Lecture 4
Estimating the Covariance Function

Dennis Sun
Stanford University
Stats 253

June 26, 2015

1. Review
2. The Hack
3. Model-Based Approach
4. Course Logistics

1. Review
2. The Hack
3. Model-Based Approach
4. Course Logistics

Generalized Least Squares

We saw that if the model is

$$\mathbf{y} = X\boldsymbol{\beta} + \boldsymbol{\epsilon},$$

where $\mathrm{E}[\boldsymbol{\epsilon}|X] = \mathbf{0}$ and $\mathrm{Var}[\boldsymbol{\epsilon}|X] = \Sigma$, then the best estimator is

$$\hat{\boldsymbol{\beta}}^{GLS} = (X^T\Sigma^{-1}X)^{-1}X^T\Sigma^{-1}\mathbf{y}.$$

Covariance Functions

- Σ is usually unknown, and there is no hope of estimating it from just n observations.
- So we parametrize Σ, i.e., $\Sigma = \Sigma_{\boldsymbol{\theta}}$.
- There is often a covariance function $\Sigma_{\boldsymbol{\theta}}(\mathbf{s}, \mathbf{s}')$ on the space where the observations lie, and we obtain $(\Sigma_{\boldsymbol{\theta}})_{ij} = \Sigma_{\boldsymbol{\theta}}(\mathbf{s}_i, \mathbf{s}_j)$.
- Examples include
  - $\Sigma_{\boldsymbol{\theta}}(\mathbf{s}, \mathbf{s}') = \max(\theta_1 - \theta_2 d(\mathbf{s}, \mathbf{s}'), 0)$
  - $\Sigma_{\boldsymbol{\theta}}(\mathbf{s}, \mathbf{s}') = \theta_1 \exp\{\theta_2 d(\mathbf{s}, \mathbf{s}')\}$.

Today

- We estimate $\boldsymbol{\theta}$ from the data.
- We will learn two ways: (1) the Hack$^{\text{TM}}$ and (2) model-based.

1. Review
2. The Hack
3. Model-Based Approach
4. Course Logistics

"The Chicken or the Egg" Problem

- To calculate $\hat{\boldsymbol{\beta}}^{GLS}$, we need an estimate of Σ.
- To estimate $\Sigma = \mathrm{Var}(\boldsymbol{\epsilon})$, we need an estimate of the error

$$\hat{\boldsymbol{\epsilon}} = \mathbf{y} - X\hat{\boldsymbol{\beta}}^{GLS},$$

so we need $\hat{\boldsymbol{\beta}}^{GLS}$.

Solution: Use $\hat{\boldsymbol{\beta}}^{OLS}$ as a "preliminary" estimate and estimate Σ using

$$\hat{\boldsymbol{\epsilon}} = \mathbf{y} - X\hat{\boldsymbol{\beta}}^{OLS}.$$

Estimating the Covariance

- Now that we have $\hat{\boldsymbol{\epsilon}}$, how do we estimate the covariance?
- $\Sigma_{ij} = \Sigma(d(\mathbf{s}_i, \mathbf{s}_j))$ is assumed to be a function of the distance between the points. So we need to estimate the covariance function $\Sigma(h)$.
- If data is regularly spaced, then estimate $\Sigma(h)$ by the covariance of the observations spaced h apart:

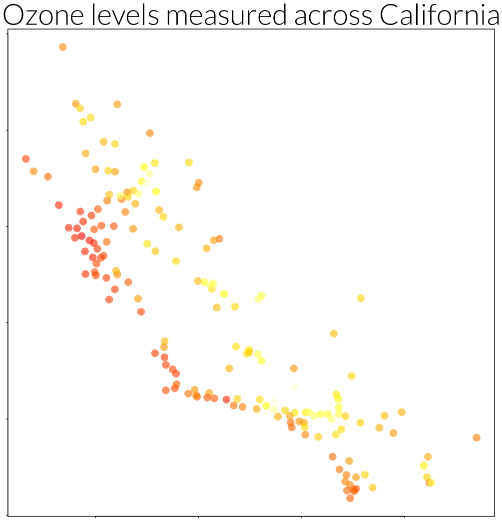
$$\hat{\Sigma}(h) = \frac{\sum_{(i,j)\in S_h} \hat{\epsilon}_i \hat{\epsilon}_j}{|S_h|},$$

  where $S_h = \{(i, j) : d(\mathbf{s}_i, \mathbf{s}_j) = h\}$.
- If data is irregularly spaced, then we look in a window around h.

