Basics of Modeling Spatial Processes

Lecture 5

SIE 512 Spatial Analysis

Fall 2009

Modeling Spatial Phenomena

We want to develop models for random variables distributed in space, and possibly evolving in time.

A stochastic process operating in a region of space can be referred to as a spatial stochastic process or a spatial random field.

We want to specifically consider statistical models for spatial stochastic processes.

Modeling Spatial Phenomena

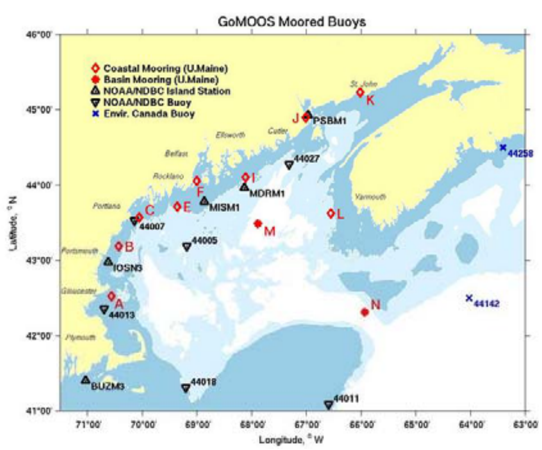
A statistical model for a **stochastic phenomena** consists of specifying a probability distribution for a random variable.

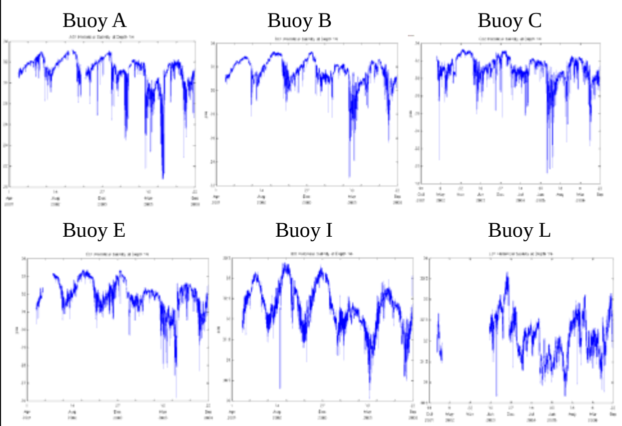
Easy for a simple phenomena like the toss of dice where we know the complete behavior

$$f_Y(y) = \frac{1}{6},$$

For complex environmental processes operating in space and evolving in time the model specification becomes more difficult.

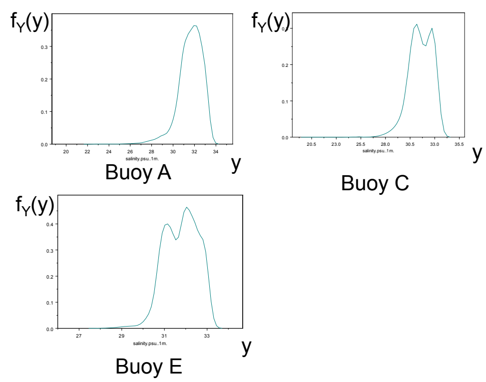
A variable can vary with different values at locations **s** over some region **R** and they can vary within a day and day to day. Values at neighboring locations may be related in various ways.





Salinity at 1m depth

Empirical Distributions of Salinity



Spatial processes model

Some notation stuff

$s \in R^d$ — ***s*** is a generic data location in d-dimensional Euclidean space

$\mathbf{s_1} = (s_{11}, s_{12})^T$ — Vector representing a point location $\begin{pmatrix} s_{11} \\ s_{12} \end{pmatrix}$

$\{Z(s) : s \in R\}$ — **Spatial stochastic process**

A set of possibly non-independent random variables

A complete statistical model requires specification of the joint probability distributions for all possible subsets of these random variables.

Specifying the Spatial Model

For complex spatial phenomena we rarely have all the theoretical knowledge to develop an appropriate statistical model.

So we rely on observation data to help specify a model

Assume a set of observations ($z_1, z_2, z_3, \ldots z_n$) taken at a set of locations ($s_1, s_2, s_3, \ldots s_n$) in R

This is referred to as a realization of a spatial process

It is simply one observation from the joint probability distribution of our spatial random variables $\{Z(s) : s \in R\}$

Alone it is not sufficient for a definitive model

Specifying the Spatial Model

Approach to model specification involves a combination of data plus some assumptions on the nature of the phenomena

Using some general knowledge on a phenomena we can specify a general form of a probability distribution with certain parameters left unspecified.

