Spatial Statistics with Image Analysis
Lecture 2

Johan Lindström

November 9, 2018

Lecture L02

Computer exercise 0

Stochastic fields
- Properties of Stochastic fields
- Stationarity
- Spectral Representation
- Example
- A Model for Spatial Data
- Multivariate Normal Distribution

Covariance functions
- Variogram
- The Matérn covariance function
- Estimation?

Computer exercise 0 – Daily Temperature

\[ y(t) = \beta_1 + \beta_2 \sin(2\pi t/365) + \beta_3 \cos(2\pi t/365) + \eta(t) \]
\[ = X(t)\beta + \eta(t) \]

\[ \eta(t) = \alpha \eta(t-1) + \nu(t), \quad \nu(t) \in \mathcal{N}(0, \sigma^2) \]
Computer exercise 0 – Prediction

Assuming that we have observed temperature up to time $t$ and know $\beta$ a prediction of the temperature at $t + \tau$ is:

$$E(y(t + \tau) | Y_{1:t}, \beta) = E(X(t + \tau) \beta + \eta(t + \tau) | Y_{1:t}, \beta) = X(t + \tau) \beta + \alpha^2 \eta(t),$$

with prediction variance:

$$V(y(t + \tau) | Y_{1:t}, \beta) = V(X(t + \tau) \beta + \alpha^2 \eta(t + \tau) | Y_{1:t}, \beta) = \sigma^2 \frac{1 - \alpha^2 \tau}{1 - \alpha^2}.$$
A stochastic field \( Y(s) \), \( s \in D \), is a random function defined on some index set \( D \).

- In image analysis and related applications, typically \( D \subseteq \mathbb{N}^2 \) or \( D \subseteq \mathbb{R}^2 \).
- The collection of all finite-dimensional distributions determines the field distribution:
  \[
p(Y(s_1), Y(s_2), \ldots, Y(s_n)), \quad s_1, \ldots, s_n \in D, \quad n \text{ finite}.
\]

Properties of Stochastic fields

- For a stochastic field \( Y(s) \), the expectation function
  \[
  \mu_Y(s) = E(Y(s))
  \]
  collects the point-wise expectations of the field.
- For the covariance between different locations, we write
  \[
  r(s_1, s_2) = C(Y(s_1), Y(s_2))
  = E \left( [Y(s_1) - \mu_Y(s_1)][Y(s_2) - \mu_Y(s_2)] \right)
  \]
  \( r(s_1, s_2) \) is called the covariance function.

Properties of Covariance functions

For a collection of finite-dimensional points, \( \{s_i\}_{i=1}^n \), the covariance function defines the covariance matrix as

\[
\Sigma = \begin{bmatrix}
  r(s_1, s_1) & r(s_1, s_2) & \cdots & r(s_1, s_n) \\
  r(s_2, s_1) & r(s_2, s_2) & \cdots & r(s_2, s_n) \\
  \vdots & \vdots & \ddots & \vdots \\
  r(s_n, s_1) & r(s_n, s_2) & \cdots & r(s_n, s_n)
\end{bmatrix}
\]

Properties of the covariance matrix translate to covariance functions:

- Symmetric: \( \Sigma = \Sigma^\top \iff r(s, t) = r(t, s) \)
- Pos. def. \( a^\top \Sigma a > 0 \iff \int_{\mathbb{R}^2} r(s, t)f(s)f(t) \, ds \, dt > 0 \)
  if \( a \neq 0 \) and \( f(s) \neq 0 \).
2nd order (weak) stationarity

A field is said to be 2nd order stationary if the expectation and covariance are unchanged under translation.

\[ \mu_Y(s) = \mu_Y(s + h) = \text{const.} \]

\[ r(s_1, s_2) = r(s_1 + h, s_2 + h) \Rightarrow r(s, s + h) = r(0, h) = r(h) \]

A stationary field is sometimes said to be **homogeneous**.

If a stationary covariance depends only on the distance between points, \( r(h) = r(||h||) \), the field is said to be **isotropic**.

An anisotropic field can be created from an isotropic covariance by a linear transformation of the coordinates:

\[ r(h) = r(||Ah||) \]

Strong stationarity

A field is said to be (strongly/strictly) stationary if all the finite-dimensional densities are unchanged under translation, i.e.

\[ p(Y(s_1), Y(s_2), \ldots, Y(s_n)) = p(Y(s_1 + h), Y(s_2 + h), \ldots, Y(s_n + h)) \]

for any \( h \).

- If \( V(Y(s)) < \infty \): strong stationarity \( \implies \) weak stationarity.
- For a Gaussian field:
  strong stationarity = weak stationarity.

