Lecture 16: Autocorrelation

Perry Williams, PhD

NRES 779

Bayesian Hierarchical Modeling in Natural Resources

Learning Objectives/Tasks

- Understand what autocorrelation is, why it's important statistically, and when it occurs in ecological data.
- Learn how to address autocorrelation in temporal models.
- Learn how to address autocorrelation in spatial models.

What is autocorrelation?

- Observations that are close together in proximity (either in space, in time, in genetics, etc...) tend to be more similar than observations further apart.
- If we don't account for this phenomenon using our covariates, x, then the similarities will manifest in the error terms, ϵ_i

$$y_i = \beta_0 + \beta_1 x_i + \epsilon_i$$
$$\epsilon_i \sim \text{Normal}(0, \sigma^2)$$

How have scienists dealt with this in the past?

Assumptions of Linear Regression

$$y_i = \beta_0 + \beta_1 x_i + \epsilon_i$$
$$\epsilon_i \sim \mathrm{Normal}(0, \sigma^2)$$

This assumes the residuals, ϵ_i, are:

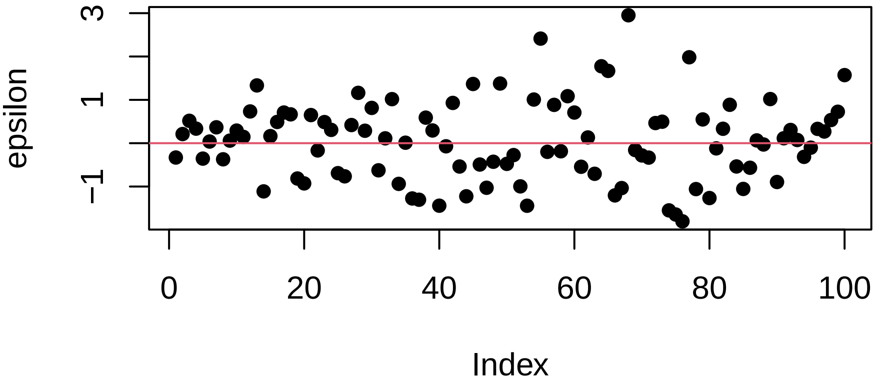
- Independent
- Identically distributed

Example of simulating iid data

```
n=100
sigma=1
epsilon=matrix(NA,n,1)
for(i in 1:n){
  epsilon[i]=rnorm(1,mean=0,sd=sigma)
}
pdf("iid.pdf",width=5,height=3)
plot(epsilon,pch=16)
abline(h=0,col=2)
dev.off()

## pdf
##   2
```

Each epsilon is independent of every other epsilon



epsilon
3
1
−1
0
20
40
60
80
100
Index

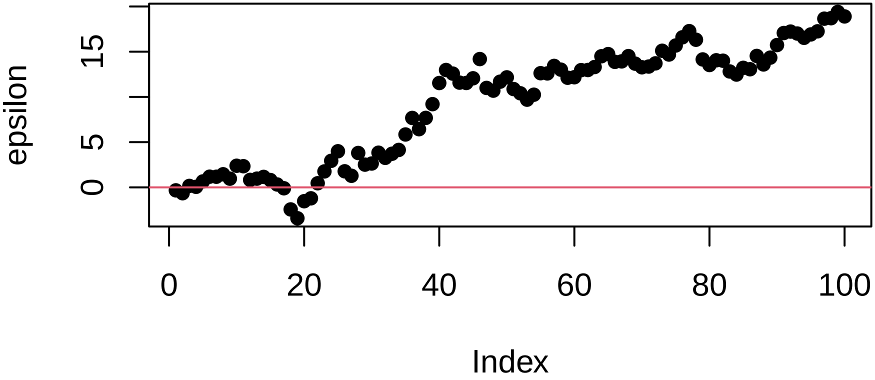
Example of simulating NON-iid data

```
for(i in 2:n){
  epsilon[i]=rnorm(1,mean=epsilon[i-1],sd=sigma)
}
pdf("noniid.pdf",width=5,height=3)
plot(epsilon,pch=16)
abline(h=0,col=2)
dev.off()

## pdf
##   2
```

Each epsilon depends on the value of epsilon right before it. In Ecology, we might see this when ϵ_i and ϵ_{i+1} are in close proximity. The errors around the mean will be similar.

