Lecture 1
Intro to Spatial and Temporal Processes

Dennis Sun
Stats 253

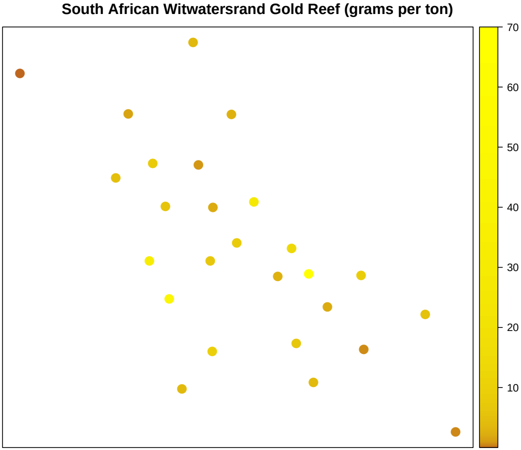
June 23, 2014

Outline of Lecture

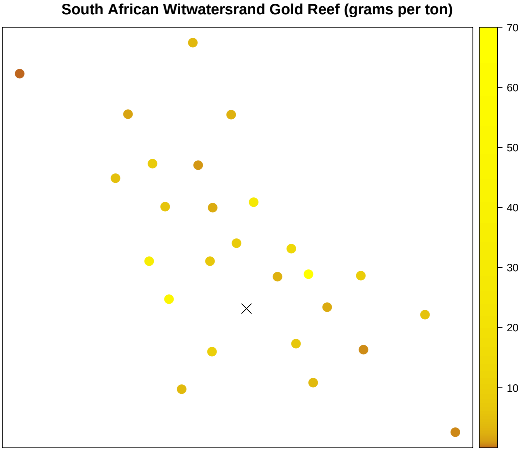
1. What is Spatial and Temporal Data?
   - Spatial Data
   - Temporal Data
   - Discussion
2. Course Logistics
3. Linear Regression
4. Autoregressions
5. Recap

Where are we?

