Chapter 4 - Fundamentals of spatial processes
Lecture notes, part 2, Theory

Odd Kolbjørnsen and Geir Storvik

February 6, 2017

Last time, spatial process

- Spatial correlation
  - Stationary
  - Isotropic/Anisotropic
  - Permutation test for independence
- Variogram
  - Link to covariance
  - Nugget effect
  - Sill
- Spatial prediction
  - Interpolation
  - Minimum Mean Squared Prediction Error (Kriging)
  - Prediction in a Gaussian model
- Kriging
  - Simple Kriging (known mean)
  - Bayesian Kriging (prior on mean)
  - Universal Kriging (unknown mean)
  - Ordenary Kriging (Special case of UK)

Plan ahead

- To day computations
  - General set up for "Kriging type problems"
    - Introductory example
    - Kriging case
    - General case
  - Change of support
  - Spatial moving average models
    - Construction
    - Correlation function
  - Non gaussian observations
    - Monte Carlo
    - Laplace approximation
    - INLA
- Next time how to do it with a computer

General setup for Kriging type problems

- What is the process vs data relation ?
- Set up the prior for the process model.
- Do the relevant computations to derive joint distribution/second order moments

Recall:

$$\begin{pmatrix} \mathbf{X}_1 \\ \mathbf{X}_2 \end{pmatrix} \sim MVN \left(\begin{pmatrix} \boldsymbol{\mu}_1 \\ \boldsymbol{\mu}_2 \end{pmatrix}, \begin{pmatrix} \boldsymbol{\Sigma}_{11} & \boldsymbol{\Sigma}_{12} \\ \boldsymbol{\Sigma}_{21} & \boldsymbol{\Sigma}_{22} \end{pmatrix} \right)$$

$$\begin{aligned} E(\mathbf{X}_1|\mathbf{X}_2) =& \boldsymbol{\mu}_1 + \boldsymbol{\Sigma}_{12}\boldsymbol{\Sigma}_{22}^{-1}(\mathbf{X}_2 - \boldsymbol{\mu}_2) \\ \text{var}(\mathbf{X}_1|\mathbf{X}_2) =& \boldsymbol{\Sigma}_{11} - \boldsymbol{\Sigma}_{12}\boldsymbol{\Sigma}_{22}^{-1}\boldsymbol{\Sigma}_{21} \end{aligned}$$

A standard problem

$$Z = aY + \varepsilon, \ \varepsilon \sim N(0, \sigma^2)$$

$$Y \sim N(\mu, \tau^2)$$

$$\begin{pmatrix} Y \\ Z \end{pmatrix} \sim MVN\left(\begin{pmatrix} \mu \\ a\mu \end{pmatrix}, \begin{pmatrix} \tau^2 & a\tau^2 \\ a\tau^2 & a^2\tau^2 + \sigma^2 \end{pmatrix}\right)$$

$$E(Y|Z) = \mu + \frac{a\tau^2}{a^2\tau^2 + \sigma^2}(Z - a\mu) = \mu\left(\frac{\sigma^2}{a^2\tau^2 + \sigma^2}\right) + \frac{Z}{a}\left(\frac{a^2\tau^2}{a^2\tau^2 + \sigma^2}\right)$$

$$\text{var}(Y|Z) = \tau^2 - \tau^2\left(\frac{a^2\tau^2}{a^2\tau^2 + \sigma^2}\right)$$