This general form of the model is then refined by fitting it to the observed data

Evaluation of the fitted model can lead to revised assumptions and a revised or adjusted model

Specifying the Spatial Model

Suppose we have some reason to propose a linear regression model for our spatial stochastic process

By proposing this model we make certain assumptions:

1. The random variables $\{Z(s) : s \in R\}$ are independent.
2. Their probability distributions only differ in their mean value
3. The mean value is a simple linear function of location

$$E(Z(s)) = (\beta_0 + \beta_1 s_1 + \beta_2 s_2)$$

4. Each Z(s) has a normal distribution about this mean and a constant variance

We can say the probability distribution is thus

$$Z(s) \sim N(\beta_0 + \beta_1 s_1 + \beta_2 s_2, \sigma^2)$$

Specifying the Spatial Model

The assumption of independence in the model removes the need to specify a joint probability distribution for every subset of the random variables Z(s).

The distributions just differ in their means as a function of the parameters $\beta_0, \beta_1, \beta_2$

We now use our data to estimate these parameters for our proposed model

Parameter Estimation

Method of Moments Estimators

Equates the sample moments of the data with the moments of the probability distribution

The rth moment of the distribution of a random variable X is often denoted as:

$$\mu_r^{'} = E(X^r)$$

First moment - mean

$$\mu = \mu_1^{'} = E(X)$$

$$\hat{\mu}_r^{'} = \frac{1}{n}\sum_{i=1}^{n} x_i^r$$

Maximum Likelihood Estimation

The aim of maximum likelihood estimation is to find the parameter value(s) that make the observed data most likely.

If the probability of a event X dependent on a model with parameter p is written

P (X | p)

then we can talk about the likelihood

L (p | X)

that is, the likelihood of the parameter given the data.

Maximum Likelihood Estimation

Say we toss a coin 100 times and observe 56 heads and 44 tails. Instead of assuming that p is 0.5, we want to find the MLE for p based on our observed data. Then we could ask whether or not this value differs significantly from 0.50.

How do we do this?

We find the value for p that makes the observed data most likely.

The observed data are considered fixed

They will be constants that are plugged into our proposed model - a binomial probability model :

n = 100 (total number of tosses)

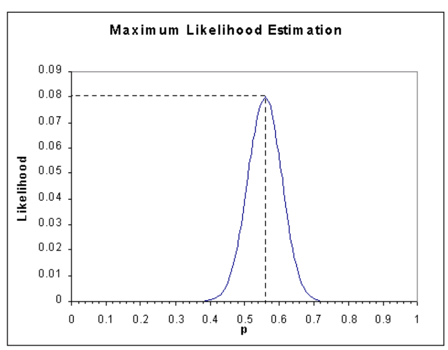
h = 56 (total number of heads)

$$L(p=.5\,|\,data)\frac{100!}{56!44!}.5^{56}.5^{44}=0.389$$

$$L(p=.52\,|\,data)\frac{100!}{56!44!}.52^{56}.48^{44}=0.581$$

p	L
0.48	0.0222
0.50	0.0389
0.52	0.0581
0.54	0.0739
0.56	0.0801
0.58	0.0738
0.60	0.0576
0.62	0.0378

If we graph these data across the full range of possible values for p we see the following likelihood surface.



Maximum Likelihood Estimation

Suppose we obtain a random sample of *m* rose blossums.

On each blossom we count the number of beetles present. The number of beetles observed on a given blossom is a discrete random variable, *X*.

In our random sample then we observe the values of *m* random variables, X_1, X_2, ... , X_m, one for each blossom in our sample

We observe x_1 beetles on blossom 1, x_2 beetles on blossom 2, etc.

We can describe a probability for this as $P = (X_1 = x_1, X_2 = x_2, X_3 = x_3, \ldots X_m = x_m)$

Maximum Likelihood Estimation

This expresses the joint probability

$$P = (X_1 = x_1, X_2 = x_2, X_3 = x_3, \ldots X_m = x_m)$$

Which we can also write as

$$P = ((X_1 = x_1) \cap (X_2 = x_2) \cap (X_3 = x_3) \cap \ldots \cap (X_m = x_m))$$

If events A and B are independent

$$P(A \cap B) = P(A)P(B)$$

$$P = (X_1 = x_1, X_2 = x_2, X_3 = x_3, \ldots X_m = x_m)$$
$$= P(X_1 = x_1) * P(X_2 = x_2) * P(X_3 = x_3), \ldots * P(X_m = x_m)$$
$$= \prod_{i=1}^{m} P(X_i = x_i)$$

Maximum Likelihood Estimation

Assume each of the probability terms in this product is a Poisson probability. We assume a Poisson distribution is a sensible model for the counts of beetles on blossoms.

