Confidence Bands for Smoothers

Ben Powell

Durham

May 12, 2010

Notation

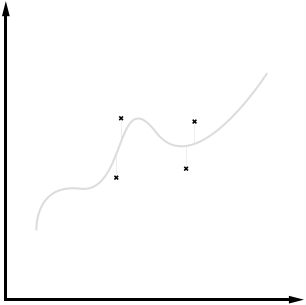
- $\mathbf{y} = \mathbf{f} + \epsilon$
- $\hat{\mathbf{f}} = \mathbf{S}\mathbf{y}$

Where can we use variance estimates?

- Confidence intervals
- Prediction intervals
- Curiosity

Local estimates

$$\hat{\sigma}_L^2 = \frac{1}{2(n-1)} \sum_{i=2}^{n} (y_i - y_{i-1})^2$$



Local estimates

$\hat{\sigma}_L^2$ is postively biased

$$
\begin{aligned}
E\left[\hat{\sigma}_L^2\right] &= E\left[\frac{1}{2(n-1)}\sum_{i=2}^{n}(y_i - y_{i-1})^2\right] \\
&= \frac{1}{2(n-1)}\sum_{i=2}^{n} E\left[(y_i - y_{i-1})^2\right] \\
&= \frac{1}{2(n-1)}\sum_{i=2}^{n} E\left[(f_i + \epsilon_i)^2 - 2(f_i + \epsilon_i)(f_{i-1} + \epsilon_{i-1}) + \cdots\right] \\
&\vdots \\
&= \sigma^2 + \frac{1}{2(n-1)}\sum_{i=2}^{n}(f_i - f_{i-1})^2
\end{aligned}
$$

Local estimates

(At least) two ways around this:

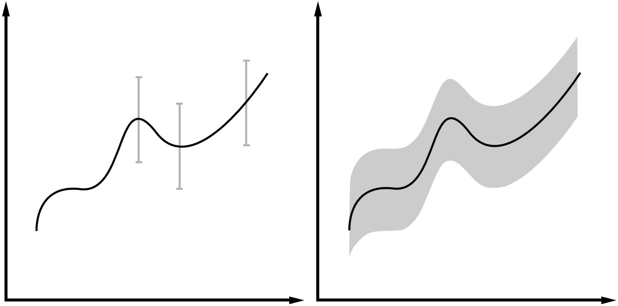
- Introduce different lags.
- Try: $\hat{\sigma}_R^2 = \frac{1}{n-2}\sum_{i=2}^{n-1}(y_i - \tilde{y}_i)^2$ where $\tilde{y}_i$ is the fitted value at x_i for a linear regression on the three closest data points.

Global estimates

$$\hat{\sigma}_\lambda^2 = \frac{1}{n-p}\sum_{i=1}^{n}(y_i - \hat{f}_i)^2,$$

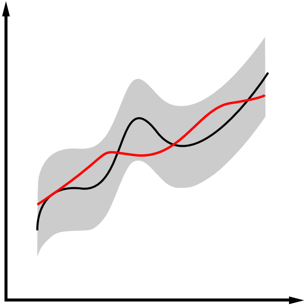
$p =$ degrees of freedom for smoother

Confidence bands

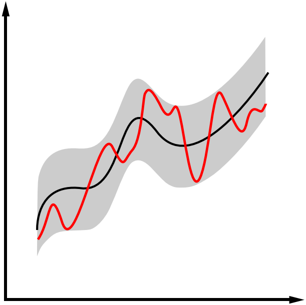


(a) Conf. intervals (b) Conf. band

Confidence bands



(c) sample smoother 1



(d) sample smoother 2

Notation

$$\hat{\mathbf{f}} = \mathbf{S}\mathbf{y}$$

$$\mathbf{y} = \mathbf{f} + \epsilon$$

$$\epsilon \sim N(\mathbf{0}, \mathbf{I}_n \sigma^2)$$

$$\epsilon^t \epsilon \sigma^{-2} \sim \chi^2_n$$

$$\epsilon^t \mathbf{S}^t (\mathbf{S}\mathbf{S}^t \sigma^2)^{-1} \mathbf{S}\epsilon \sim \chi^2_{rank(\mathbf{S})}$$

$$(\mathbf{S}\epsilon)^t \, (\mathbf{S}\mathbf{S}^t \sigma^2)^{-1} \, (\mathbf{S}\epsilon) \sim \chi^2_n$$

$$(\mathbf{S}f + \mathbf{S}\epsilon - \mathbf{S}f)^t \, (\mathbf{S}\mathbf{S}^t \sigma^2)^{-1} \, (\mathbf{S}f + \mathbf{S}\epsilon - \mathbf{S}f) \sim \chi^2_n$$

$$\left(\hat{\mathbf{f}} - \mathbf{S}f\right)^t (\mathbf{S}\mathbf{S}^t \sigma^2)^{-1} \left(\hat{\mathbf{f}} - \mathbf{S}f\right) \sim \chi^2_n$$

$$\hat{\epsilon} = \mathbf{y} - \hat{\mathbf{f}} = (\mathbf{I} - \mathbf{S})\mathbf{y} = (\mathbf{I} - \mathbf{S})\mathbf{f} + (\mathbf{I} - \mathbf{S})\epsilon$$

$$\mathbf{S}\mathbf{f} = \mathbf{f} \; ?$$

$$\hat{\epsilon} = (\mathbf{I} - \mathbf{S})\epsilon$$

Detour

Recall...

