Nonparametric Regression and Bonferroni joint confidence intervals

Frank Wood

February 25, 2010

Nonparametric Regression Curves

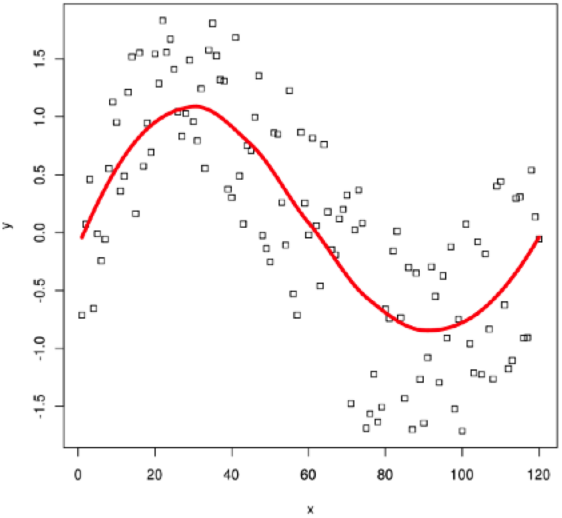
- So far: parametric regression approaches
  - Linear
  - Linear with transformed inputs and outputs
  - etc.
- Other approaches
  - Method of moving averages : interpolate between mean outputs at adjacent inputs
  - Lowess : "locally weighted scatterplot smoothing"

Lowess Method

- Intuition
  - Fit low-order polynomial (linear) regression models to points in a neighborhood
    - The neighborhood size is a parameter Determining the neighborhood is done via a nearest neighbors algorithm

    Produce predictions by weighting the regressors by how far the set of points used to produce the regressor is from the input point for which a prediction is wanted
- While somewhat ad-hoc, it is a method of producing a nonlinear regression function for data that might seem otherwise difficult to regress

Lowess Method Example



Bonferroni Joint Confidence Intervals

- Calculation of Bonferroni joint confidence intervals is a general technique
- We highlight its application in the regression setting
  - Joint confidence intervals for β_0 and β_1
- Intuition
  - Set each statement confidence level to greater than $1 - \alpha$ so that the family coefficient is at least $1 - \alpha$

Ordinary Confidence Intervals

- Start with ordinary confidence intervals for β_0 and β_1

$$b_0 \pm t(1-\alpha/2; n-2)s\{b_0\}$$
$$b_1 \pm t(1-\alpha/2; n-2)s\{b_1\}$$

- And ask what the probability that one or both of these intervals is incorrect

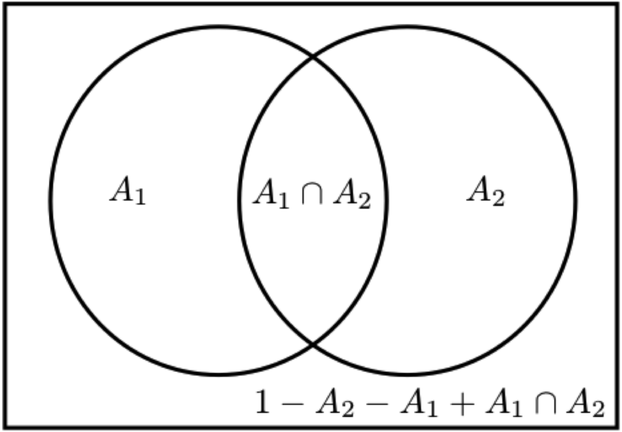
Remember

$$\begin{aligned} s^2\{b_0\} &= MSE\left[\frac{1}{n} + \frac{\bar{X}^2}{\sum(X_i - \bar{X})^2}\right] \\ s^2\{b_1\} &= \frac{MSE}{\sum(X_i - \bar{X})^2} \end{aligned}$$

General Procedure

- Let A_1 denote the event that the first confidence interval does not cover β_0, i.e. $P(A_1) = \alpha$
- Let A_2 denote the event that the second confidence interval does not cover β_1, i.e. $P(A_2) = \alpha$
- We want to know the probability that both estimates fall in their respective confidence intervals, i.e. $P(\bar{A}_1 \cap \bar{A}_2)$
- How do we get there from what we know?