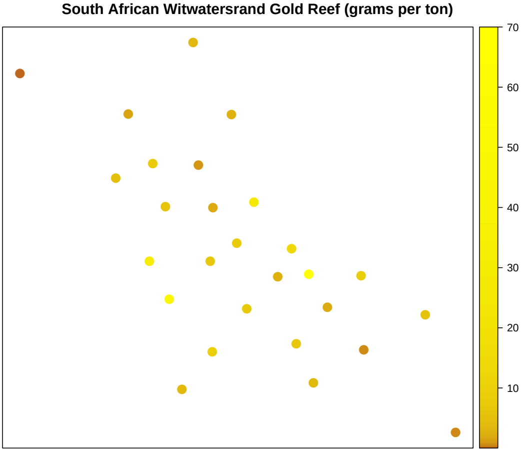
Geostatistics



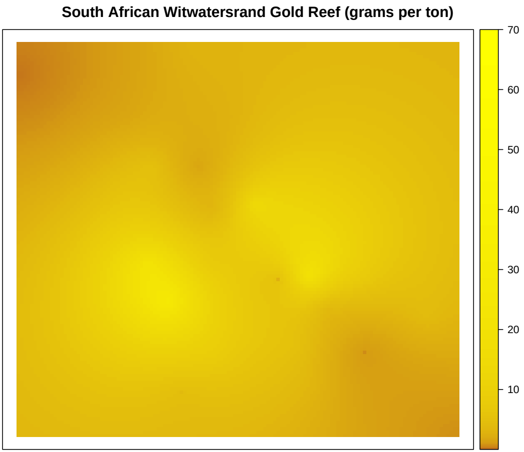
Geostatistics



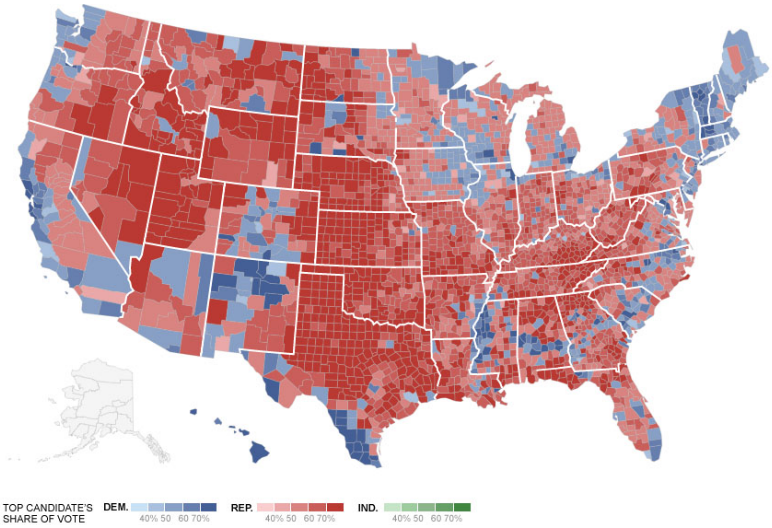
Geostatistics



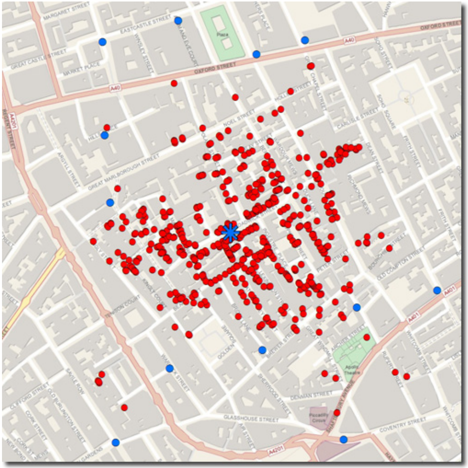
Geostatistics



Lattice (Areal) Data

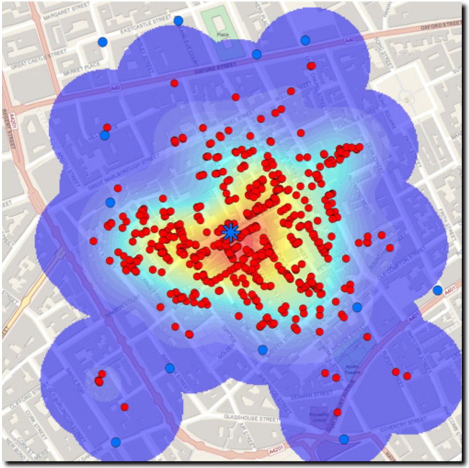


Point Processes



John Snow: 1854 Broad Street Cholera Outbreak

Point Processes



John Snow: 1854 Broad Street Cholera Outbreak

The Division of Spatial Statistics

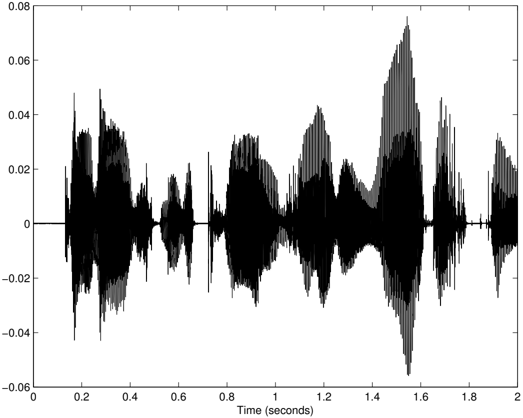
Cressie (1993) organizes spatial statistics into these three categories.

I Geostatistics: continuous space, labeled observations, goal is prediction

II Lattice (areal) data: discrete space, labeled observations, goal is inference

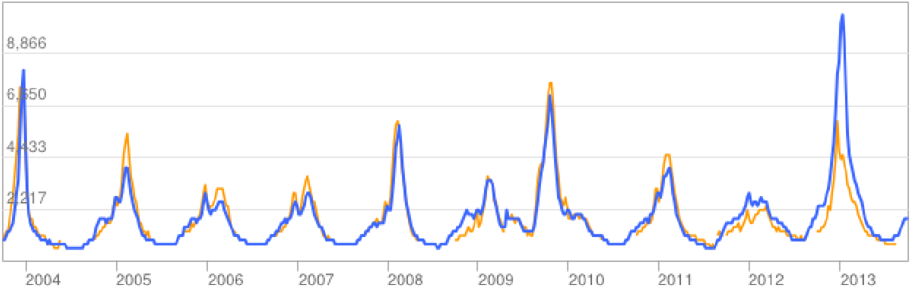
III Point processes: continuous space, unlabeled observations, goal is inference

Time Series Example 1



Human Speech

Time Series Example 2



Google Flu Trends

What do space and time have in common?

