Applied Spatial Statistics: Lecture1

Douglas Nychka,
National Center for Atmospheric Research

Supported by the National Science Foundation *Boulder, Spring 2013*

Outline

- The additive model
- Colorado springtime temperatures
- Correlations and covariances
- Linear operations on random vectors
- Error in a Kernel estimator.

Organize several random variables in a vector.

Organize correlations/covariances for more than one random variables in a matrix.

Estimating a curve or surface.

The additive statistical model:

Given n pairs of observations (x_i, y_i), $i = 1, \ldots, n$

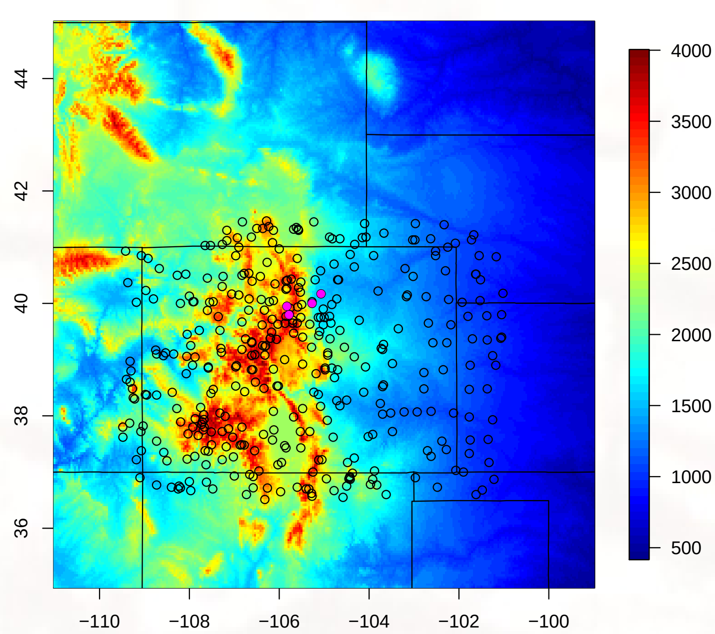
$$y_i = g(x_i) + \epsilon_i$$

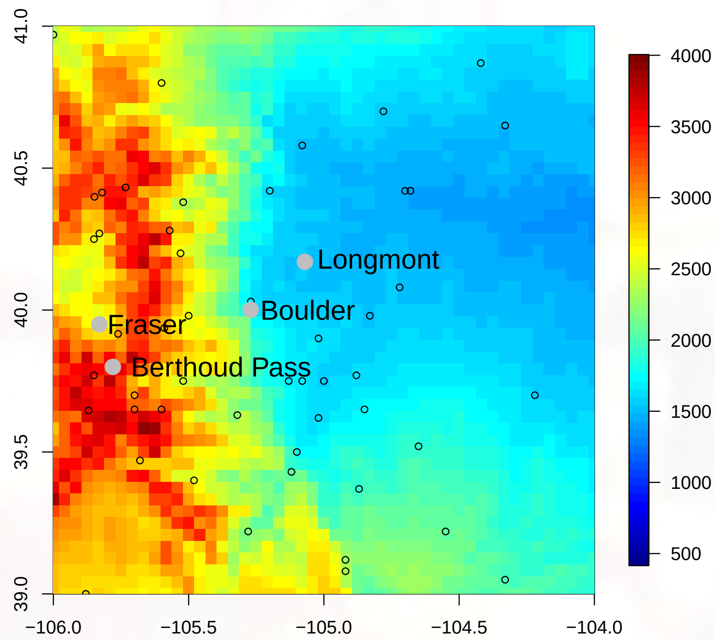
ϵ_i's are random errors
and g is an unknown smooth function.

