Lecture 6
Covariance Estimation

Dennis Sun
Stats 253

July 9, 2014

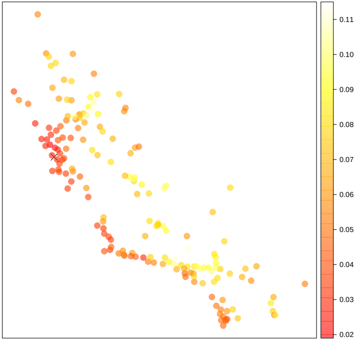
Outline of Lecture

1. Last Class
2. Stationarity and Covariance Estimation
3. Variogram Estimation
4. Putting it all together
5. Wrapping up

Where are we?

1. Last Class
2. Stationarity and Covariance Estimation
3. Variogram Estimation
4. Putting it all together
5. Wrapping up

Geostatistics / Kriging

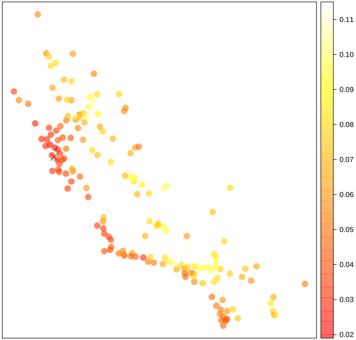


California Ozone Data

- $y(\boldsymbol{s})$ is Gaussian process with mean $\mu(\boldsymbol{s})$ and covariance $\Sigma(\boldsymbol{s}, \boldsymbol{s}')$.
- Assume $y(\boldsymbol{s}) \leftarrow y(\boldsymbol{s}) - \mu(\boldsymbol{s})$ so that $y(\boldsymbol{s})$ has mean $\mathbf{0}$.
- Observe $y(\boldsymbol{s}_i)$ or $z(\boldsymbol{s}_i) = y(\boldsymbol{s}_i) + \delta_i$ at locations $\boldsymbol{s}_1, ..., \boldsymbol{s}_n$.
- Goal is to estimate $y_0 \stackrel{def}{=} y(\boldsymbol{s}_0)$.
- The MMSE estimator is:

$$\mathrm{E}(y_0|\boldsymbol{y}) = \Sigma_{y_0,\boldsymbol{y}}\Sigma_{\boldsymbol{yy}}^{-1}\boldsymbol{y}$$
$$\mathrm{E}(y_0|\boldsymbol{z}) = \Sigma_{y_0,\boldsymbol{y}}(\Sigma_{\boldsymbol{yy}} + \tau^2 I)^{-1}\boldsymbol{y}$$

Geostatistics / Kriging



California Ozone Data

- If we do not assume normality, the above estimator is also the best linear estimator, i.e., of the form

$$f(\boldsymbol{y}) = \boldsymbol{w}^T \boldsymbol{y} = \sum_{i=1}^{n} w_i y_i$$

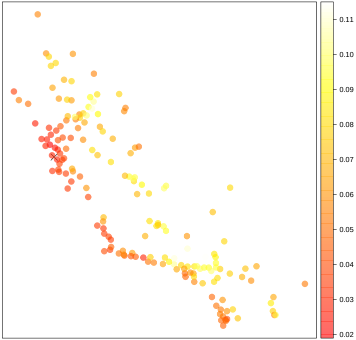
- To see this, write

$$\mathrm{E}(y_0 - \boldsymbol{w}^T \boldsymbol{y})^2 = \Sigma_{y_0 y_0} - 2\boldsymbol{w}^T \Sigma_{y_0, \boldsymbol{y}} + \boldsymbol{w}^T \Sigma_{\boldsymbol{y}\boldsymbol{y}} \boldsymbol{w}$$

- Optimizing over $\boldsymbol{w}$, we obtain

$$\boldsymbol{w}^T = \Sigma_{y_0, \boldsymbol{y}} \Sigma_{\boldsymbol{y}\boldsymbol{y}}^{-1}$$

Geostatistics / Kriging



California Ozone Data

Today's question: what if we don't know $\Sigma(s, s')$ in advance? Can we estimate it from the data?