Simple Kriging

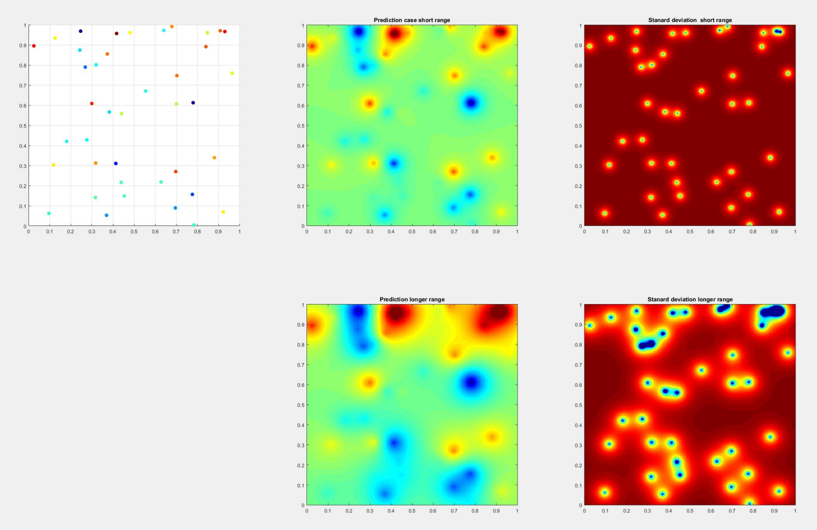
$$Z_i = Y(\boldsymbol{s}_i) + \varepsilon_i,\ \varepsilon_i \sim N(0, \sigma^2), iid$$

$$Y(\boldsymbol{s}) \sim N(\mu(\boldsymbol{s}), C_Y(\boldsymbol{s}_1, \boldsymbol{s}_2))$$

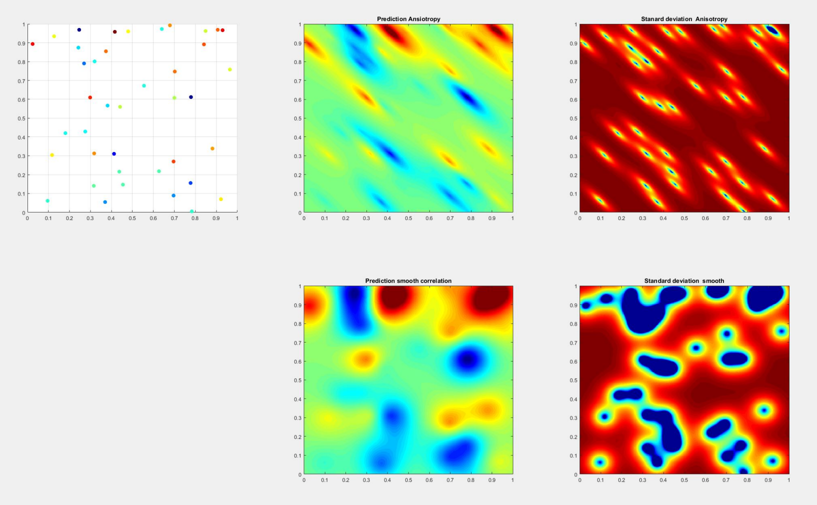
$$\begin{pmatrix} Y(\boldsymbol{s}_0) \\ \boldsymbol{Z} \end{pmatrix} \sim MVN\left(\begin{pmatrix} \mu(\boldsymbol{s}_0) \\ \boldsymbol{\mu}(''\boldsymbol{s}_i'') \end{pmatrix}, \begin{pmatrix} C_Y(\boldsymbol{s}_0, \boldsymbol{s}_0) & C_Y(\boldsymbol{s}_0, ''\boldsymbol{s}_i'') \\ C_Y(''\boldsymbol{s}_i'', \boldsymbol{s}_0) & C_Y(''\boldsymbol{s}_i'', ''\boldsymbol{s}_i'') + \sigma^2 \boldsymbol{I} \end{pmatrix}\right)$$

$$\begin{aligned} E(Y(s_0)|\boldsymbol{Z}) =& \mu(s_0) + C_Y(\boldsymbol{s}_0, ''\boldsymbol{s}_i'')(C_Y(''\boldsymbol{s}_i'', ''\boldsymbol{s}_i'') + \sigma^2 \boldsymbol{I})^{-1}(Z - \mu(''\boldsymbol{s}_i'')) \\ \text{var}(Y(\boldsymbol{s}_0)|Z) =& C_Y(\boldsymbol{s}_0, \boldsymbol{s}_0) \\ & - C_Y(\boldsymbol{s}_0, ''\boldsymbol{s}_i'')(C_Y(''\boldsymbol{s}_i'', ''\boldsymbol{s}_i'') + \sigma^2 \boldsymbol{I})^{-1} C_Y(''\boldsymbol{s}_i'', \boldsymbol{s}_0) \end{aligned}$$