The observations y_t (or y_s) are correlated:

$$\mathrm{Cov}(y_t, y_{t'}) \neq 0 \text{ for } t \neq t'$$

Compare with the first assumption you see in most statistics courses:

Let y_i be i.i.d....

How are they different?

Time data is **ordered**, whereas there is no clear ordering for spatial data.

Where are we?

Organization

- Lectures: Mondays and Wednesdays 2:15-3:30pm in Education 334
- Instructors:
  - Dennis Sun
  - Edgar Dobriban
  - Jingshu Wang
- Contact? Office Hours? Sections?
- All information can be found on the course website: `stats253.stanford.edu`.

Content

- This is a new course.
- The material that we will be covering has not really been synthesized—because it is at the frontiers of statistics!
- We will be loosely following the books
  - Shumway and Stoffer. Time Series Analysis and Applications (with R Applications).
  - Sherman. Spatial Statistics and Spatio-Temporal Data.
- You don't have to purchase these books: they are available for free for Stanford students. (Link on course website.)
- Other useful references:
  - Bivand et al. Applied Spatial Data Analysis with R. (also available free)
  - Cressie and Wikle. Statistics for Spatio-Temporal Data.

Homeworks

- There will be about 4 short homeworks. Each homework will be a case study (data analysis).
- We will provide support for R, but you are free to use any computing environment (e.g., Python, Matlab, C....)
- You may work in pairs. If you do this, please turn in only one copy with both of your names on the front page.
- They will be graded on effort and completion only.

Project

- The goal of this course is to make a small but useful contribution to the world: e.g., a conference publication, open-source code, etc.
- This may sound intimidating, but there's actually a lot of low-hanging fruit in this subject!
- Grading rubric: produce something useful $\Rightarrow$ A+.
- I have no qualms about giving everyone an A+ if you all earn it!

Course Requirements

- The grade will be based on the final project.
- If taking this class CR/NC, the final project is optional, but you are required to complete all homeworks.
- Please enroll in this class if you are able: I promise that you will get much more out of it! The homeworks will be short and instructional.

Where are we?

Review of Linear Regression

$$\textbf{Model:} \quad y_i = \beta_1 x_{1i} + ... + \beta_p x_{pi} + \epsilon_i, \;\; \epsilon_i \overset{iid}{\sim} N(0, \sigma^2)$$

- Ordinary least squares: choose $\beta_1, ..., \beta_p$ by solving

$$\textbf{Estimation:} \quad \underset{\beta_1,...,\beta_p}{\operatorname{argmin}} \sum_{i=1}^{n} (y_i - (\beta_1 x_{1i} + ... + \beta_{pi} x_{pi}))^2.$$

- Write in vector notation as:

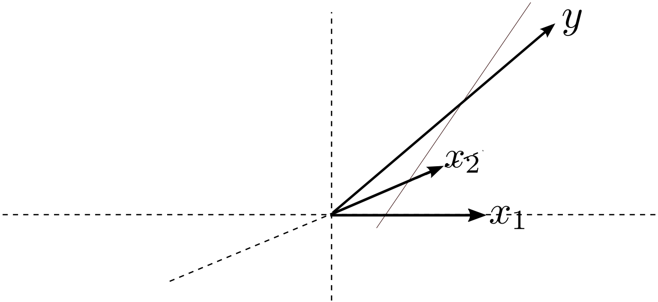
$$\textbf{Model:} \quad \boldsymbol{y} = X\boldsymbol{\beta} + \boldsymbol{\epsilon}, \;\; \boldsymbol{\epsilon} \sim N(0, \sigma^2 I)$$

$$\textbf{Estimation:} \quad \hat{\boldsymbol{\beta}} = \underset{\boldsymbol{\beta}}{\operatorname{argmin}} \; ||\boldsymbol{y} - X\boldsymbol{\beta}||^2.$$