$$
\begin{aligned}
F(g) &= (\mathbf{y}-\mathbf{f})^t(\mathbf{y}-\mathbf{f}) + \lambda \mathbf{f}^t\mathbf{K}\mathbf{f} \\
&= (\mathbf{y}-\phi\beta)^t(\mathbf{y}-\phi\beta) + \lambda(\phi\beta)^t\mathbf{K}\phi\beta \\
&\ldots \\
&\Rightarrow \hat{\beta} = \left(\phi^t\phi + \lambda\phi^t\mathbf{K}\phi\right)^{-1}\phi^t\mathbf{y}
\end{aligned}
$$

- Using Sherman-Morrison-Woodbury formula and ϕ symmetric

$$
\begin{aligned}
\mathbf{S} &= \phi\left(\phi^t\phi + \lambda\phi^t\mathbf{K}\phi\right)^{-1}\phi^t \\
&= (\mathbf{I}-\lambda\mathbf{K})^{-1}
\end{aligned}
$$

$$\begin{aligned} &\mathbf{Sf} = \mathbf{f} \\ \Rightarrow &(\mathbf{I} - \lambda\mathbf{K})\mathbf{f} = \mathbf{f} \\ \Rightarrow &\mathbf{Kf} = \mathbf{0} \end{aligned}$$

...back to business

$$\hat{\epsilon}^t\hat{\epsilon} = \epsilon^t(\mathbf{I} - \mathbf{S})^t(\mathbf{I} - \mathbf{S})\epsilon$$

$$\frac{\hat{\epsilon}^t\hat{\epsilon}}{\sigma^2} = \frac{\epsilon^t(\mathbf{I} - \mathbf{S})^t(\mathbf{I} - \mathbf{S})\epsilon}{\sigma^2} = ?$$

$$\chi^2_{tr(\mathbf{I} - \mathbf{S} - \mathbf{S}^t + \mathbf{S}^t\mathbf{S})}$$

$$n - p = tr(\mathbf{I} - \mathbf{S} - \mathbf{S}^t + \mathbf{S}^t\mathbf{S})$$

$$\hat{\sigma}^2_\lambda = \frac{\hat{\epsilon}^t\hat{\epsilon}}{n-p} \sim \frac{\sigma^2\chi^2_{n-p}}{n-p}$$

Constructing the test

$$
\begin{aligned}
&\left(\hat{\mathbf{f}} - \mathbf{S}f\right)^t (\mathbf{S}\mathbf{S}^t \hat{\sigma}^2_\lambda)^{-1} \left(\hat{\mathbf{f}} - \mathbf{S}f\right) \\
&\sim \frac{n-p}{\sigma^2 \chi^2_{n-p}} \left(\hat{\mathbf{f}} - \mathbf{S}f\right)^t (\mathbf{S}\mathbf{S}^t)^{-1} \left(\hat{\mathbf{f}} - \mathbf{S}f\right) \\
&\sim \frac{n-p}{\chi^2_{n-p}} \chi^2_n \\
&\sim \frac{n-p}{\chi^2_{n-p}} (\chi^2_{n-p} + \chi^2_p) \\
&\sim (n-p)(1 + \frac{\chi^2_p}{\chi^2_{n-p}}) \\
&\sim (n-p) + pF_{p,n-p} \qquad \text{(test distribution)}
\end{aligned}
$$

Constructing the test

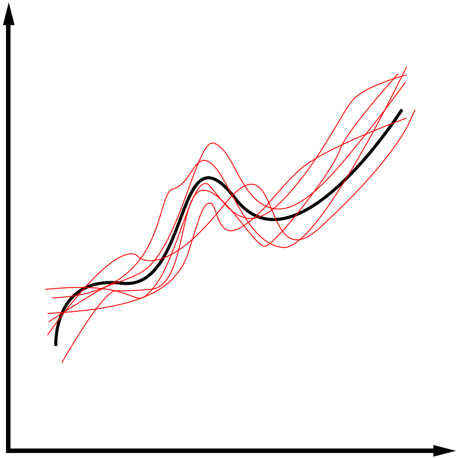
Nice result but assumptions may cause test to be misleading

- $\mathbf{Sf} = \mathbf{f}$
- $(\mathbf{I} - \mathbf{S})$ behaves like $(\mathbf{I} - \mathbf{H})$ from linear regression.

Viewing functions from the interval

- Draw MVN vectors from $N(\hat{\mathbf{f}}, \hat{\sigma}^2_\lambda)$
- Accept or reject depending on the test distribution statistic.

Viewing functions from the interval



Model Selection

Use F-test completely analogously to the linear regression (ANOVA) context.

If the fitted model $\hat{\mathbf{f}}_2 = \mathbf{S}_2\mathbf{y}$ at least contains all the terms of $\hat{f}_1 = \mathbf{S}_1\mathbf{y}$, with residual sums of squares RSS_2 and RSS_1 and effective degrees of freedom p_2 and p_1, then:

$$\frac{(RSS_1 - RSS_2)/(p_2 - p_1)}{RSS_2/(n - p_2)} \sim F_{p_2 - p_1, n - p_2}$$

References

- Hastie, T.J. & Tibshirani, R.J. (1990), **Generalized Additive Models**, Chapman and Hall.
- Green, P.J. & Silverman, B.W. (1994), **Nonparametric regression and generalized linear models**, Chapman and Hall.
- Rice, J. (1984), *Bandwidth Choice for Nonparametric Regression*, Annals of Statistics **12(4)**, pp 1215-1230.
- Tong, T. & Wang, Y. (2005), *Estimating residual variance in nonparametric regression using least squares*, Biometrika **92(4)**, pp 821-830.