The errors are correlated.



Why this is important statistically

Recall that using the chain rule of probability, that:

$$[\epsilon_1, \epsilon_2, \ldots, \epsilon_n|\boldsymbol{\theta}] = [\epsilon_1|\epsilon_2, \ldots, \epsilon_n, \boldsymbol{\theta}][\epsilon_2|, \epsilon_3, \ldots, \epsilon_n, \boldsymbol{\theta}] \tag{1}$$

When $\epsilon_1, \epsilon_2, \ldots, \epsilon_n$ are *independent*, Then, $[\epsilon_1|\epsilon_2, \ldots, \epsilon_n, \boldsymbol{\theta}]$ simplifies to $[\epsilon_1|\boldsymbol{\theta}]$, and the previous equation simplifies to:

$$[\epsilon_1, \epsilon_2, \ldots, \epsilon_n|\boldsymbol{\theta}] = [\epsilon_1|\boldsymbol{\theta}][\epsilon_2|\boldsymbol{\theta}]\times, \ldots, \times[\epsilon_n|\boldsymbol{\theta}] \tag{2}$$

and we can calculate the likelihood as

$$[\epsilon_1, \epsilon_2, \ldots, \epsilon_n|\boldsymbol{\theta}] = \prod_{i=1}^{n}[\epsilon_i|\boldsymbol{\theta}] \tag{3}$$

or in R and our Metropolis-Hastings ratio:

```
sum(dnorm(epsilon,mean=0,sd=sigma,log=TRUE))

## [1] -6428.064
```

Lack of independence

When the residuals are not independent, then the likelihood does not simplify!

How do we address this?

Build a model that accounts for this.

Model autocorrelation in data

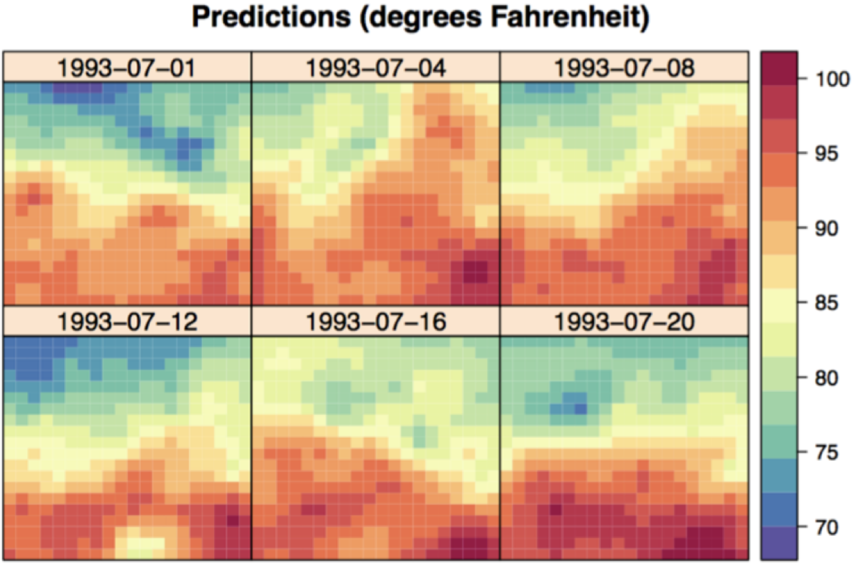
$$
\begin{aligned}
y_t &= \beta_0 + \beta_1 x_t + \epsilon_t \\
\epsilon_t &\sim \text{Normal}(\epsilon_{t-1}, \sigma^2) \\
\beta_0 & \text{ Normal}(0, 10000) \\
\beta_1 & \text{ Normal}(0, 10000) \\
\sigma^2 & \text{ IG}(\text{shape} = 0.0001, \text{rate} = 0.0001) \\
\epsilon_1 &\sim \text{Normal}(0, \sigma^2)
\end{aligned}
$$