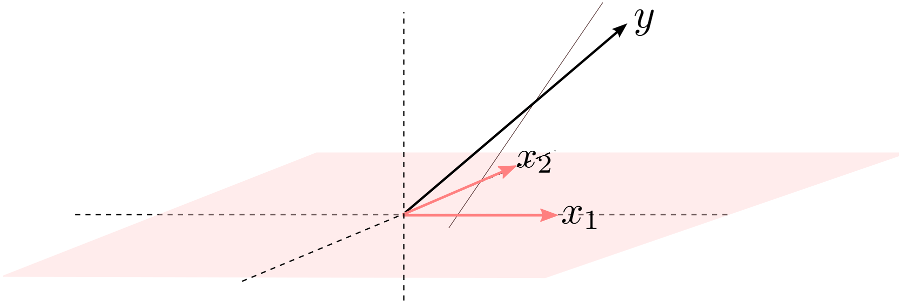
- Solve by differentiating: $\hat{\beta}$ must satisfy

$$2X^T(\boldsymbol{y} - X\hat{\boldsymbol{\beta}}) = 0 \quad \Rightarrow \quad \hat{\boldsymbol{\beta}} = (X^T X)^{-1} X^T \boldsymbol{y}.$$

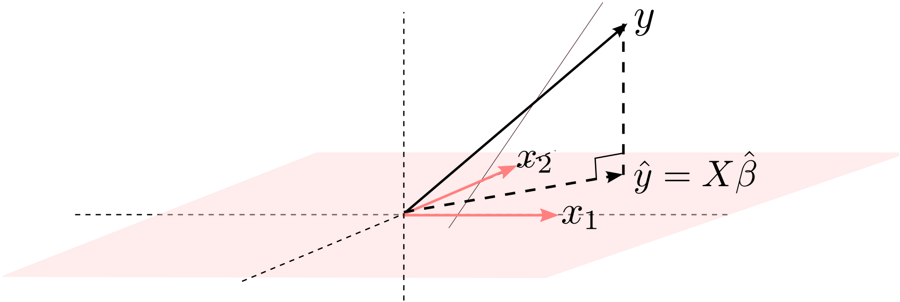
Regression: A Geometric Perspective



Regression: A Geometric Perspective



Regression: A Geometric Perspective



Properties of the Estimator

Recall that the model is

$$\boldsymbol{y} = X\boldsymbol{\beta} + \boldsymbol{\epsilon},\ \boldsymbol{\epsilon} \sim N(0, \sigma^2 I).$$

Is the estimator $\hat{\boldsymbol{\beta}}$ any good for estimating $\boldsymbol{\beta}$?

$$\hat{\boldsymbol{\beta}} = (X^T X)^{-1} X^T \boldsymbol{y} = \boldsymbol{\beta} + (X^T X)^{-1} X^T \boldsymbol{\epsilon}$$

$$\mathrm{E}(\hat{\boldsymbol{\beta}}) = \boldsymbol{\beta}$$

$$\mathrm{Var}(\hat{\boldsymbol{\beta}}) = \sigma^2 (X^T X)^{-1}$$

It is the **best linear unbiased predictor** (i.e., with the smallest variance).

Where are we?

Introducing Dependence

- In linear regression:

$$\mathrm{Cov}(y_i, y_j) = \mathrm{Cov}(\boldsymbol{x}_i^T \boldsymbol{\beta} + \epsilon_i, \boldsymbol{x}_j^T \boldsymbol{\beta} + \epsilon_j) = \mathrm{Cov}(\epsilon_i, \epsilon_j) = 0$$

- How can we introduce dependence?

$$y_t = \boldsymbol{x}_t^T \boldsymbol{\beta} + \phi y_{t-1} + \epsilon_t$$

- How do we estimate ϕ?

Autoregression

- **Idea:** Write as

$$\underbrace{\begin{bmatrix} y_2 \\ \vdots \\ y_n \end{bmatrix}}_{\tilde{\boldsymbol{y}}} = \underbrace{\begin{bmatrix} - & X_{2:n} & - & y_1 \\ & & & \vdots \\ & & & y_{n-1} \end{bmatrix}}_{\tilde{X}} \begin{bmatrix} | \\ \boldsymbol{\beta} \\ | \\ \phi \end{bmatrix} + \boldsymbol{\epsilon}$$

- Regress $\tilde{\boldsymbol{y}}$ on $\tilde{X}$ to obtain estimates of $\boldsymbol{\beta}$ and ϕ.
- Hence, we call this an **auto - regressive** (AR) model, meaning "regress on itself."
- Does this method "work"?

Does it work?