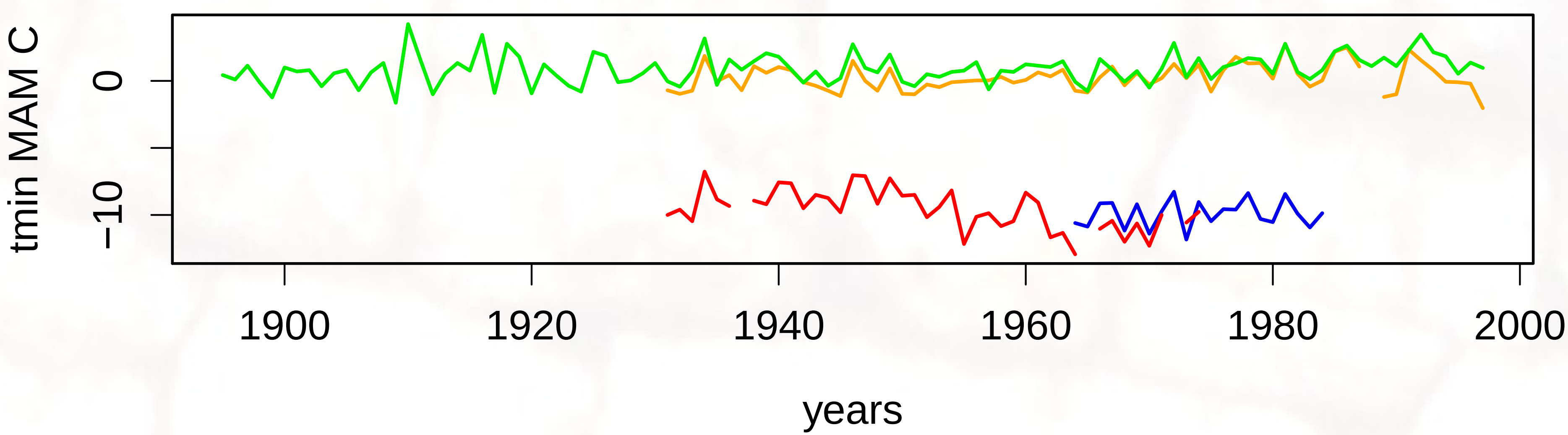
The goal is to estimate a function g based on the observations