Spectral Density

For a stochastic process in time, \( x(t) \) with stationary covariance function, \( r(t) \) the **spectral density** and covariance function form a **Fourier transform pair**

\[ f(\omega) = \frac{1}{2\pi} \int_{-\infty}^{\infty} r(t) e^{-i\omega t} dt = \mathcal{F}r \]

\[ r(t) = \int_{-\infty}^{\infty} f(\omega) e^{i\omega t} d\omega = \mathcal{F}^{-1}f \]
Spectral Density

For a stochastic field, \( x(s) \), in \( \mathbb{R}^d \) with **stationary covariance** function, \( r(h) \) the spectral density is

\[
f(\omega) = \frac{1}{(2\pi)^d} \int_{\mathbb{R}^d} r(h) e^{-i\omega h} \, dh,
\]

\[
r(h) = \int_{\mathbb{R}^d} f(\omega) e^{i\omega h} \, d\omega
\]

**Bochner's Theorem**

A real valued continuous function is **positive definite** iff it is the Fourier transformation of a **symmetric, non-negative measure** on \( \mathbb{R}^d \).

\[
r(h) \text{ is pos. def. } \iff \quad r(h) = \int_{\mathbb{R}^d} f(\omega) e^{i\omega h} \, d\omega
\]

\[
f(\omega) = f(-\omega), \quad f(\omega) \geq 0, \quad \int f(\omega) \, d\omega < \infty
\]

**Bessel functions**

\[
J_\alpha(x) = \sum_{m=0}^{\infty} \frac{(-1)^m}{m!(m+\alpha+1)} \left( \frac{x}{2} \right)^{2m+\alpha} \quad 1^{st} \text{ kind}
\]

\[
K_\nu(x) = \frac{(2x)^\nu}{\sqrt{\pi}} \Gamma(\nu + \frac{1}{2}) \int_0^{\infty} \frac{\cos(t)}{(t^2 + x^2)^{\nu+1/2}} \, dt \quad \text{modified 2}^{nd}
\]
Spectral Representation

For a stationary isotropic covariance function we have:

\[
\begin{align*}
    r(h) &= \int_{\mathbb{R}^d} f(\omega) e^{i \omega \cdot h} \, d\omega = \left[ \omega \cdot h = \omega h \cos(\theta) \right] = \\
    &= \int_0^\infty \omega^{d-1} f(\omega) \int_0^{\pi} \sin(\theta) e^{i \omega h \cos(\theta)} \, d\theta \, d\omega = \\
    &= \int_0^\infty f(\omega) \left( 2\pi \right)^{d/2} \omega^{d-1} J_{(d-2)/2}(\omega h) \frac{1}{(\omega h)^{(d-2)/2}} \, d\omega
\end{align*}
\]

where \( \omega = |\omega| \) and \( h = |h| \).

Spectral Density — Isotropy

For a stochastic field, \( x(s) \), in \( \mathbb{R}^d \) with isotropic stationary covariance function, \( r(h) \) the spectral density is

\[
    f(\omega) = \frac{1}{(2\pi)^{d/2}} \int_0^\infty r(h) h^{d-1} J_{(d-2)/2}(\omega h) \frac{1}{(\omega h)^{(d-2)/2}} \, dh
\]

\[
    r(h) = \int_0^\infty f(\omega) \omega^{d-1} J_{(d-2)/2}(\omega h) \frac{1}{(\omega h)^{(d-2)/2}} \, d\omega,
\]

where \( J_\alpha \) is a Bessel function of the first kind.

Example: The AR(1) process

In time an autoregressive (AR) model of order one is given by:

\[
    y(t) = \alpha y(t-1) + \nu(t), \quad \nu(t) \in \mathcal{N} \left( 0, \sigma^2 \right),
\]

with independent innovations \( \nu(t) \) and \( |\alpha| < 1 \)

This is a stationary process with

\[
    E(y(t)) = 0, \quad r(\tau) = C(y(t), y(t+\tau)) = \frac{\sigma^2}{1-\alpha^2} \delta(\tau), \quad V(y(t)) = r(0) = \frac{\sigma^2}{1-\alpha^2}
\]
Example: Estimating the covariance function

Given observations of a mean-zero stationary process an estimate of the covariance function can be obtained as

\[ \hat{r}(\tau) = \frac{1}{n \tau} \sum_{t} y(t) y(t + \tau) \approx E(y(t) y(t + \tau)) \]

A Model for Spatial Data

For modelling a stochastic process is often divided into

\[ y(s) = \mu(s) + \eta(s) + \varepsilon(s), \]

where \( \eta(s) \) is assumed to be stationary with one of the above covariance functions. \( \varepsilon(s) \) is called the nugget and represents small scale variability and measurement noise.