Kriging example



Kriging example



General Case

$$\boldsymbol{Z} = \boldsymbol{G}\boldsymbol{Y} + \boldsymbol{\varepsilon},\ \varepsilon_i \sim N(0, \sigma^2), iid$$

$$Y(\boldsymbol{s}) \sim N(\mu(\boldsymbol{s}), C_Y(\boldsymbol{s}_1, \boldsymbol{s}_2)) = N(\boldsymbol{\mu}_Y, \boldsymbol{C}_Y)$$

$$\begin{pmatrix}\boldsymbol{Y}\\ \boldsymbol{Z}\end{pmatrix} \sim MVN\left(\begin{pmatrix}\boldsymbol{\mu}_Y\\ \boldsymbol{G}\boldsymbol{\mu}_Y\end{pmatrix}, \begin{pmatrix}\boldsymbol{C}_Y & \boldsymbol{G}\boldsymbol{C}_Y\\ \boldsymbol{C}_Y\boldsymbol{G}^T & \boldsymbol{G}\boldsymbol{C}_Y\boldsymbol{G}^T + \sigma^2\boldsymbol{I}\end{pmatrix}\right)$$

$$E(\boldsymbol{Y}|\boldsymbol{Z}) = \boldsymbol{\mu}_Y + \boldsymbol{C}_Y\boldsymbol{G}^T(\boldsymbol{G}\boldsymbol{C}_Y\boldsymbol{G}^T + \sigma^2\boldsymbol{I})^{-1}(Z - \boldsymbol{G}\boldsymbol{\mu}_Y)$$

$$\text{var}(\boldsymbol{Y}|\boldsymbol{Z}) = C_Y - \boldsymbol{C}_Y\boldsymbol{G}^T(\boldsymbol{G}\boldsymbol{C}_Y\boldsymbol{G}^T + \sigma^2\boldsymbol{I})^{-1}\boldsymbol{G}\boldsymbol{C}_Y$$

Change of support, Area support on Data

$$
\begin{aligned}
Z_i &= \int_{A_i} Y(s)d\mathbf{s} + \varepsilon_i,\ \varepsilon_i \sim N(0,\sigma^2), iid \\
\mathbf{G}^T &= [g_1^T, ..., g_n^T],\ \text{with } g_i = I(\mathbf{s} \in A_i)
\end{aligned}
$$

$$
\begin{aligned}
\text{E}(Z_i) &= \int_{A_i} \mu_Y(\mathbf{s})\, d\mathbf{s} \\
\text{Cov}(Y(\mathbf{s}_0), Z_i) &= \int_{A_i} C_Y(\mathbf{s}_0, \mathbf{s})\, d\mathbf{s} \\
\text{Cov}(Z_i, Z_j) &= \int_{A_i}\int_{A_j} C_Y(\mathbf{s}_1, \mathbf{s}_2)\, d\mathbf{s}_1 d\mathbf{s}_2 + \sigma^2 I(i = j)
\end{aligned}
$$

Change of support,Prediction of a a region

$$V = \int_A Y(\mathbf{s})d\mathbf{s}$$

$$E(V|\mathbf{Z}) = \int_A E(Y(\mathbf{s})|\mathbf{Z})d\mathbf{s}$$

$$Var(V|\mathbf{Z}) = \int_A \int_A \text{Cov}(Y(\mathbf{s}_1), Y(\mathbf{s}_2)|\mathbf{Z})\, d\mathbf{s}_1 d\mathbf{s}_2$$