The previous model is one example! There are a lot of ways to model autocorrelation in time! This is the field of time-series statistics. Here is another example:

$$
\begin{aligned}
y_t &= \mu_t + \epsilon_t \\
\mu_t &\sim \text{Normal}(\rho\mu_{t-1}, \sigma^2_\mu) \\
\epsilon_t &\sim \text{Normal}(0, \sigma^2) \\
\sigma^2 &\sim \text{IG}(\text{shape} = 0.0001, \text{rate} = 0.0001) \\
\sigma^2_\mu &\sim \text{IG}(\text{shape} = 0.0001, \text{rate} = 0.0001) \\
\mu_1 &\sim \text{Normal}(0, \sigma^2_{\mu_1})
\end{aligned}
$$

This is a specific time-series model known as an AR(1) model.

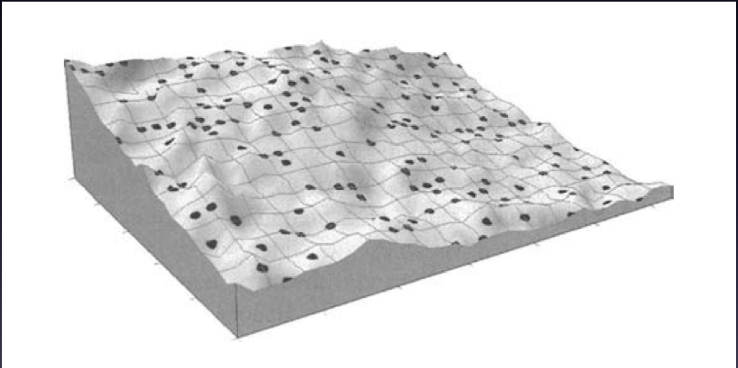
What about space?



Spatial Processes

1. **Spatial Point Processes:** Random locations are of interest, sometimes associated point characteristics ("marks").
2. **Continuous Spatial Processes:** Random measurements at fixed locations are of interest.
3. **Areal Spatial Processes:** Random measurements in fixed regions are of interest.

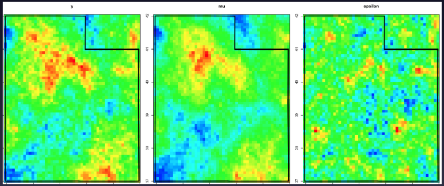
Continuous Spatial Processes



Imagine a smooth 2-D function

$y(\mathrm{s}) = \mu(\mathrm{s}) + \varepsilon(\mathrm{s})$, where $\mathrm{s} \in \Re$

1. μ: First order, the mean effect, a trend.
2. ε: Second order, often thought of as correlated error.



Gaussian Spatial Regression

$$\mathbf{y} = \mathbf{X}\beta + \varepsilon, \quad \varepsilon \sim N(\mathbf{0}, \mathbf{\Sigma})$$

Gaussian Spatial Regression

$$\mathbf{y} = \mathbf{X}\beta + \varepsilon, \quad \varepsilon \sim N(\mathbf{0}, \mathbf{\Sigma})$$

1. **First Order Structure:** $\mathbf{X}\beta$, the trend.
2. **Second Order Structure:** ε, where $\mathbf{\Sigma}$ can explain various forms of spatial autocorrelation.
3. **Prediction:** Kriging.

Note: This is referred to as "model-based geostatistics."

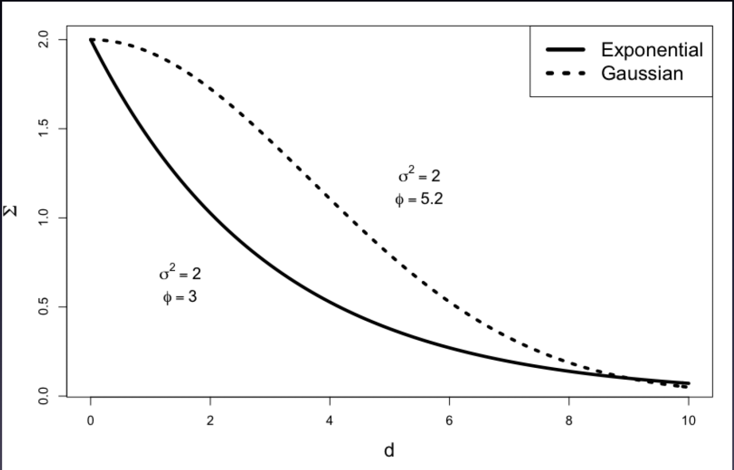
Covariance function: Covariogram

Parametric Covariance Functions:

