Lecture 4: Likelihood-based parameter estimation

Spatial Statistics and Image Analysis

David Bolin
University of Gothenburg

Gothenburg
April 3, 2019

Project report

- Parts 1 and 2 should be documented as lab reports, describing what you did and the results you obtained but it does not need a detailed introduction and discussion.
- The ideal form of the report for Part 3 is in principle a journal paper, containing:
  1. Project title, author names, course name, date of report.
  2. Abstract/Summary: about 10-15 lines
  3. Introduction: Statement of problem, earlier work with references.
  4. Data description and source.
  5. Methods: Mathematical, statistical, computational, image analysis.
  6. Results with tables and figures.
  7. Discussion: include if possible here also comparison with results from literature.
  8. Conclusions, suggested continued studies.
  9. References.

Project work

The project consists of three parts:

1. Image reconstruction (based on computer exercises 1-4).
2. Image segmentation (based on computer exercises 5-8).
3. Problem of your own choice.

For the third part:

- A list of project suggestions is available at `http://www.math.chalmers.se/Stat/Grundutb/CTH/tms016/1819/projects2019.html`
- Send a planning report for your project to me via email, no later than May 3. It should contain name of project, your project group number, purpose of project, data description including source of data, if suitable also a sketch of methodology and a few literature references.

Deadlines

Sign up for a project group in PINGPONG as soon as possible (Before the deadline for the planning report).

For Parts 1 and 2:

- A PDF containing these two parts should be submitted via PINGPONG at the latest May 17, 23:55. Include the Matlab code as a zip-file.

For Part 3:

- Email with planning report , at the latest April 25, 23:55.
- A PDF containing a preliminary version of the report should be submitted together with Parts 1 and 2 at the latest May 17, 23:55. It does not need to be complete.
- The final version of the report should be submitted at the latest May 31, 23:55, together with the Matlab code as a zip-file, as a revision of the submitted project in PINGPONG.

Project seminars

- The final two lectures (and possible also one of the computer exercises) will be devoted to project seminars.
- In the seminar there will be 9 minutes allotted to each student.
- You are expected to use 7 minutes (not more) for each participant in your project describing what you have done so far and what you plan to do.
- After this there will be $2N$ minutes left for discussion, where N is the number of students in the project.
- The object of the seminar is that you should get feedback from the audience, both to point out parts that are less clear and suggestions on what to do, the goal being to help you to write as good a report as possible.

Least-squares parameter estimation

Model

$$Y_i = \mathbf{B}(\mathbf{s}_i)\boldsymbol{\beta} + X(\mathbf{s}_i) + \varepsilon_i, \quad i = 1, \ldots, N$$

where $\varepsilon_i \sim \mathsf{N}(0, \sigma_e^2)$ and $X(\mathbf{s})$ is a Gaussian random field.

In the classical geostatistical approach, the parameters are estimated in four steps:

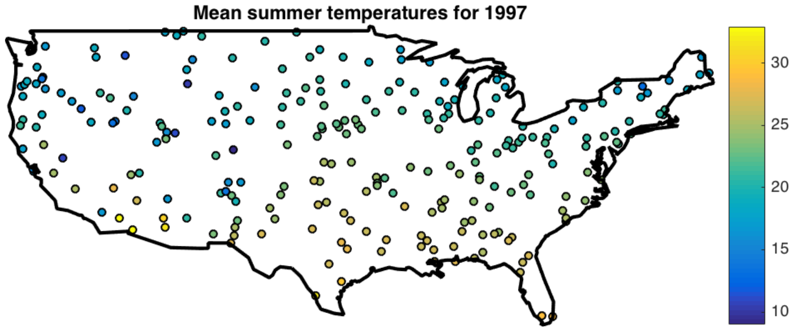
1. Obtain an initial estimate $\hat{\boldsymbol{\beta}}_{ols}$ of $\boldsymbol{\beta}$ using OLS.
2. Obtain a non-parametric estimate $\hat{\gamma}(h)$ of the semivariogram of the residuals $\mathbf{Y} - \mathbf{B}\hat{\boldsymbol{\beta}}_{ols}$.
3. Estimate the covariance parameters $\boldsymbol{\theta}$ by WLS estimation of the parametric semivariogram $\gamma(h; \boldsymbol{\theta})$ to $\hat{\gamma}(h)$.
4. Update the estimate of the regression parameters using GLS.