Spatial moving average models

Multivariate setting:

$$\boldsymbol{Y} = \boldsymbol{\mu} + \boldsymbol{K}\boldsymbol{W}, W_i \sim N(0,1), iid$$

$$\boldsymbol{Y} \sim N(\boldsymbol{\mu}, \boldsymbol{K}\boldsymbol{K}^T)$$

Continious random field: Moving average kernel: $k(\boldsymbol{s}, \boldsymbol{u}) : R^d \times R^d \to R$

$$Y(\boldsymbol{s}) = \mu(\boldsymbol{s}) + \int_{R^d} k(\boldsymbol{s}, \boldsymbol{u}) W(d\boldsymbol{u}), \text{ where } W(d\boldsymbol{u}) \sim N(0, d\boldsymbol{u}), iid$$

$$\boldsymbol{Y} \sim N(\mu(\boldsymbol{s}), C_Y(\boldsymbol{s}_1, \boldsymbol{s}_2))$$

$$C_Y(\boldsymbol{s}_1, \boldsymbol{s}_2) = \int_{R^d} k(\boldsymbol{s}_1, \boldsymbol{u}) k(\boldsymbol{s}_2, \boldsymbol{u}) d\boldsymbol{u}$$

Note: distinct moving average kernes may give the same correlation function.

Spectral expansion

Eigen-representation of moving average kernel:

$$k(\boldsymbol{s}, \boldsymbol{u}) = \sum_{i=0}^{\infty} \lambda_i \phi_i(\boldsymbol{s}) \phi_i(\boldsymbol{u})$$

$\Rightarrow$ Spectral representation random field:

$$Y(\boldsymbol{s}) = \sum_{i=0}^{\infty} y_i \phi_i(\boldsymbol{s}), \text{ with } y_i \sim N(0, \lambda_i^2), \textit{independent}$$

Spectral representation korrelation function (Karhunen-Loéve expansion):

$$C_Y(\boldsymbol{s}, \boldsymbol{u}) = \sum_{i=0}^{\infty} \lambda_i^2 \phi_i(\boldsymbol{s}) \phi_i(\boldsymbol{u}) \quad (\Sigma = \boldsymbol{V} \boldsymbol{\Lambda}^2 \boldsymbol{V}^T)$$

Common to use Fourier transform, Fast Fourier Transform (FFT)

Estimation

Likelihood:

$$L(\theta) = p(\boldsymbol{Z};\theta) = \int_{\boldsymbol{Y}} p(\boldsymbol{Z}|\boldsymbol{Y}\boldsymbol{\theta})\, d\boldsymbol{Y}$$

Gaussian process: can be derived analytically

- Optimization can still be problematic
- Many routines in R available

Estimation

Likelihood:

$$L(\theta) = p(\boldsymbol{Z};\theta) = \int_{\boldsymbol{Y}} p(\boldsymbol{Z}|\boldsymbol{Y}\boldsymbol{\theta})\,d\boldsymbol{Y}$$