$\boldsymbol{y} = (y_1, y_2, y_3, \ldots, y_n)$ *is a random vector*

Introduction



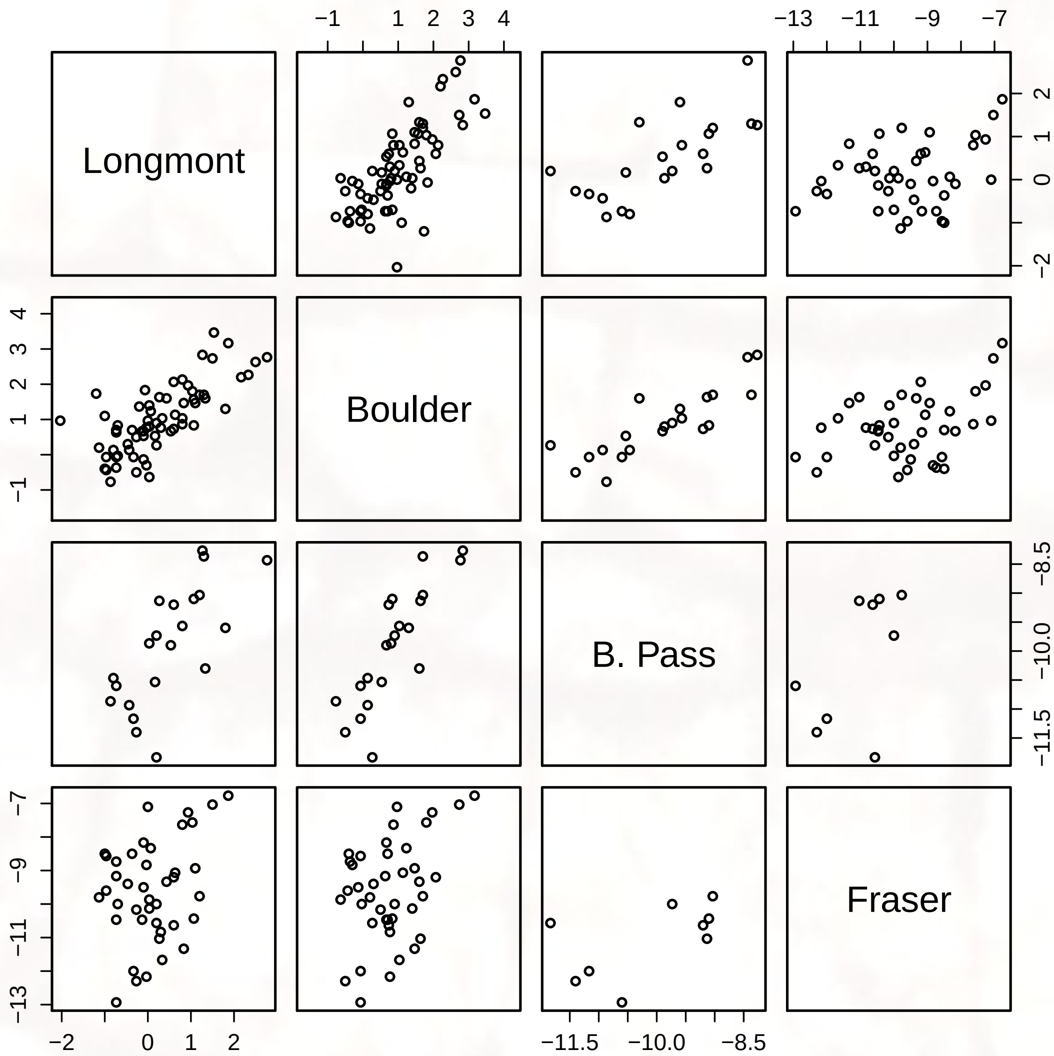




Some stats

	Longmont	Boulder	Berthoud Pass	Fraser
N	66.000	103.000	21.000	41.000
mean	0.261	0.890	-9.919	-9.604
Std.Dev.	0.989	1.092	1.020	1.507
min	-2.033	-1.633	-11.833	-12.933
Q1	-0.417	0.133	-10.600	-10.467
median	0.050	0.800	-9.867	-9.600
Q3	0.908	1.567	-9.133	-8.567
max	2.767	4.233	-8.267	-6.767
missing values	37.000	0.000	82.000	62.000

Scatter plots



Covariances and correlations

Sample mean variance:

$$\hat{\mu}_U = \frac{1}{n}\sum_{j=1}^{n} U_j$$

$$\hat{\sigma}_U^2 = \frac{1}{n-1}\sum_{j=1}^{n} (U_j - \hat{\mu}_U)^2$$

Sample covariance

$$\widehat{cov(U,V)} = \frac{1}{n-1}\sum_{j=1}^{n} (U_j - \hat{\mu}_U)(V_j - \hat{\mu}_V)$$

Sample correlation

$$\widehat{cor(U,V)} = \frac{\widehat{cov(U,V)}}{\hat{\sigma}_U \hat{\sigma}_V}$$

statistics for temperatures

Covariance matrix

	Longmont	Boulder	B. Pass	Fraser
Longmont	0.978	0.699	0.686	0.334
Boulder	0.699	1.192	0.819	0.531
B. Pass	0.686	0.819	1.041	0.687
Fraser	0.334	0.531	0.687	2.271

Correlation matrix

	Longmont	Boulder	B. Pass	Fraser
Longmont	1.000	0.731	0.720	0.302
Boulder	0.731	1.000	0.843	0.399
B. Pass	0.720	0.843	1.000	0.560
Fraser	0.302	0.399	0.560	1.000

As random variables

Assuming we have a large (or infinite population) the probability density function is the limiting case of a histogram for a sample.

Mean and variance for a random variable:

$$\mu = \int u f(u) du$$

$$\sigma^2 = \int (u - \mu)^2 f(u) du$$

f(u) = standard euponential

$$\mu = \int_0^\infty u e^{-u} du = 1, \quad \sigma^2 = \int_0^\infty (u-1)^2 e^{-u} du = 1$$

f(u) = standard normal

$$\mu = \int u \frac{1}{\sqrt{2\pi}} e^{-u^2/2} du = 0, \quad \sigma^2 = \int u^2 \frac{1}{\sqrt{2\pi}} e^{-u^2/2} du = 1$$

Covariance

f(u,y) 2-dimensional joint density for U and V

$$COV(U,V) = \int (u - \mu_U)(v - \mu_V) f(u,v) du dv$$

Mean vector and covariance matrix

For two random variables U, V:

$$\boldsymbol{\mu} = \begin{bmatrix} \mu_U \\ \mu_V \end{bmatrix}$$

$$\Sigma = \begin{bmatrix} \mathrm{VAR}(U) & \mathrm{COV}(U,V) \\ \mathrm{COV}(V,U) & \mathrm{VAR}(V) \end{bmatrix}$$

for U_1 and U_2

$$\boldsymbol{\mu} = \begin{bmatrix} \mu_1 \\ \mu_2 \end{bmatrix}$$

$$\Sigma = \begin{bmatrix} \mathrm{VAR}(U_1) & \mathrm{COV}(U_1,U_2) \\ \mathrm{COV}(U_2,U_1) & \mathrm{VAR}(U_2) \end{bmatrix}$$

This generalizes to more than two RVs!

