)\%$ CI for β_1.

The reflection method

The reflection method computes

$$\begin{aligned} d_1 &= b_1 - b_1^*(\alpha/2) \\ d_2 &= b_1^*(1-\alpha/2) - b_1 \end{aligned}$$

The $100(1-\alpha)\%$ CI is then $(b_1 - d_2, b_1 + d_1)$. Why does this work?

The reflection method

With probability $1 - \alpha$, b_1 will fall between the percentiles of its sampling distribution:

$$P\left[b_1(\alpha/2) \leq b_1 \leq b_1(1-\alpha/2)\right] = 1 - \alpha$$

The distances between these percentiles and β_1, the mean of the sampling distribution of b_1, are:

$$\begin{aligned} D_1 &= \beta_1 - b_1(\alpha/2) \\ D_2 &= b_1(1-\alpha/2) - \beta_1 \end{aligned}$$

Rearranging, we have

$$\begin{aligned} b_1(\alpha/2) &= \beta_1 - D_1 \\ b_1(1-\alpha/2) &= \beta_1 + D_2 \end{aligned}$$

Example

The **cars** data set in R studies stopping distance of cars as a function of speed. The data is not quite linear, and the variation in stopping distance increases with speed, but we will set aside those issues for now.

R has numerous libraries (**boot** is popular, though it has its peculiarities) to bootstrap models. We can use a hand-constructed function to bootstrap residuals from the regression of `dist` on `speed`. We will want to compare the percentile and reflection bootstrap confidence intervals to the 95% confidence interval obtained from the regular normal errors model: (3.097, 4.768).