Plugging these in and regrouping terms yields the following.

$$\prod_{i=1}^{m} P(X_i = x_i) = \prod_{i=1}^{m} \frac{\lambda^{x_i} e^{-\lambda}}{x_i!}$$
$$= \frac{\lambda^{x_1} e^{-\lambda}}{x_1!} * \frac{\lambda^{x_2} e^{-\lambda}}{x_2!} * \cdots \frac{\lambda^{x_m} e^{-\lambda}}{x_m!}$$
$$= \frac{\lambda^{\sum_{i=1}^{m} x_i} e^{-m\lambda}}{x_1! x_2! \ldots x_m!}$$

Maximum Likelihood Estimation

Using notation for probability functions with parameters we can write

$$f(x_1, x_2, \ldots x_m; \lambda) = \frac{\lambda^{\sum_{i=1}^{m} x_i} e^{-m\lambda}}{x_1! x_2! \ldots x_m!}$$

This is the probability of our data. If we knew λ we could calculate the probability of obtaining any set of values x_1, x_2, ... , x_m.

For fixed λ if we summed this expression over all possible values of x_1, x_2, ... , x_m we would get 1.

Maximum Likelihood Estimation

Now consider a different perspective. Since it's the data that are observed and the parameter that is unknown it makes more sense to think of this probability as a function of λ for fixed data, i.e.,

$$L(\lambda; x_1, x_2, \ldots x_m) = \frac{\lambda^{\sum_{i=1}^{m} x_i} e^{-m\lambda}}{x_1! x_2! \ldots x_m!}$$

We call this function the **likelihood function.** It is still the joint probability function for our data under the assumed probability model only by another name.

For both practical and theoretical reasons, it is preferable to work with the natural logarithm of the likelihood function, i.e., the **loglikelihood**. Starting with a generic probability model and proceeding to our independent Poisson model, the loglikelihood takes the following form.

$$\begin{aligned}\ell(\lambda; x_1, x_2, \ldots, x_m) &= \log\left(L(\lambda; x_1, x_2, \ldots, x_m)\right)\\ &= \log\left(\prod_{i=1}^{m} P(X_i = x_i)\right)\\ &= \sum_{i=1}^{m} \log\left(P(X_i = x_i)\right)\\ &= \sum_{i=1}^{m} \log\left(\frac{e^{-\lambda}\lambda^{x_i}}{x_i!}\right)\end{aligned}$$

Recall some of the basic properties of the logarithm function. For positive numbers *a* and *b*, and real number *n* we have the following.

1. $\log ab = \log a + \log b$ so the logarithm turns multiplication into addition.
2. $\log \frac{a}{b} = \log a - \log b$, so the logarithm turns division into subtraction.
3. $\log a^n = n \log a$, so the logarithm turns exponentiation into scalar multiplication

$$\begin{aligned}\ell(\lambda; x_1, x_2, \ldots, x_m) &= \sum_{i=1}^{m} \log\left(\frac{e^{-\lambda}\lambda^{x_i}}{x_i!}\right)\\ &= \sum_{i=1}^{m}\left(\log e^{-\lambda} + \log \lambda^{x_i} - \log x_i!\right)\\ &= \sum_{i=1}^{m}\left(-\lambda + x_i \log \lambda - \log x_i!\right)\\ &= -m\lambda + (\log \lambda)\sum_{i=1}^{m} x_i - \sum_{i=1}^{m} \log x_i!\end{aligned}$$

Maximum Likelihood Estimation

Going back to the example of our spatial observed data ($z_1, z_2, z_3, \ldots z_n$) suppose a joint probability distribution or density is $f(z_1, z_2, \ldots z_n; \theta)$

Where θ represents the parameters β of the proposed model

It represents the probability of the data occurring given the proposed model

Maximum Likelihood Estimation

The probability distribution of our proposed model was

$$Z(s) \sim N(\beta_0 + \beta_1 s_1 + \beta_2 s_2, \sigma^2)$$

$$L(\mu, \sigma \mid z_1, z_2 ... z_n) = \prod_{i=1}^{n} \left(\frac{1}{\sqrt{2\pi\sigma}} e^{\frac{(z_i - \mu)^2}{2\sigma^2}} \right)$$

$$\ln L(\mu, \sigma \mid z_1, z_2 ... z_n) = \frac{n}{2} \ln(2\pi) + n \ln(\sigma) - \sum_{i=1}^{n} \frac{(z_i - \mu)^2}{2\sigma^2}$$

Maximum Likelihood Estimation

For models with assumed independently distributed random variables that are normally distributed with constant variance maximum likelihood reduces to ordinary least squares.

Maximum Likelihood Estimation

Maximum likelihood estimation provides standard errors for the estimated parameters.

Related to how peaked the likelihood function is at its maximum. Sharp peaks provide more reliable estimates – smaller standard errors.