Where are we?

Stationarity and Isotropy

- If $\boldsymbol{s} \in \mathbb{R}^k$, then $\Sigma(\boldsymbol{s}, \boldsymbol{s}')$ is a function on $\mathbb{R}^k \times \mathbb{R}^k$.
- We can reduce the dimensionality if we assume $\Sigma(\boldsymbol{s}, \boldsymbol{s}') = C(\boldsymbol{s} - \boldsymbol{s}')$. $C(\boldsymbol{h})$ is a function on only $\mathbb{R}^k$.
- If the covariance of y can be written in the form $\Sigma(\boldsymbol{s}, \boldsymbol{s}') = C(\boldsymbol{s} - \boldsymbol{s}')$, we say that y is **stationary**. (We've already assumed y is mean 0.)
- We can further assume that $\Sigma(\boldsymbol{s}, \boldsymbol{s}') = C(||\boldsymbol{s} - \boldsymbol{s}'||)$ so that $C(h)$ is a function on $\mathbb{R}$. In this case, we say that y is **isotropic**.

Issues in Estimation

- If data is regularly spaced, then estimating $C(\boldsymbol{h})$ is easy:

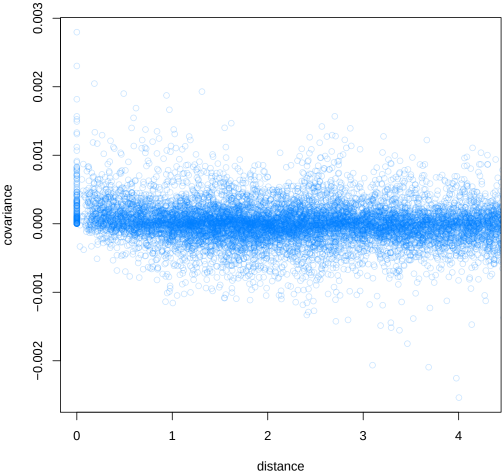
$$\hat{C}(\boldsymbol{h}) = \frac{1}{|N(\boldsymbol{h})|} \sum_{(\boldsymbol{s}_i, \boldsymbol{s}_j) \in N(\boldsymbol{h})} y(\boldsymbol{s}_i) y(\boldsymbol{s}_j)$$

  where $N(\boldsymbol{h}) = \{(\boldsymbol{s}_i, \boldsymbol{s}_j) : \boldsymbol{s}_i - \boldsymbol{s}_j = \boldsymbol{h}\}$.
- What if data is not regularly spaced?
- Simple solution: binning

$$N(d) = \{(\boldsymbol{s}_i, \boldsymbol{s}_j) : d - \epsilon \leq ||\boldsymbol{s}_i - \boldsymbol{s}_j|| \leq d + \epsilon\}.$$

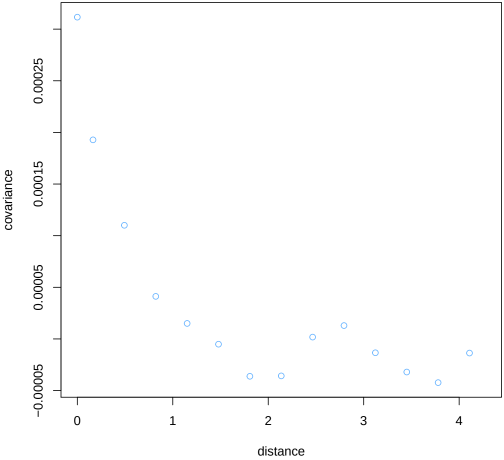
Sample Covariance

All $y(\boldsymbol{s}_i)y(\boldsymbol{s}_j)$, plotted against $||\boldsymbol{s}_i - \boldsymbol{s}_j||$