Gaussian process: can be derived analytically

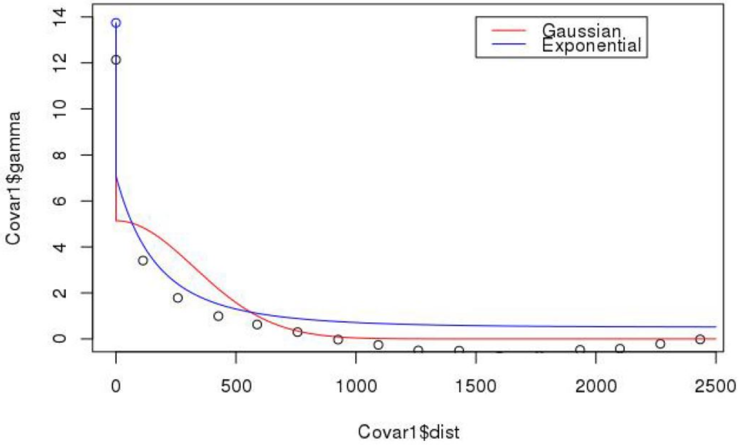
- Optimization can still be problematic
- Many routines in R available

```
> gls(Bor~Wet, correlation=corGaus(form=~x+y, nugget=TRUE), data=Boreality)
Generalized least squares fit by REML
  Model: Bor ~ Wet
  Data: Boreality
  Log-restricted-likelihood: -1363.146

Coefficients:
(Intercept)         Wet
   15.06705    77.66940

Correlation Structure: Gaussian spatial correlation
 Formula: ~x + y
 Parameter estimate(s):
      range      nugget
460.3941829   0.6112509
Degrees of freedom: 533 total; 531 residual
Residual standard error: 3.636527
```

Boreality - fitted covariance function



Hierarchical and Bayesian model

Hierarchical model

	Variable	Densities	Notation in book
Data model:	$\mathbf{Z}$	$p(\mathbf{Z}\|\mathbf{Y},\boldsymbol{\theta})$	$[\mathbf{Z}\|\mathbf{Y},\boldsymbol{\theta}]$
Process model:	$\mathbf{Y}$	$p(\mathbf{Y}\|\boldsymbol{\theta})$	$[\mathbf{Y}\|\boldsymbol{\theta}]$ (Gaussian in 4.1)
Parameter:	$\boldsymbol{\theta}$		

Hierarchical and Bayesian model

Hierarchical model

	Variable	Densities	Notation in book
Data model:	$\mathbf{Z}$	$p(\mathbf{Z}\vert\mathbf{Y},\boldsymbol{\theta})$	$[\mathbf{Z}\vert\mathbf{Y},\boldsymbol{\theta}]$
Process model:	$\mathbf{Y}$	$p(\mathbf{Y}\vert\boldsymbol{\theta})$	$[\mathbf{Y}\vert\boldsymbol{\theta}]$ (Gaussian in 4.1)
Parameter:	$\boldsymbol{\theta}$		

Simultaneous model: $p(\mathbf{y},\mathbf{z}|\boldsymbol{\theta}) = p(\mathbf{z}|\mathbf{y},\boldsymbol{\theta})p(\mathbf{y}|\boldsymbol{\theta})$
Marginal model: $L(\boldsymbol{\theta}) = p(\mathbf{z}|\boldsymbol{\theta}) = \int_{\mathbf{y}} p(\mathbf{z},\mathbf{y}|\boldsymbol{\theta})d\mathbf{y}$

Inference: $\widehat{\boldsymbol{\theta}} = \text{argmax}_{\boldsymbol{\theta}} L(\boldsymbol{\theta})$

Hierarchical and Bayesian model

Hierarchical model

	Variable	Densities	Notation in book
Data model:	$\mathbf{Z}$	$p(\mathbf{Z}\|\mathbf{Y},\boldsymbol{\theta})$	$[\mathbf{Z}\|\mathbf{Y},\boldsymbol{\theta}]$
Process model:	$\mathbf{Y}$	$p(\mathbf{Y}\|\boldsymbol{\theta})$	$[\mathbf{Y}\|\boldsymbol{\theta}]$ (Gaussian in 4.1)
Parameter:	$\boldsymbol{\theta}$		

Simultaneous model: $p(\mathbf{y},\mathbf{z}|\boldsymbol{\theta}) = p(\mathbf{z}|\mathbf{y},\boldsymbol{\theta})p(\mathbf{y}|\boldsymbol{\theta})$
Marginal model: $L(\boldsymbol{\theta}) = p(\mathbf{z}|\boldsymbol{\theta}) = \int_{\mathbf{y}} p(\mathbf{z},\mathbf{y}|\boldsymbol{\theta})d\mathbf{y}$
Inference: $\widehat{\boldsymbol{\theta}} = \text{argmax}_{\boldsymbol{\theta}} L(\boldsymbol{\theta})$