Venn Diagram



Bonferroni inequality

- We can see that
  $P(\bar{A}_1 \cap \bar{A}_2) = 1 - P(A_2) - P(A_1) + P(A_1 \cap A_2)$
  - Size of set is equal to area is equal to probability in a Venn diagram.
- It also is clear that $P(A_1 \cap A_2) \geq 0$
- So, $P(\bar{A}_1 \cap \bar{A}_2) \geq 1 - P(A_2) - P(A_1)$ which is the Bonferroni inequality.
- In words, in our example
  - $P(A_1) = \alpha$ is the probability that β_0 is *not* in A_1
  - $P(A_2) = \alpha$ is the probability that β_1 is *not* in A_2
  - $P(\bar{A}_1 \cap \bar{A}_2)$ is the probability that β_0 is in A_1 *and* β_1 is in A_2
  - So $P(\bar{A}_1 \cap \bar{A}_2) \geq 1 - 2\alpha$

Using the Bonferroni inequality

- Forward (less interesting) :
  - If we know that β_0 and β_1 are lie within intervals with 95% confidence, the Bonferroni inequality guarantees us a family confidence coefficient (i.e. the probability that *both* random variables lie within their intervals simultaneously) of at least 90% (if both intervals are correct). This is

$$P(\bar{A}_1 \cap \bar{A}_2) \geq 1 - 2\alpha$$

- Backward (more useful):
  - If we know what to *specify* a family confidence interval of 90%, the Bonferroni procedure instructs us how to adjust the value of α for each interval to achieve the overall family confidence desired

Using the Bonferroni inequality cont.

- To achieve a $1-\alpha$ *family* confidence interval for β_0 and β_1 (for example) using the Bonferroni procedure we know that both individual intervals most shrink.
- Returning to our confidence intervals for β_0 and β_1 from before

$$b_0 \pm t(1-\alpha/2; n-2)s\{b_0\}$$
$$b_1 \pm t(1-\alpha/2; n-2)s\{b_1\}$$

- To achieve a $1-\alpha$ *family* confidence interval these intervals must *widen* to

$$b_0 \pm t(1-\alpha/4; n-2)s\{b_0\}$$
$$b_1 \pm t(1-\alpha/4; n-2)s\{b_1\}$$

- Then
$P(\bar{A}_1 \cap \bar{A}_2) \geq 1 - P(A_2) - P(A_1) = 1 - \alpha/4 - \alpha/4 = 1 - \alpha/2$

Using the Bonferroni inequality cont.

- The Bonferroni procedure is very general. To make joint confidence statements about multiple simultaneous predictions remember that

$$\begin{aligned} \hat{Y}_h &\pm t(1-\alpha/2; n-2)s\{pred\} \\ s^2\{pred\} &= MSE\left[1 + \frac{1}{n} + \frac{(X_h - \bar{X})^2}{\sum_i (X_i - \bar{X})^2}\right] \end{aligned}$$

- If one is interested in a $1-\alpha$ confidence statement about g predictions then Bonferroni says that the confidence interval for each individual prediction must get wider (for each h in the g predictions)

$$\hat{Y}_h \pm t(1-\alpha/2g; n-2)s\{pred\}$$

Note: if a sufficiently large number of simultaneous predictions are made, the width of the individual confidence intervals may become so wide that they are no longer useful.

A few notes on regression through the origin

- Sometimes it is known that the regression function is linear and that it *must* go through the origin.
- The normal error model for this case is $Y_i = \beta_1 X_i + \epsilon_i$
- The least squares and maximum likelihood estimators for β_1 coincide as before, the estimator is $b_1 = \frac{\sum X_i Y_i}{\sum X_i^2}$
- In regression through the origin there is only one free parameter (β_1) so the number of degrees of freedom of the MSE

  $$s^2 = MSE = \frac{\sum e_i^2}{n-1} = \frac{\sum (Y_i - \hat{Y}_i)^2}{n-1}$$

  is increased by one.
- This is because this is a "reduced" model in the general linear test sense and because the number of parameters estimated from the data is less by one.
- Care must be taken in interval estimation for parameters in this model to account for this difference.