$$\hat{\Sigma}(h) = \frac{\sum_{(i,j)\in S_{h,\delta}} \hat{\epsilon}_i \hat{\epsilon}_j}{|S_{h,\delta}|},$$

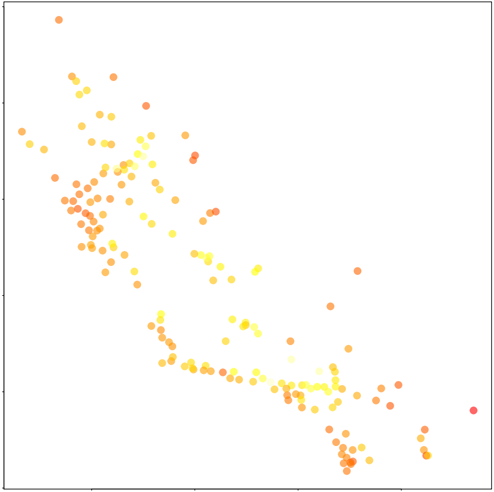
  where $S_{h,\delta} = \{(i, j) : d(\mathbf{s}_i, \mathbf{s}_j) \in [h - \delta, h + \delta]\}$.

California Ozone Example



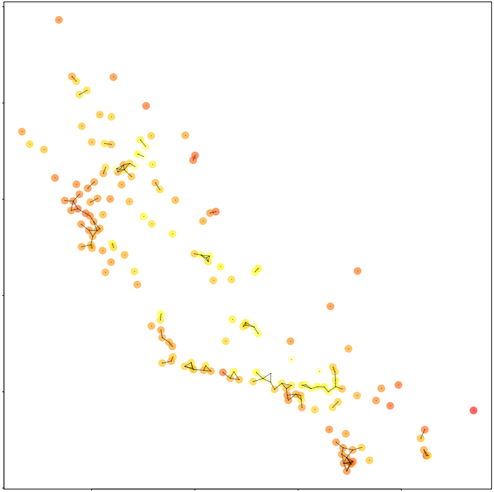
California Ozone Example

Residuals $\hat{\boldsymbol{\epsilon}}$ after regressing out latitude and longitude



California Ozone Example

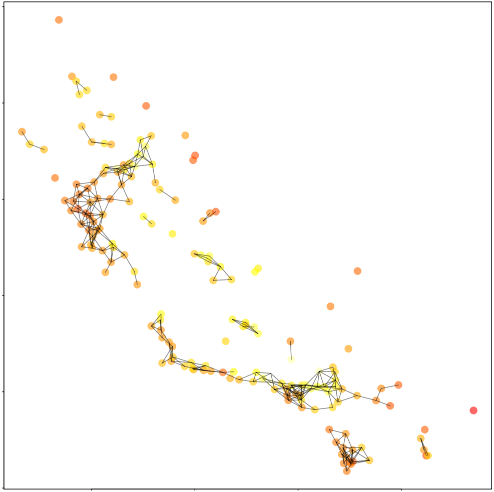
Pairs of observations where $0 \leq d(\mathbf{s}_i, \mathbf{s}_j) < .2$



$\hat{\Sigma}(.1) = 2.7 \times 10^{-4}$

California Ozone Example

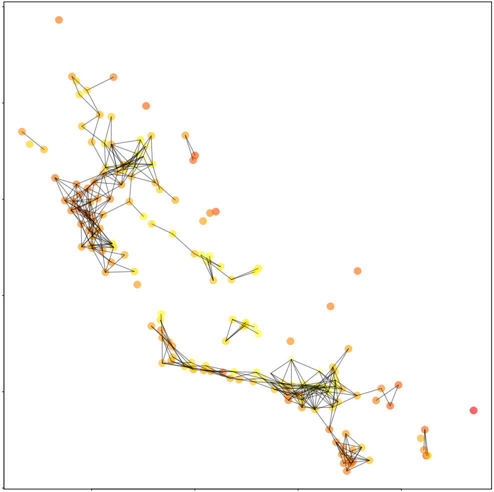
Pairs of observations where $.2 \leq d(\mathbf{s}_i, \mathbf{s}_j) < .4$