Code to simulate a Gaussian Matérn field

```
n = 50;
L = 1;
xx = linspace(0,L,n);
[XX, YY] = meshgrid(xx,xx);
loc = [XX(:) YY(:)];
D = squareform(pdist(loc));

rho = 0.3*L;
nu = 2;
kappa = sqrt(8*nu)/rho;
sigma = 2;
Sigma = matern_covariance(D,sigma,kappa,nu);
R = chol(Sigma);
x = R'*randn(n^2,1);
imagesc(reshape(x,[n n]))
```


Example: US temperatures

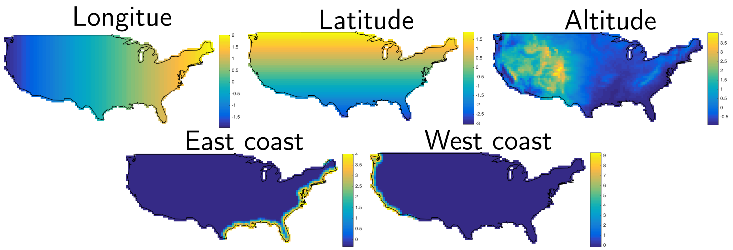


- Mean summer temperatures (June-August) in the continental US 1997 recorded at 250 weather stations.
- We want to estimate all US temperatures based on the data.

Model

$$Y(\mathbf{s}) = \sum_{i=1}^{k} \beta_i B_i(\mathbf{s}) + X(\mathbf{s}) + \varepsilon_{\mathbf{s}},$$

where $\varepsilon_{\mathbf{s}}$ are iid $\mathsf{N}(0, \sigma^2)$, $X(\mathbf{s})$ is a mean-zero Gaussian Matérn field, and we use the covariates



UNIVERSITY OF GOTHENBURG CHALMERS

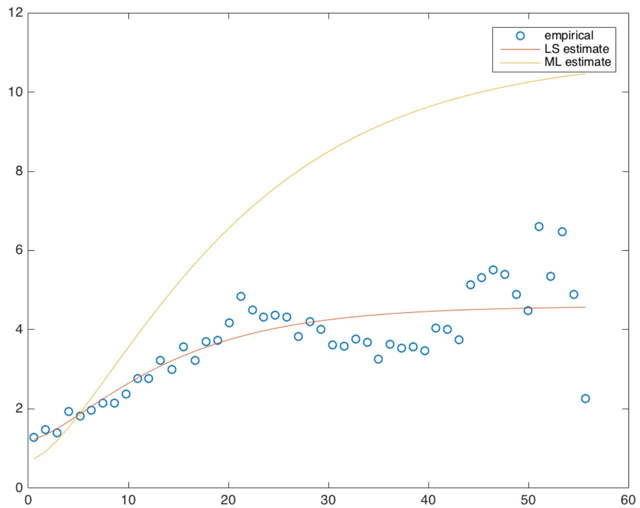
Estimates of regression parameters

	OLS	GLS	ML
Intercept	$21.6317*$	$20.4688*$	$19.5425*$
Longitude	$-1.2897*$	-1.0022	-0.4192
Latitude	$-2.6959*$	$-2.6845*$	$-2.6477*$
Altitude	$-2.6693*$	$-4.2177*$	$-4.3520*$
East coast	-0.0952	-0.0096	$0.0170*$
West coast	$-1.3064*$	$-1.0139*$	$-0.9261*$

Example David Bolin

UNIVERSITY OF GOTHENBURG CHALMERS

Likelihood-based analysis



Example David Bolin