- **Exponential**: $\Sigma_{i,j} = \sigma^2 \exp\left(-\frac{d_{i,j}}{\phi}\right)$
- **Gaussian**: $\Sigma_{i,j} = \sigma^2 \exp\left(-\frac{d_{i,j}^2}{\phi^2}\right)$

Note: $d_{i,j} =$ distance between locations i and j.

Parameteric Covariance Functions



Important Assumptions

- **Stationarity**: spatial structure does not vary with location.
- **Isotropy**: spatial structure does not vary with direction.

Two Sources of Error

Random Effects Approach:

$$\mathbf{y} = \mathbf{X}\beta + \varepsilon + \eta$$

1. **Correlated Error:** $\varepsilon \sim N(\mathbf{0}, \mathbf{\Sigma})$
2. **Uncorrelated Error:** $\eta \sim N(\mathbf{0}, \sigma_\eta^2 \mathbf{I})$

Two Sources of Error

Hierarchical Approach:

$$\mathbf{y} \sim \mathrm{N}(\mathbf{X}\beta + \epsilon, \sigma_\eta^2 \mathbf{I})$$

$$\varepsilon \sim \mathrm{N}(\mathbf{0}, \mathbf{\Sigma})$$

Two Sources of Error

Hierarchical Approach:

$$\mathbf{y} \sim \mathrm{N}(\mathbf{X}\beta + \epsilon, \sigma_\eta^2 \mathbf{I})$$

$$\varepsilon \sim \mathrm{N}(\mathbf{0}, \mathbf{\Sigma})$$

These both imply:

$$\mathbf{y} \sim \mathrm{N}(\mathbf{X}\beta, \mathbf{\Sigma} + \sigma_\eta^2 \mathbf{I})$$

Simulate a correlated continuous spatial process

1. Choose locations s_i for $i = 1, \ldots, n$.
2. Choose the mean $\boldsymbol{\mu}$. This could be a scalar or it could vary spatially.
3. Choose range parameter ϕ and variance component σ^2.
4. Compute distance matrix $\mathbf{D}$ between all n locations of interest.
5. Calculate covariance matrix $\boldsymbol{\Sigma} = \sigma^2 \exp\left(-\frac{\mathbf{D}}{\phi}\right)$.
6. Sample the n-dimensional vector $\mathbf{y} \sim \mathrm{N}(\boldsymbol{\mu}, \boldsymbol{\Sigma})$.

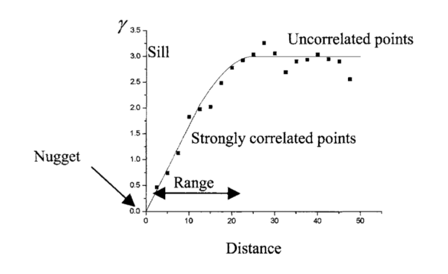
Assess the spatial correlation in a data set

1. Assume $\mathbf{y}$ is measured at n spatial locations.
2. Compute the residuals: $\mathbf{e} = \mathbf{y} - \boldsymbol{\mu}$.
3. Examine the residuals $\mathbf{e}$ for spatial correlation (i.e., autocorrelation).

Estimating spatial correlation

Empirical Semi-Variogram:

$$\hat{\gamma}(d) = \frac{\sum (e_i - e_j)^2}{2N(d)}$$



Fitted Variogram

Classical Estimation:

- After the empirical variogram is estimated at several bins for d, one can fit a parametric model to it.
- In this case, use $\hat{\gamma}(d)$ as the response variable and d as the covariate in weighted least squares or nonlinear regression to estimate σ^2, ϕ.