Sample Covariance

$\hat{C}(d_j)$, using a bin width $.33$



Parametric Modeling

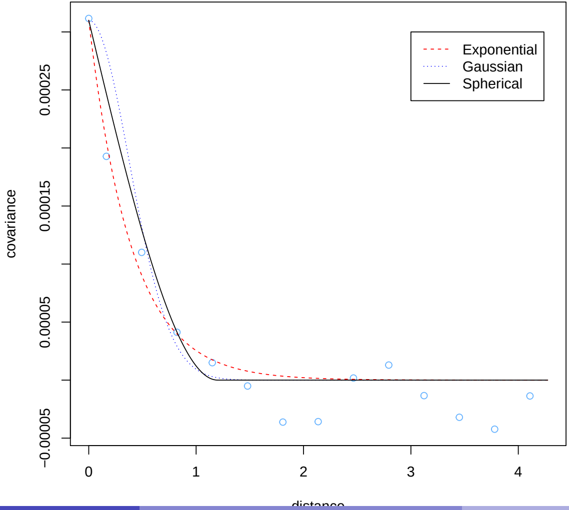
- $\hat{C}$ is not necessarily a valid covariance function.
- Use parametric models that guarantee positive-definiteness, e.g.,
  - Exponential: $C_{\sigma^2,\nu}(\boldsymbol{h}) = \sigma^2 \exp(-\nu||\boldsymbol{h}||)$
  - Gaussian: $C_{\sigma^2,\nu}(\boldsymbol{h}) = \sigma^2 \exp(-\nu||\boldsymbol{h}||^2)$
  - Spherical: $C_{\sigma^2,m}(\boldsymbol{h}) = \sigma^2 \left(1 - \frac{3||\boldsymbol{h}||}{2m} + \frac{||\boldsymbol{h}||^3}{2m^3}\right)$ for $||\boldsymbol{h}|| \leq m$.
  - Matérn: $C_{\sigma^2,\nu}(\boldsymbol{h}) = \frac{\sigma^2}{\Gamma(\eta + 1/2)} \left(\frac{||\boldsymbol{h}||}{2\nu}\right)^\eta K_\eta(\nu||\boldsymbol{h}||)$
- Procedure: Calculate $\hat{C}(d_j)$ for $d_1, ..., d_m$. Use iterative procedure to minimize

$$\sum_{j=1}^{m} w_j(\hat{C}(d_j) - C_\theta(d_j))^2$$

over parameters θ.

Parametric Modeling

$\hat{C}(d)$, fit using various parametric models



Where are we?

Variogram

- Historically, spatial statisticians have worked with the **variogram** instead of the covariance.

$$2\Gamma(\boldsymbol{s}, \boldsymbol{s}') = \mathrm{E}(y(\boldsymbol{s}) - y(\boldsymbol{s}'))^2$$

(or $\Gamma(\boldsymbol{s}, \boldsymbol{s}')$, the **semivariogram**).

- We say a process is (variogram-)stationary if $\Gamma(\boldsymbol{s}, \boldsymbol{s}') = \gamma(\boldsymbol{s} - \boldsymbol{s}')$ is a function of the difference only.
- If process is stationary, the variogram is easy to estimate:

$$2\hat{\gamma}(\boldsymbol{h}) = \frac{1}{|N(\boldsymbol{h})|} \sum_{(\boldsymbol{s}_i, \boldsymbol{s}_j) \in N(\boldsymbol{h})} (y(\boldsymbol{s}_i) - y(\boldsymbol{s}_j))^2$$

- Why prefer the variogram over the covariance?
  - More processes have a stationary variogram than a stationary covariance.
  - If we don't assume $y(s)$ has mean 0 but has mean μ, variogram doesn't require an estimate of μ.