Bayesian approach: Include model on $\boldsymbol{\theta}$

	Variable	Densities	Notation in book
Data model:	$\mathbf{Z}$	$p(\mathbf{Z}\|\mathbf{Y},\boldsymbol{\theta})$	$[\mathbf{Z}\|\mathbf{Y},\boldsymbol{\theta}]$
Process model:	$\mathbf{Y}$	$p(\mathbf{Y}\|\boldsymbol{\theta})$	$[\mathbf{Y}\|\boldsymbol{\theta}]$ (Gaussian in 4.1)
Parameter model:	$\boldsymbol{\theta}$	$p(\boldsymbol{\theta})$	$[\boldsymbol{\theta}]$

Simultaneous model: $p(\mathbf{y},\mathbf{z},\boldsymbol{\theta})$
Marginal model: $p(\mathbf{z}) = \int_{\boldsymbol{\theta}}\int_{\mathbf{y}} p(\mathbf{z},\mathbf{y}|\boldsymbol{\theta})d\mathbf{y}d\boldsymbol{\theta}$
Inference: $\widehat{\boldsymbol{\theta}} = \int_{\boldsymbol{\theta}} \boldsymbol{\theta} p(\boldsymbol{\theta}|\mathbf{z})d\boldsymbol{\theta}$

Non-Gaussian data

- Counts, binary data: Gaussian assumption inappropriate
- Can still have Gaussian assumption on *latent* process, but non-Gaussian data-distribution

$$Y(\mathbf{s}) = \mathbf{x}(\mathbf{s})^T\boldsymbol{\beta} + \varepsilon(\mathbf{s}), \{\varepsilon(\mathbf{s})\} \text{ Gaussian process}$$
$$Z(\mathbf{s}_i)|Y(\mathbf{s}_i), \theta_1 \sim \text{ind.f}(Y(\mathbf{s}_i), \theta_1)$$

- Best linear predictor still possible, but is it reasonable ?
- Conditional expectation $E[Y(\mathbf{s}_0)|\mathbf{Z}]$ still optimal under square loss
  Not easy to compute anymore

Exponential-family model (EFM)

$$f(z) = \exp\{(z\eta - b(\eta))/a(\theta_1) + c(z, \theta_1)\}$$
$$\eta = \mathbf{x}^T\boldsymbol{\beta}$$

Include Binomial, Poisson, Gaussian, Gamma

Hierarcical model ans non-Gaussian observations

Likelihood:

$$L(\theta) = p(\boldsymbol{Z};\theta) = \int_{\boldsymbol{Y}} p(\boldsymbol{Z}|\boldsymbol{Y}\boldsymbol{\theta})\, d\boldsymbol{Y}$$

$E[Y(\mathbf{s}_0)|\mathbf{Z}] = ?$

- Plug in / Empirical bayes. $\hat{\theta}$- Maximumlikelihood. $E[Y(\mathbf{s}_0)|\mathbf{Z},\hat{\theta}]$
- Monte Carlo: Sample a "subset of variables" to get full distribution
- Laplace approximation: Approximate $[\mathbf{Y}|\mathbf{Z}]$ by a Gaussian distribution
- Computer software available for both approahces.
- Also give uncertainty measures var$[Y(\mathbf{s}_0)|\mathbf{Z}]$.