Linear operations

If linear transformations are applied to random vectors there are simple formulas to find the new mean vector and new covariance matrix.

U a random vector with mean $\boldsymbol{\mu}$ and covariance Σ

Adding a constant vector: $\boldsymbol{b}$ a fixed vector $Z = U + \boldsymbol{b}$

mean of Z is $\boldsymbol{\mu} + \boldsymbol{b}$

covariance matrix of Z is Σ (why?)

Multiplying by a matrix: A a fixed matrix $Z = AU$

mean of Z is $A\boldsymbol{\mu}$

covariance matrix of Z is $A\Sigma A^T$

In general

Using some compact notation:

$\sim$ *(mean vector , covariance matrix)*

$Z = AU + \boldsymbol{b}$

$$U \sim (\boldsymbol{\mu}, \Sigma)$$

then

$$Z \sim (A\mu + \boldsymbol{b}, A\Sigma A^T)$$

This works for a single variable as a special case: $Z = aU + b$

$$U \sim (\mu, \sigma^2)$$

$$Z \sim (a\mu + b, a^2\sigma^2)$$

An example

Correlated variables (U_1, U_2) has a mean vector $(0, 0)$ and covariance matrix

$$\Sigma = \begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix}$$

$$A = \begin{bmatrix} 1 & 0 \\ .8 & .6 \end{bmatrix}$$

$Z = AU$ has mean of zero and covariance $A\Sigma A^T$

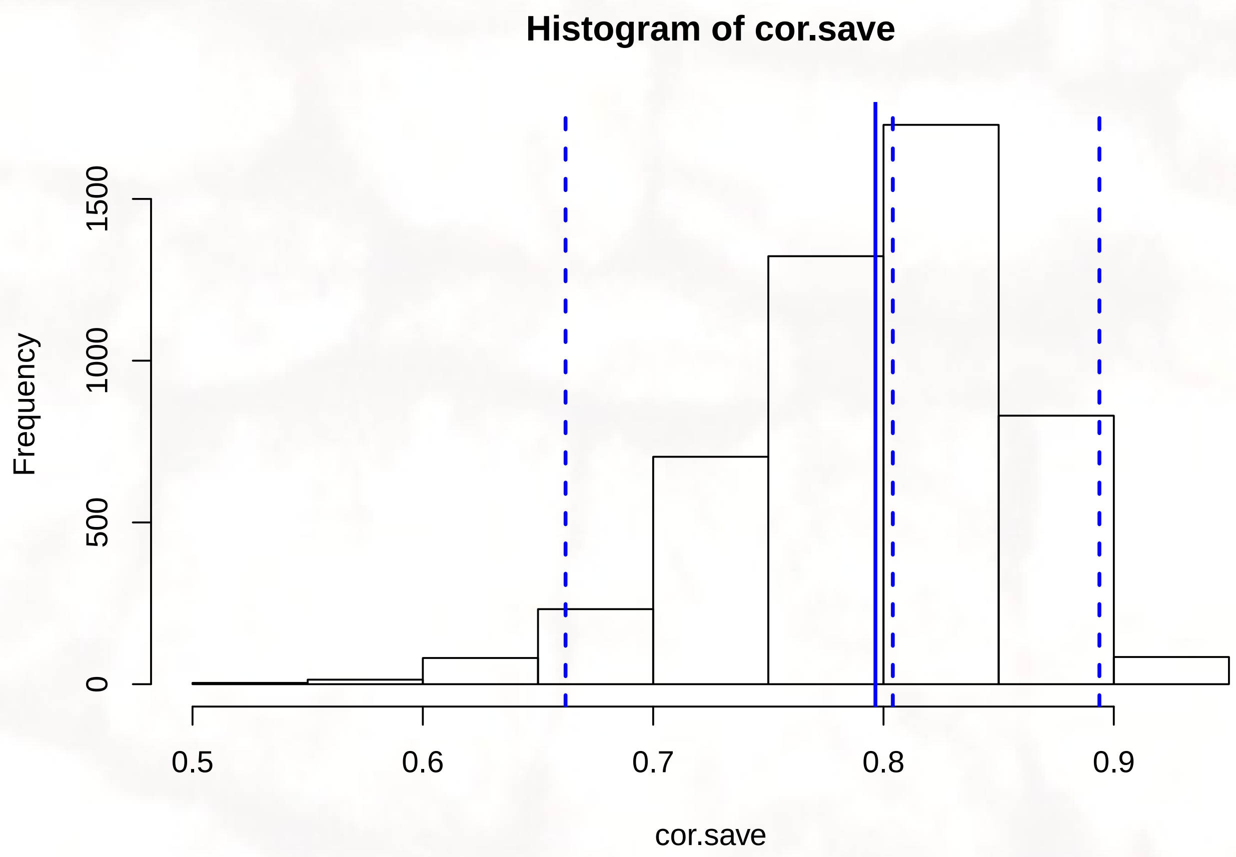
(worked out on board) In R:

```
Sigma<- matrix( c( 1,  0, 0,  1), nrow=2, ncol=2)
    A<- matrix( c( 1, .8, 0, .6), nrow=2, ncol=2)
A%*%Sigma%*%t(A)
     [,1] [,2]
[1,]  1.0  0.8
[2,]  0.8  1.0
```