Fitted Variogram

Classical Estimation:

- After the empirical variogram is estimated at several bins for d, one can fit a parametric model to it.
- In this case, use $\hat{\gamma}(d)$ as the response variable and d as the covariate in weighted least squares or nonlinear regression to estimate σ^2, ϕ.

Semi-Variogram, Variogram, and Covariogram

- Semi-Variogram: $\gamma(d)$
- Variogram: $2\gamma(d)$
- Covariogram: $\Sigma(d) = \Sigma(0) - \gamma(d)$

Bayesian Geostatistical Model

- **Goal**: use Bayesian methods to estimate $\boldsymbol{\beta}$, σ^2, and ϕ.

$$\mathbf{y} \sim N(\mathbf{X}\boldsymbol{\beta}, \boldsymbol{\Sigma})$$

- $\Sigma_{i,j} = \sigma^2 \exp\left(-\frac{d_{i,j}}{\phi}\right)$.
- $\boldsymbol{\beta} \sim N(\boldsymbol{\mu}, \boldsymbol{\Sigma})$.
- $\sigma^2 \sim IG(r, q)$.
- Many choices for $\phi \sim [\phi]$.

Posterior:

$$[\boldsymbol{\beta}, \sigma^2, \phi|\mathbf{y}] = c \times [\mathbf{y}|\boldsymbol{\beta}, \sigma^2, \phi][\boldsymbol{\beta}][\sigma^2][\phi]$$

Prior Selection

Choices for range parameter ϕ:

- $\phi \sim \text{Gamma}(\gamma_1, \gamma_2)$
- $\log(\phi) \sim \text{N}(\mu_\phi, \sigma^2_\phi)$
- $\phi \sim \text{DiscUnif}(\Phi)$
- $\phi \sim \text{Half-Cauchy}(\gamma)$

Bayesian Kriging

- **Goal**: predict $y(\mathrm{s}_u)$ at unobserved location s_u, given the model and the data $y(\mathrm{s}_i)$ for $i = 1, \ldots, n$.

$$y(\mathrm{s}_i) = \mathbf{x}(\mathrm{s}_i)'\boldsymbol{\beta} + \varepsilon(\mathrm{s}_i)$$

- We need the posterior predictive distribution:

$$[y_u|\mathbf{y}] = \int\int\int [y_u|\mathbf{y}, \boldsymbol{\beta}, \sigma^2, \phi][\boldsymbol{\beta}, \sigma^2, \phi|\mathbf{y}]d\boldsymbol{\beta}d\sigma^2 d\phi$$

Predictive Full-Conditional

Notice that:

- $[y_u|\mathbf{y}, \boldsymbol{\beta}, \sigma^2, \phi] = N(\tilde{\mu}, \tilde{\sigma}^2)$

where,

- $\tilde{\mu} = \mathbf{x}_u'\boldsymbol{\beta} + \mathbf{c}'\Sigma^{-1}(\mathbf{y} - \mathbf{X}\boldsymbol{\beta})$
- $\tilde{\sigma}^2 = \sigma^2 - \mathbf{c}'\Sigma^{-1}\mathbf{c}$

and,

- $\mathbf{c} = (c_1, \ldots, c_n)'$
- $c_i = cov(\varepsilon_u, \varepsilon_i)$

- In MCMC, sample $y_u^{(k)} \sim N(\tilde{\mu}^{(k)}, \tilde{\sigma}^{2(k)})$.
- The Bayesian Kriging predictor is: $\mathrm{E}(y_u|\mathbf{y}) \approx \sum_{k=1}^{K} y_u^{(k)}/K$.

Generalized Spatial Models

- Binary:

$$y_i = \text{Bern}(p_i)$$
$$\text{logit}(p_i) = \mathbf{x}_i'\boldsymbol{\beta} + \varepsilon_i$$

- Count:

$$y_i = \text{Pois}(\lambda_i)$$
$$\log(\lambda_i) = \mathbf{x}_i'\boldsymbol{\beta} + \varepsilon_i$$