Variogram

- All the calculations that we did before carry over, except with the covariance replaced by the variogram.
- For example, to predict y_0, the optimal weights are :

$$\boldsymbol{w}^T = \Gamma_{y_0,\boldsymbol{y}} \Gamma_{\boldsymbol{yy}}^{-1}$$

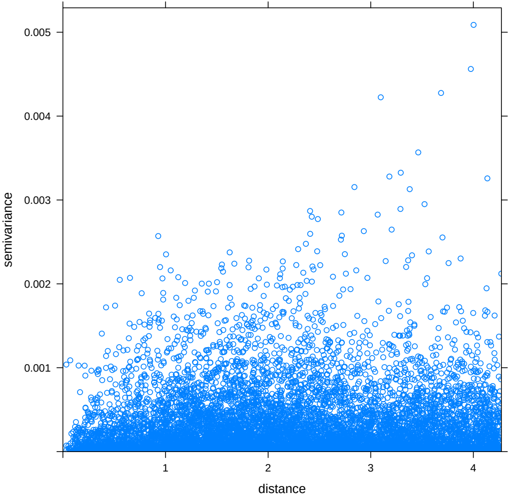
- We use parametric models for the variogram, e.g., Exponential: $\gamma_{\sigma^2,\nu}(\boldsymbol{h}) = \sigma^2(1 - \exp(-v||\boldsymbol{h}||))$
- And we choose parameters θ to minimize

$$\sum_{i=1}^{n} w_j(\hat{\gamma}(d_j) - \gamma_\theta(d_j))^2$$

- In keeping with tradition, we'll use the variogram, but keep in mind that everything could be done with the covariance.

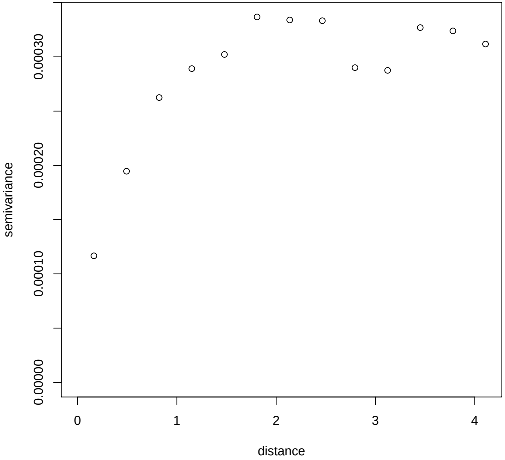
Sample (Semi)Variogram for Ozone Data

All $(y(\boldsymbol{s}_i) - y(\boldsymbol{s}_j))^2$, plotted against $||\boldsymbol{s}_i - \boldsymbol{s}_j||$