Monte Carlo results

Histogram of 5000 correlation estimates:

Assuming correct correlation is .8, sample size of 40, and a normal distribution.



See R script for how this was done.

Dealing with two random vectors

One reason for working this out is to find the error of a kernel estimator.

Adding together two vectors

$$U \sim (0, \Sigma_U) and V \sim (0, \Sigma_V)$$

$COV(U,V) = \Sigma_{U,V}$ and U and V have same length (dimension).

$$Z = U + V \sim (0, \Sigma_U + 2\Sigma_{U,V} + \Sigma_V)$$

This handy formula follows from the general formula using partioned matrices.

If the cross covariance, Σ_{UV} is zero the covariance matrices just add.

$$Z \sim (0, \Sigma_U + \Sigma_V)$$

More on the cross covariance matrix

In general $\Sigma_{U,V}$ has the dimensions from U and V

U length m and V length n

- $\Sigma_{U,V}$ has m rows and n columns (i.e. called $m \times n$)
- i, j element of this matrix is $COV(U_i, V_j)$

An example

- A spatial data set of 200 observations (the y_is)
- evaluate the estimate at a grid of 1000 points (other $g(x)$s)
- Cross covariance matrix between the gs and the $y's$ is 1000×200.

Weighted sums

Adding together two vectors with different weights and sizes

$$U \sim (\mu_U, \Sigma_U) \text{ and } V \sim (\mu_V, \Sigma_V)$$

$COV(U,V) = \Sigma_{U,V}$

$$Z = A_U U + A_V V$$

$$Z \sim ($$

$$A_U \mu_U + A_V \mu_V,$$

$$A_U \Sigma_U A_U^T + A_U \Sigma_{U,V} A_V^T + A_V \Sigma_{V,U} A_U^T + A_V \Sigma_V A_V^T$$

$$)$$

Mean vectors add

Add covariances and but also include the cross covariances

The finale – kernel estimates

The estimate of $g(\boldsymbol{x})$ is a weighted sum of the observations.

Predict g at $\boldsymbol{x}$ as

$$\widehat{g}(\boldsymbol{x}) = \sum_i w_i y_i$$

In matrix/vector notation:

$$\widehat{g}(\boldsymbol{x}) = \boldsymbol{w}^T \boldsymbol{y}$$

One way to quantify the error is mean and variance of

$$g(\boldsymbol{x}) - \boldsymbol{w}^T \boldsymbol{y} = g(\boldsymbol{x}) + \left(-\boldsymbol{w}^T\right) \boldsymbol{y}$$

compare to

$$Z = A_U U + A_V V$$

to match the roles of U, V, A_U, A_V