The resulting covariance function for \( Y(s) \) is:

\[
r_Y(h) = r_\eta(h) + \begin{cases} 
\sigma^2 + \sigma^2_\varepsilon, & \|h\| = 0, \\
\sigma^2_\eta(h), & \|h\| > 0.
\end{cases}
\]

Discrete representations

- If the true field is defined on (a subset of) \( \mathbb{R}^2 \), some form of discretisation is needed for practical computations.
- For a given set of points, \( \{s_1, \ldots, s_n\} \), the full field is represented by the random variables \( Y(s_i), i = 1, \ldots, n \).
- Almost all numerical calculations are performed for this discretely indexed field.
The Multivariate Normal distribution

The Gaussian (Normal) distribution will be used extensively.
- \( Y \in \mathcal{N}(\mu, \Sigma) \)
- The expectation is \( \mu \), with \( \mu_i = \mathbb{E}(Y(s_i)) = \mathbb{E}(Y_i) \).
- The covariance matrix is \( \Sigma \):
  \[
  \Sigma_{ij} = C(Y_i, Y_j), \quad \Sigma = C(Y, Y) = \mathbb{E}\left( [Y - \mu] [Y - \mu]^\top \right)
  \]
- The density is given by
  \[
  p(Y) = \frac{1}{(2\pi)^{N/2} |\Sigma|^{1/2}} \exp\left( -\frac{1}{2} (Y - \mu)^\top \Sigma^{-1} (Y - \mu) \right)
  \]
  where \( Y \) and \( \mu \) are column vectors of length \( N \), and \( \Sigma \) is a \( N \times N \)-matrix.

Covariance functions

- We often need to estimate the \( N \times N \) covariance matrix \( \Sigma \) based on one sample (\( N \) observations) of the field.
- This implies estimating \( N(N + 1)/2 \) unknowns from \( N \) observations...
- The requirements (positive definite) on \( \Sigma \) implies restrictions on the estimation.
- The covariance matrix is often assumed to come from a parametric family of covariance functions. Reducing the problem to estimation of the covariance parameters.
- The resulting estimation problem does not have a closed form solution, leading to numerical optimisation.

Matérn:

\[
 r(h) = \frac{\sigma^2}{\Gamma(\nu) 2^{\nu-1}} (\kappa ||h||)^{\nu} K_\nu(\kappa ||h||),
\]

Exponential: Matérn with \( \nu = 1/2 \)

\[
 r(h) = \sigma^2 \exp(-\kappa ||h||)
\]

Gaussian: Matérn with \( \nu \to \infty \)

\[
 r(h) = \sigma^2 \exp(-2||h||^2/\rho^2)
\]

Cauchy:

\[
 r(h) = \frac{\sigma^2}{1 + (||h||/\rho)^2}^\pi
\]

Spherical:

\[
 r(h) = \sigma^2 \begin{cases} 
 1 - 1.5(||h||/\rho) + 0.5(||h||/\rho)^2, & ||h|| \leq \rho \\
 0, & ||h|| > \rho 
\end{cases}
\]
Examples of Covariance functions

- Spherical
- Gaussian
- Exponential
- Matern, $k=1$
- Matern, $k=3$
- Cauchy, $k=1$
- Cauchy, $k=5$

Semi-variogram

For a stationary, isotropic field the semi-variogram is defined as

$$\gamma(h) = \frac{1}{2} V(Y(s + h) - Y(s)) = r(0) - r(h)$$

The link between variogram and covariance function is:

$$\gamma(h) = \sigma^2 + \sigma^2_z - r_0(|h|) \quad \rightarrow \sigma^2 + \sigma^2_z, \quad \text{as } |h| \rightarrow \infty$$

$$r(h) = r_0(h) + \mathbb{I}(|h| = 0)\sigma^2_z \quad \rightarrow \sigma^2 + \sigma^2_z, \quad \text{as } |h| \rightarrow 0$$

with nugget $\sigma^2_z$, partial sill $\sigma^2$, and sill $\sigma^2 + \sigma^2_z$. 
Matérn covariance

\[ r_M(h) = \frac{\sigma^2}{\Gamma(\nu) 2^{\nu-1}} (\kappa \|h\|)^\nu K_\nu(\kappa \|h\|), \quad h \in \mathbb{R}^d, \]

where \( K_\nu \) is a modified Bessel function of the second kind.

Parameters, \( \theta \), of the covariance are:
- **variance** \( (\sigma^2 \geq 0) \),
- **scale** \( (\kappa > 0) \),
- **shape** \( (\nu > 0) \).

A measure of the **range** is given by \( \rho = \sqrt{8\nu}/\kappa \).
If $\theta = \{\mu, \Sigma\}$ are known, how can we predict $Y$ at unobserved locations?
- If $\mu$ is unknown?
- If $\Sigma$ in unknown?