$$\hat{\Sigma}(.3) = 1.5 \times 10^{-4}$$

California Ozone Example

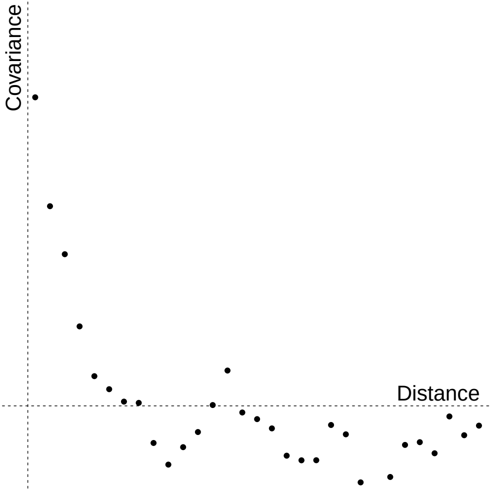
Pairs of observations where $.4 \leq d(\mathbf{s}_i, \mathbf{s}_j) < .6$



$$\hat{\Sigma}(.5) = 1.2 \times 10^{-4}$$

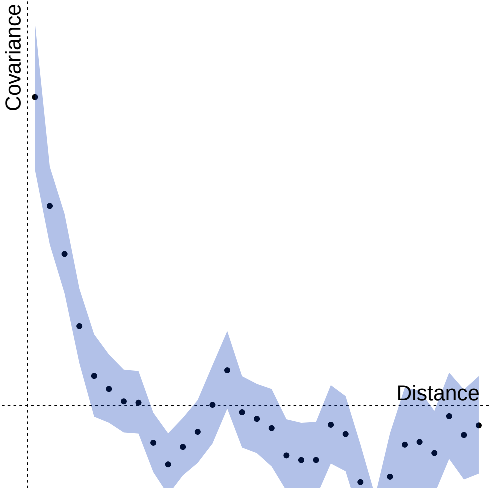
California Ozone Example

Plot of $\hat{\Sigma}(h)$



California Ozone Example

Calculate $SE = SD/\sqrt{n}$ to add error bars:



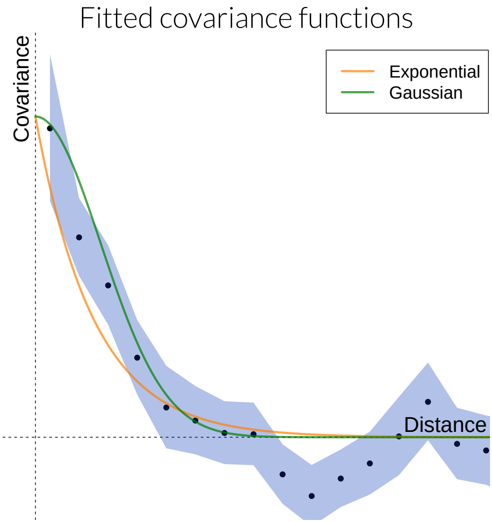
Estimating θ

- Can we just connect the dots and call that the covariance function?
- No! Not guaranteed to be positive definite.
- Parametrizing the covariance as $\Sigma_{\boldsymbol{\theta}}(h)$ helps ensure that the covariance is positive definite.
- Choose $\boldsymbol{\theta}$ to minimize the difference between the observed and theoretical covariance function:

$$\hat{\boldsymbol{\theta}} = \underset{\boldsymbol{\theta}}{\operatorname{argmin}} \sum_h w_h (\hat{\Sigma}(h) - \Sigma_{\boldsymbol{\theta}}(h))^2.$$

- The w_h are weights. We may want to downweight bins for which we have less data, i.e., $w_h \propto N_h$.

California Ozone Example



California Ozone Example

Now that we've estimated the covariance function as

$$\hat{\Sigma}(h) = .00025e^{-2.496h^2}$$

we can go back and fit generalized least squares!

1. Review
2. The Hack
3. Model-Based Approach
4. Course Logistics

Model-Based Approach

- The Hack suffers from two drawbacks:
  - We used $\hat{\boldsymbol{\beta}}^{OLS}$ to get a preliminary estimate of $\boldsymbol{\epsilon}$ when we really should be using $\hat{\boldsymbol{\beta}}^{GLS}$.
  - The final fitting of the covariance function to the data requires manual tuning of parameters: bin size, weights w_h, etc.
- Another approach is to assume a parametric model and estimate the parameters by maximum likelihood.