An approximate 95 percent confidence interval for a parameter estimate can be found as two standard deviations from the maximum likelihood estimate.

Modeling Spatial Phenomena

The behavior of spatial phenomena is a combination of first and second order effects

First Order effects

Describe large scale variation in the mean due to location or other explanatory variables.

Second order effects

Spatial dependence on a process - correlation in the deviations of values of the process from the mean

Example of distributions of metal filings

Modeling Spatial Phenomena

The mean function (the stochastic expectation) characterizes trends and systematic structures in space/time.

The covariance function on the residuals $C(s_i, s_j) = E((Z(s_i)-\mu(s_i))(Z(s_j)-\mu(s_j)))$ expresses correlations and dependencies – small scale interactions in space.

Spatial Phenomena

Because of combined first and second order effects – the assumption of independently distributed random spatial variables is usually violated

We will typically need to replace the independence assumption with an alternative that incorporates covariance structure to accommodate the second order effects

The second order component is concerned with behavior of stochastic deviations from the mean

The second order component is often modeled as a stationary spatial process

Stationarity

Stationarity is a form of location invariance

Stationarity is the quality of a process in which the statistical parameters (mean and standard deviation) of the process do not change with space or time.

Some degree of stationarity must be assumed to make inferences about the data

Stationarity

A spatial process $\{Z(s) : s \in R\}$ is said to be stationary if its statistical properties are independent of absolute location in R.

Implies the mean E(Z(s)), and Variance Var(Z(s)) are constant in R and do not depend on location.

Also the covariance COV($Z(s_i)$, $Z(s_j)$) between any two locations depends only on the relative location of the sites, the distance and direction between them, and not the absolute location in R.

Stationarity

First order stationarity $E(Y(s)) = \mu \quad \text{for all } s \in R$

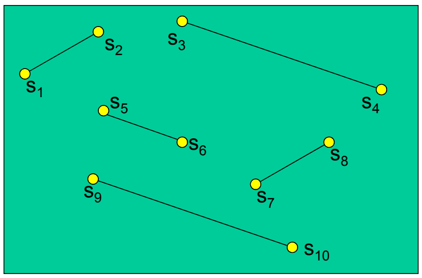
Second order stationarity

Depends only on relative distance and direction between observations

$Cov(Y(s_1),Y(s_2)) = Cov(Y(s_7),Y(s_8))$

$Cov(Y(s_3),Y(s_4)) = Cov(Y(s_9),Y(s_{10}))$

$Cov(Y(s_1),Y(s_2)) \neq Cov(Y(s_5),Y(s_6))$



Isotropic

Refers to a spatial process that evolves the same in all directions

We say a process is isotropic if in addition to being stationary, the covariance depends only on distance between locations and not the direction in which they are separated.

Anisotropic

A spatial process in which the correlation and covariance differs with direction

Most methods assume spatial correlation is isotropic

Stationary and Isotropic

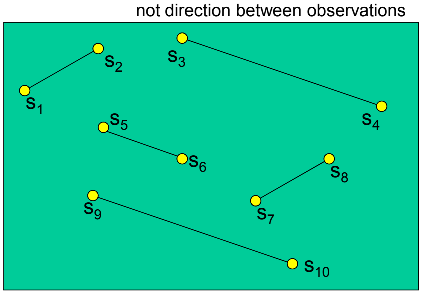
First order stationarity $E(Y(s)) = \mu \quad \text{for all } s \in R$

Second order stationarity and isotropy

Depends only on relative distance not direction between observations

$Cov(Y(s_1),Y(s_2)) = Cov(Y(s_7),Y(s_8))$

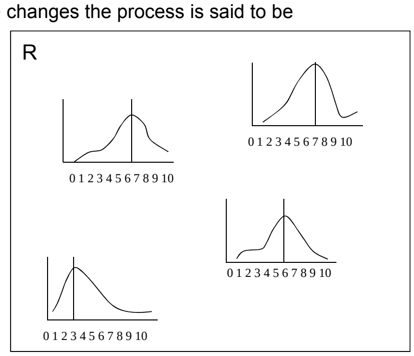
$Cov(Y(s_1),Y(s_2)) = Cov(Y(s_5),Y(s_6))$



Non Stationarity

If the process drifts over R, i.e the mean, variance, or covariance structure changes the process is said to be non-stationary.

Non-stationarity in the mean



Increasing west to east

Modeling Spatial Processes

For modeling purposes the general approach assumes non-stationarity (heterogenity) in the mean but stationarity in the second order effects.

Without an assumption of stationarity in the covariance, it becomes very difficult to fit a spatial model – too many parameters to estimate and not enough data

Assumptions

- Heterogeneity in the mean
- Covariance in the deviations from the mean is stationary