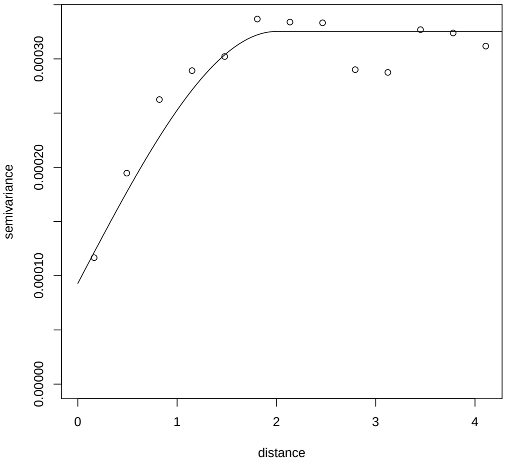
Sample (Semi)Variogram for Ozone Data

$\hat{\gamma}(d_j)$, using a bin width of $.33$

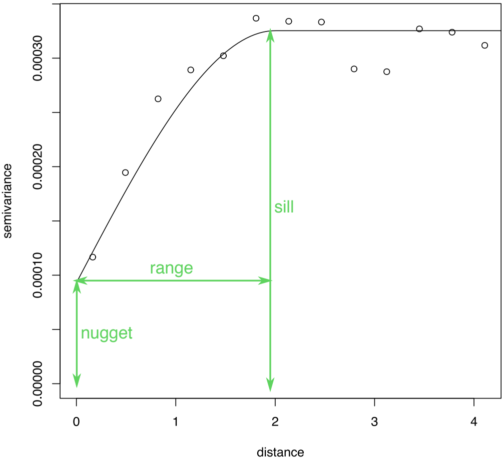


Sample (Semi)Variogram for Ozone Data

$\hat{\gamma}(d)$, fit using a spherical covariance

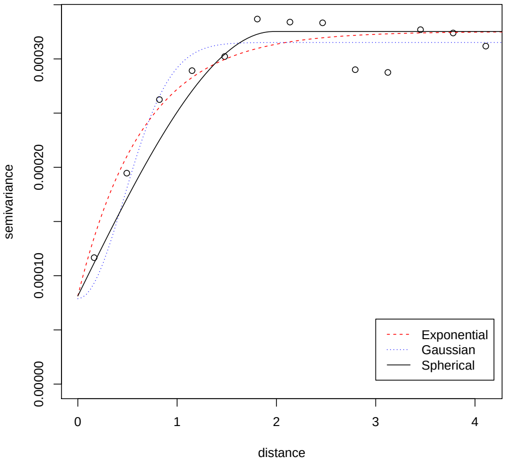


Sample (Semi)Variogram for Ozone Data



Sample (Semi)Variogram for Ozone Data

$\hat{\gamma}(d)$, fit using different parametric models



Where are we?

Summary so far

- We have come up with a way to estimate the covariance $\Sigma(\boldsymbol{s}, \boldsymbol{s}')$ from the data:

$$\hat{\Sigma}(\boldsymbol{s}, \boldsymbol{s}') = \hat{C}(\boldsymbol{s} - \boldsymbol{s}').$$

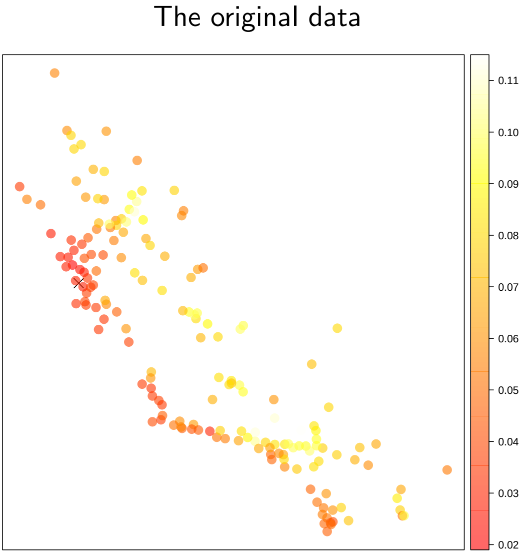
- Equivalently, we have a way to estimate the semivariogram $\Gamma(\boldsymbol{s}, \boldsymbol{s}')$ from the data:

$$\hat{\Gamma}(\boldsymbol{s}, \boldsymbol{s}') = \hat{\gamma}(\boldsymbol{s} - \boldsymbol{s}')$$

- Now we just need to plug $\hat{\Sigma}$ (or $\hat{\Gamma}$) into the kriging equations:

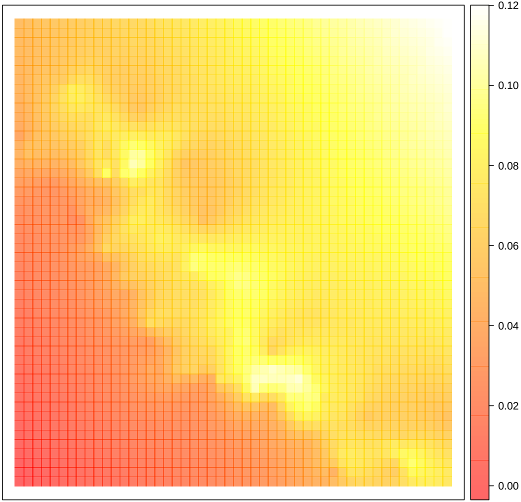
$$\hat{y}_0 = \hat{\Sigma}_{y_0,\boldsymbol{y}} \hat{\Sigma}_{\boldsymbol{y}\boldsymbol{y}}^{-1} \boldsymbol{y}$$
$$\hat{y}_0 = \hat{\Gamma}_{y_0,\boldsymbol{y}} \hat{\Gamma}_{\boldsymbol{y}\boldsymbol{y}}^{-1} \boldsymbol{y}$$