Monte Carlo approximation of conditional expectation

$$Z_i \sim f(z|Y(\boldsymbol{s}_i), \theta), \text{ independent }, i = 1, ..., n$$
$$Y(\boldsymbol{s}) \sim N(\mu(\boldsymbol{s}), C_Y(\boldsymbol{s}_1, \boldsymbol{s}_2))$$

If we only knew $Y(\boldsymbol{s}_i)$ i=1,...,m, and θ.
Then everything would be linear and nice.
Solution: Sample $\boldsymbol{Y}$, and θ using Monte Carlo techniqies:
McMC, Rejection sampling, SIR ...

Obtain samples:
$\{\boldsymbol{Y}, \theta\}^m$, m=1,...M from $p(\boldsymbol{Y}, \theta|\boldsymbol{Z})$.

Then compute:

$$E[E[Y(\mathbf{s}_0)|\mathbf{Y}, \mathbf{Z}]] \approx \frac{1}{M}\sum_{m=1}^{M} E[Y(\mathbf{s}_0)|\mathbf{Y}^m] \text{ or } E[\theta|\boldsymbol{Z}] \approx \frac{1}{M}\sum_{m=1}^{M}\theta^m$$

Laplace approximation

$$L(\boldsymbol{\theta}) = p(\mathbf{z};\boldsymbol{\theta}) = \int_{\mathbf{y}} p(\mathbf{z}|\mathbf{y};\boldsymbol{\theta})p(\mathbf{y};\boldsymbol{\theta})d\mathbf{y} = \int_{\mathbf{y}} \exp\{f(\mathbf{y};\boldsymbol{\theta})\}d\mathbf{y}$$

where

$$f(\mathbf{y};\boldsymbol{\theta}) = \log p(\mathbf{z}|\mathbf{y};\boldsymbol{\theta}) + \log p(\mathbf{y};\boldsymbol{\theta})$$

Taylor approximation, $\mathbf{y}_0$ max-point of $f(\mathbf{y};\boldsymbol{\theta})$ (depending on $\boldsymbol{\theta}$)

$$\begin{aligned}
f(\mathbf{y}) \approx & f(\mathbf{y}_0) - \frac{1}{2}(\mathbf{y}-\mathbf{y}_0)^T\mathbf{H}(\mathbf{y}-\mathbf{y}_0), \quad \mathbf{H} = -\frac{\partial^2}{\partial\mathbf{y}\partial\mathbf{y}^T}f(\mathbf{y})|_{\mathbf{y}=\mathbf{y}_0} \\
L(\boldsymbol{\theta}) \approx & \int_{\mathbf{y}} \exp\{f(\mathbf{y}_0) - \frac{1}{2}(\mathbf{y}-\mathbf{y}_0)^T\mathbf{H}(\mathbf{y}-\mathbf{y}_0)\}d\mathbf{y} \\
= & \exp\{f(\mathbf{y}_0)\}(2\pi)^{m/2}|\mathbf{H}|^{1/2}
\end{aligned}$$

Integrated nested Laplace approximations

- Difficult to compute:

$$L(\boldsymbol{\theta}|\boldsymbol{Z}) = p(\boldsymbol{Z};\boldsymbol{\theta}) = \int_{\boldsymbol{Y}} p(\boldsymbol{Z}|\boldsymbol{Y}\boldsymbol{\theta})\,d\boldsymbol{Y}$$

- Note:

$$p(\boldsymbol{Y}|\boldsymbol{\theta},\boldsymbol{Z}) = \frac{p(\boldsymbol{Y},\boldsymbol{\theta},\boldsymbol{Z})}{p(\boldsymbol{\theta},\boldsymbol{Z})} = \frac{p(\boldsymbol{Y},\boldsymbol{\theta},\boldsymbol{Z})}{p(\boldsymbol{\theta}|\boldsymbol{Z})p(\boldsymbol{Z})}$$