The Model

The model is still

$$\mathbf{y} = X\boldsymbol{\beta} + \boldsymbol{\epsilon},$$

with $\mathrm{E}[\boldsymbol{\epsilon}|X] = \mathbf{0}$ and $\mathrm{Var}[\boldsymbol{\epsilon}|X] = \Sigma_{\boldsymbol{\theta}}$, except now we further assume that $\boldsymbol{\epsilon}$ is normal.

We can now write down a likelihood for our data:

$$\frac{1}{(2\pi)^{n/2}(\det \Sigma_{\boldsymbol{\theta}})^{1/2}} \exp\left\{-\frac{1}{2}(\mathbf{y} - X\boldsymbol{\beta})^T \Sigma_{\boldsymbol{\theta}}^{-1}(\mathbf{y} - X\boldsymbol{\beta})\right\},$$

which we can optimize over $\boldsymbol{\beta}$ and $\boldsymbol{\theta}$ simultaneously.

Calculating the MLE

$$\frac{1}{(2\pi)^{n/2}(\det \Sigma_{\boldsymbol{\theta}})^{1/2}} \exp\left\{-\frac{1}{2}(\mathbf{y} - X\boldsymbol{\beta})^T \Sigma_{\boldsymbol{\theta}}^{-1}(\mathbf{y} - X\boldsymbol{\beta})\right\}.$$

The log-likelihood is

$$-\frac{1}{2}\log\det\Sigma_{\boldsymbol{\theta}} - \frac{1}{2}(\mathbf{y} - X\boldsymbol{\beta})^T \Sigma_{\boldsymbol{\theta}}^{-1}(\mathbf{y} - X\boldsymbol{\beta}).$$

Let's first optimize over $\boldsymbol{\beta}$ for any fixed $\boldsymbol{\theta}$. That is, what is $\hat{\boldsymbol{\beta}}(\boldsymbol{\theta})$?

Why, it's the GLS estimator! $\hat{\boldsymbol{\beta}}(\boldsymbol{\theta}) = (X^T\Sigma_{\boldsymbol{\theta}}^{-1}X)^{-1}X^T\Sigma_{\boldsymbol{\theta}}^{-1}\mathbf{y}$.

Let's plug this into the log-likelihood:

$$-\frac{1}{2}\log\det\Sigma_{\boldsymbol{\theta}} - \frac{1}{2}(\mathbf{y} - X\hat{\boldsymbol{\beta}}(\boldsymbol{\theta}))^T \Sigma_{\boldsymbol{\theta}}^{-1}(\mathbf{y} - X\hat{\boldsymbol{\beta}}(\boldsymbol{\theta})).$$

Now we "just" have to optimize this over $\boldsymbol{\theta}$.

Calculating the MLE

$$
\begin{aligned}
-\frac{1}{2}\log\det\Sigma_{\boldsymbol{\theta}} - \frac{1}{2}(\mathbf{y} - X\hat{\boldsymbol{\beta}}(\boldsymbol{\theta}))^T\Sigma_{\boldsymbol{\theta}}^{-1}(\mathbf{y} - X\hat{\boldsymbol{\beta}}(\boldsymbol{\theta}))& \\
= \ldots& \\
= -\frac{1}{2}\log\det\Sigma_{\boldsymbol{\theta}} - \frac{1}{2}\mathbf{y}^T\Sigma_{\boldsymbol{\theta}}^{-1}(I - X(X^T\Sigma_{\boldsymbol{\theta}}^{-1}X)^{-1}X^T\Sigma_{\boldsymbol{\theta}}^{-1})\mathbf{y}&
\end{aligned}
$$

To optimize this, you'll need to:

- compute derivative with respect to $\boldsymbol{\theta}$
- solve a highly non-convex problem.

However, the dimensionality of $\boldsymbol{\theta}$ is usually small, so you can do a brute-force grid search.

Summary

- The Hack is simple and fairly robust. We don't make any distributional assumptions.
- The model-based approach is more principled. But it may be difficult or impossible to compute the MLE in practice, and it is sensitive to the assumption of normality.

1 Review

2 The Hack

3 Model-Based Approach

4 Course Logistics

Logistics

- Homework 1 will be released tonight and due next Friday.
- It is a data analysis assignment that involves implementing some of the methods we've discussed.
- It is also a prediction competition. There will be prizes for the winners.
- I will provide starter code in R and maybe Python.
- Jingshu Wang, the TA for this course, will have office hours Thursdays 2-4pm.