Original data



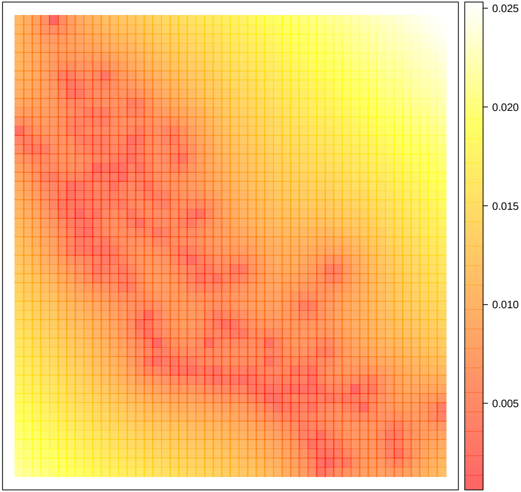
Kriging predictions

The kriging predictions $\hat{y}_0$ over a grid.



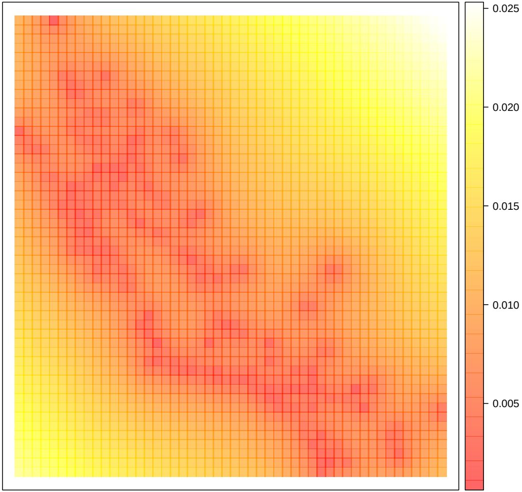
Kriging variances

The kriging standard deviations $\sqrt{\text{Var}(\hat{y}_0 - y_0)}$.



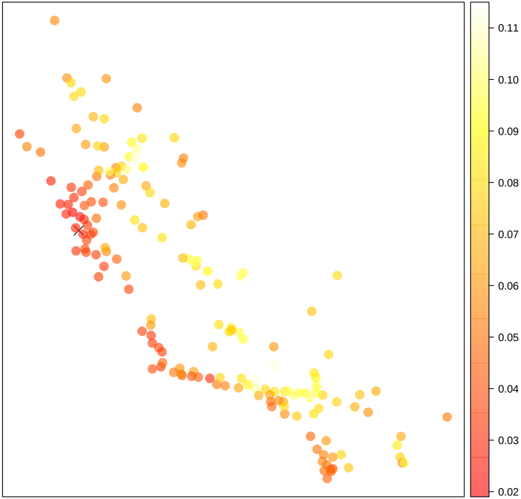
Kriging variances

Where is the variance lowest?



Original data

At the original data points!



Where are we?

1. Last Class
2. Stationarity and Covariance Estimation
3. Variogram Estimation
4. Putting it all together
5. Wrapping up

Summary

- Kriging gives optimal predictions if the covariance / variogram is known exactly.
- We've looked at how we might estimate the covariance / variogram from the data.
- This estimate can be plugged into the kriging equations, although the resulting estimate may no longer be optimal.

Administrivia

- Homework 2 is shorter; it is due Monday.
- Please work on it on your own.
- I will be holding Edgar's workshop tomorrow: I will be reviewing the Kalman filter for the first hour.