- Flip the above expression to get for any $\boldsymbol{Y}$

$$p(\boldsymbol{\theta}|\boldsymbol{Z}) = \frac{p(\boldsymbol{Y},\boldsymbol{\theta},\boldsymbol{Z})}{p(\boldsymbol{Y}|\boldsymbol{\theta},\boldsymbol{Z})p(\boldsymbol{Z})} \propto \frac{p(\boldsymbol{\theta})p(\boldsymbol{Y}|\boldsymbol{\theta})p(\boldsymbol{Z}|\boldsymbol{Y},\boldsymbol{\theta})}{p(\boldsymbol{Y}|\boldsymbol{\theta},\boldsymbol{Z})}$$

- INLA: $p(\boldsymbol{Y}|\boldsymbol{\theta},\boldsymbol{Z})$ is approximated by a Gaussian distribution. (Laplace appoximation) See paper on homepage.

Extra

Ordinary kriging, derivation

Ordinary kriging, $EY(\mathbf{s}) = \mu$, unknown (special case of UK)

$$\widehat{Y}(\mathbf{s}_0) = \{\mathbf{c}_Y(\mathbf{s}_0) + \frac{\mathbf{1}(1 - \mathbf{1}^T\mathbf{C}_Z{-}1\mathbf{c}_y(\mathbf{s}_0))}{\mathbf{1}^T\mathbf{C}_Z^{-1}\mathbf{1}}\}^T\mathbf{C}_Z^{-1}\mathbf{Z}$$

$$= \widehat{\mu}_{gls} + \mathbf{c}_Y(\mathbf{s}_0)^T\mathbf{C}_Z^{-1}(\mathbf{Z} - \mathbf{1}\widehat{\mu}_{gls})$$

$$\widehat{\mu}_{gls} = [\mathbf{1}^T\mathbf{C}_Z^{-1}\mathbf{1}]^{-1}\mathbf{1}^T\mathbf{C}_z^{-1}\mathbf{Z}$$

To minimize the MSPE, we make an unbiased estimate with the minimum variance.

$$\text{MSPE}(\boldsymbol{\lambda}) = E(Y(\mathbf{s}_0) - \boldsymbol{\lambda}^T\mathbf{Z})^2$$

Unbiased constraint:

$$E[Y(\mathbf{s}_0)] - E[\boldsymbol{\lambda}^T\mathbf{Z}] = 0$$

Prediction variance:

$$PV(\boldsymbol{\lambda}) = C_Y(\mathbf{s}_0, \mathbf{s}_0) - 2\boldsymbol{\lambda}^T c_Y(\mathbf{s}_0) + \boldsymbol{\lambda}^T\mathbf{C}_Z\boldsymbol{\lambda}$$

Ordinary kriging equations, cont

We have: $E[Y(\mathbf{s}_0)] = \mu$ and $E[\boldsymbol{\lambda}^T\mathbf{Z}] = \boldsymbol{\lambda}^T E[\mathbf{Z}] = \boldsymbol{\lambda}^T\mathbf{1}\mu$ thus

$$\mu = \boldsymbol{\lambda}^T\mathbf{1}\mu \Rightarrow 1 = \lambda^T\mathbf{1}$$

The problem then becomes:

$$\min_{\lambda} PV(\boldsymbol{\lambda})$$

subject to:

$$1 = \lambda^T\mathbf{1}$$

Solved by Lagrange multiplier, i.e. minimize

$$C_Y(\mathbf{s}_0, \mathbf{s}_0) - 2\boldsymbol{\lambda}^T\mathbf{c}_Y(\mathbf{s}_0) + \boldsymbol{\lambda}^T\mathbf{C}_Z\boldsymbol{\lambda} - 2\kappa(\boldsymbol{\lambda}^T\mathbf{1} - 1)$$

Differentiation wrt $\boldsymbol{\lambda}^T$ gives (which is combined to the final result presented on previous page):

$$\begin{aligned} -2c_Y(\mathbf{s}_0) + 2\mathbf{C}_Z\boldsymbol{\lambda} &= 2\kappa\mathbf{1} \\ \boldsymbol{\lambda}^T\mathbf{1} &= 1 \end{aligned}$$