Let's set $X \equiv 0$ for now. So the model is

$$y_t = \phi y_{t-1} + \epsilon_t$$

The least squares estimate is

$$\hat{\phi} = (\boldsymbol{y}_{1:(n-1)}^T \boldsymbol{y}_{1:(n-1)})^{-1} \boldsymbol{y}_{1:(n-1)}^T \boldsymbol{y}_{2:n}$$

- Is it true that $\mathrm{E}(\hat{\phi}) = \phi$? **No.**
- Is it true that $\mathrm{Var}(\hat{\phi}) = \sigma^2 (\boldsymbol{y}_{1:(n-1)}^T \boldsymbol{y}_{1:(n-1)})^{-1}$? **No.**
- **However**, it turns out that $\hat{\phi}$ is consistent for ϕ.

$$\hat{\phi} \xrightarrow{p} \phi \text{ as } n \to \infty.$$

Simulation Example

- Suppose we observe 1000 observations of a random walk:

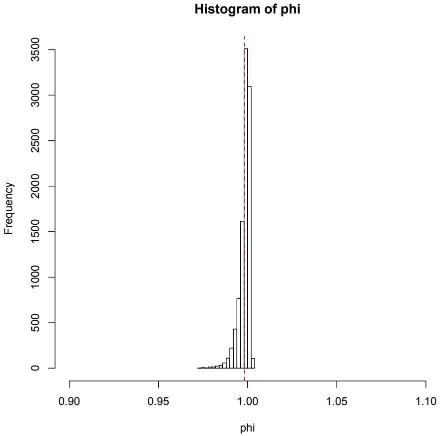
$$y_t = y_{t-1} + \epsilon_t, \;\; \epsilon_t \sim N(0,1)$$

  (In this case, $\phi = 1$.)
- Calculate $\hat{\phi}$ by regression.
- R Code:

```
phi <- sapply(1:10000, function(iter) {
    z <- cumsum(rnorm(1000))
    x <- z[1:999]
    y <- z[2:1000]
    return(sum(x*y)/sum(x*x))
})
```

Simulation Example

```
hist(phi, xlim=c(.9,1.1))
abline(v=mean(phi), col='red', lty=2)
```



sd(phi) $= .003$

Simulation Example

- The true standard error of $\hat{\phi}$ is $.003$.
- Does this agree with what linear regression would tell us?
- R Code:

```
z <- cumsum(rnorm(1000))
x <- z[1:999]
y <- z[2:1000]
model <- lm(y~x-1)
summary(model)
```

```
Call:
lm(formula = y ~ x - 1)

Residuals:
    Min      1Q  Median      3Q     Max
-3.2497 -0.6678  0.0396  0.6699  4.3311

Coefficients:
  Estimate Std. Error t value Pr(>|t|)
x 0.998757   0.001835   544.3   <2e-16 ***
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 0.9882 on 998 degrees of freedom
```

Simulation Example

- The true standard error of $\hat{\phi}$ is $.003$.
- Does this agree with what linear regression would tell us?
- R Code:

```
z <- cumsum(rnorm(1000))
x <- z[1:999]
y <- z[2:1000]
model <- lm(y~x-1)
summary(model)
```

```
Call:
lm(formula = y ~ x - 1)

Residuals:
    Min      1Q  Median      3Q     Max
-3.2497 -0.6678  0.0396  0.6699  4.3311

Coefficients:
  Estimate Std. Error t value Pr(>|t|)
x 0.998757   0.001835   544.3   <2e-16 ***
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 0.9882 on 998 degrees of freedom
```

Good

Simulation Example

- The true standard error of $\hat{\phi}$ is $.003$.
- Does this agree with what linear regression would tell us?
- R Code:

```
z <- cumsum(rnorm(1000))
x <- z[1:999]
y <- z[2:1000]
model <- lm(y~x-1)
summary(model)
```

```
Call:
lm(formula = y ~ x - 1)

Residuals:
    Min      1Q  Median      3Q     Max
-3.2497 -0.6678  0.0396  0.6699  4.3311

Coefficients:
  Estimate Std. Error t value Pr(>|t|)
x 0.998757   0.001835   544.3   <2e-16 ***
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1

Residual standard error: 0.9882 on 998 degrees of freedom
```

Good

Bad

Conclusions

- Linear regression gives good estimates for the coefficients of an AR process.
- However, it tends to **underestimate** the error (when observations are positively correlated).
- This will make effects look more significant than they really are!
- How can we fix this? *Next lecture.*

Where are we?

What We've Learned

- The similarities and differences between spatial and temporal data.
- Linear regression
- AR processes: the simplest model for correlated data
- The advantages and shortfalls of using regression to estimate parameters in AR processes.

`stats253.stanford.edu`