

Bachelor project

Filip Schjerven

Spring 2017

Contents

I	main	2
0.1	Layout	3
0.1.1	Hello	3
0.2	Theory	3
0.2.1	Spatial prediction in general	3
0.3	Case study	3
0.3.1	Case introduction	3
0.3.2	Case motivation	3
0.3.3	Case theory	4
0.3.4	Case main	5
0.3.5	Case discussion	5
0.3.6	Case code	5
0.4	Plan	5

Chapter 1

main

1.1 Layout

1.1.1 Hello

- Intro
- Theory
- Case
- Code

1.2 Theory

1.2.1 Spatial prediction in general

- data model, process model, parameter model (whats this called?)
- Two of many ways to model: Proactive and retroactive / bayesian and frequentist, difference in model assumptions and estimating parameters
- Proactive: Bayesian with prior, distribution for all parameters, hyperpriors, posteriors, issue with uncommon distributions when using mixed models. Advantage: Able to produce predictions without samples Disadvantages: Relies on a good model, parameter estimates may be sensitive to choice of model thus needing expertise, needs heavy computations for complex models
- Retroactive: Estimating the parameters of the model based on samples collected, typical MLE, other ways of doing it (e.g. method of moments). Advantage: relatively easy. Disadvantages: Needs to have obtained samples, no samples no predictions
- other differences between frequentist and bayesian way of modelling exists, but general
- write generally the distributions in a bayesian setting, retrospective used to estimate parameters -i not bayesian in that sense. integration to get the distribution that we're interested in

- correlation functions, what is available and how do they work

Correlation in spatial prediction: defines a discretization of the field as a collection of random variables, isotropic (only euclidean distance) correlation function, many to choose, used matern with 3/2 dof, matern is smooth.

1.3 Case study

1.3.1 Case introduction

temperature data, interpreted in a spatial setting, independent of time

1.3.2 Case motivation

minimize variance of prediction

1.3.3 Case theory

- notation
- Model build up and motivation of choice
- what this means in terms of expressions for all parameters
- what does this mean in relation to the goal of the case study
- proofs of theorems used?
- correlation function used, motivation for using this

For the case study of this project, several models have been used to perform predictions. However, the field that is to be predicted has been modelled as a GRF with Gaussian data error and priors in the exponential family, that is:

1. Data model: $[Z(\vec{s})|Y(\vec{s}_0)] = F \cdot Y(\vec{s}|\vec{s}_0) + \epsilon$, where ϵ is Gaussian with mean 0 and variance σ_ϵ^2
2. Process model: Y is a GRF with $[Y(\vec{s}_0)|\sigma_r^2, \tau] = X(\vec{s}_0)\vec{\beta} + C_Y(\vec{s}_0, \sigma_r^2, \tau)$
3. Parameter model: $\sigma_{range}^2 \sim Gamma(\alpha, \kappa), \tau \sim Exp(\lambda)$

The trend of the GRF model is not known, and will have to be estimated somehow. In this case, this has been done by performing an Generalized Least Squares estimate of the sampled data $Z(\vec{s}_0)$. In large enough samples, which are relatively small considering the size of the field, the distribution of the GLS estimate, $\hat{\beta}$, converges to a Gaussian with mean β and covariance $(X(\vec{s}_0)^T C_Y^{-1} X(\vec{s}_0))^{-1}$, where C_Y denotes the correlation matrix between sampled areas. Combining this result with $Y(\vec{x})$ being assumed a GRF, we may rewrite our model with their full distributions:

1. Data model: $Z(\vec{s})|Y(\vec{s}_0)$ is Gaussian with mean $F \cdot X(\vec{s})\hat{\beta}$ and variance $F(X(\vec{s})(X(\vec{s}_0)^T C_Y^{-1} X(\vec{s}_0))^{-1} X(\vec{s})^T + C_Y(\vec{s}, \sigma_r^2, \tau))F^T + \sigma_\epsilon^2$

2. Process model: Y is a GRF $[Y(\vec{s}_0)|\sigma_r^2, \tau] = X(\vec{s}_0)\hat{\beta} + C_Y(\vec{s}_0, \sigma_r^2, \tau)$
3. Parameter model: $\sigma_{range}^2 \sim \text{Gamma}(\alpha, \kappa), \tau \sim \text{Exp}(\lambda)$

The distribution of interest is that of $[Y(\vec{s})|Z(\vec{s}_0)]$, i.e. the underlying process data at locations \vec{s} when having observed noisy data at locations \vec{s}_0 . As both the data model and the process model is Gaussian with assumed known parameters, there is an exact analytical result for the ensuing distribution. First, the joint distribution of $\begin{bmatrix} Y(\vec{s}) \\ Z(\vec{s}_0) \end{bmatrix}$ is known bivariate Gaussian as both $Y(\vec{s})$ and $Z(\vec{s}_0)$ are Gaussian. Denote the expected values as $\begin{bmatrix} \mu_{Y(\vec{s})} \\ \mu_{Z(\vec{s}_0)} \end{bmatrix}$ and their variance matrix as $\begin{bmatrix} \Sigma_{Y(\vec{s})} & \Sigma_{Y(\vec{s}), Z(\vec{s}_0)} \\ \Sigma_{Z(\vec{s}_0), Y(\vec{s})} & \Sigma_{Z(\vec{s}_0)} \end{bmatrix}$. Then, from ??, we have that $[Y(\vec{s})|Z(\vec{s}_0), \sigma_r^2, \tau]$ is also Gaussian with expected value μ and covariance matrix Σ given as:

$$\begin{aligned} \mu &= \mu_Y + \Sigma_{Y(\vec{s}), Z(\vec{s}_0)} \Sigma_{Z(\vec{s}_0)}^{-1} (z(\vec{s}_0) - \mu_{Z(\vec{s}_0)}) \\ \Sigma &= \Sigma_{Y(\vec{s})} - \Sigma_{Y(\vec{s}), Z(\vec{s}_0)} \Sigma_{Z(\vec{s}_0)}^{-1} \Sigma_{Z(\vec{s}_0), Y(\vec{s})} \end{aligned}$$

Mainly, the expected value of the posterior $[Y(\vec{s})|Z(\vec{s}_0), \sigma_r^2, \tau]$ and the variance, the diagonal of Σ , will be used to assess the posterior model.

However, in this derivation of our distribution, the underlying parameters for the covariance of the process model are not part of the actual distribution of interest, $[Y(\vec{s})|Z(\vec{s}_0)]$. We may derive this final distribution by using a combination of conditional probability and integration over the domain of the prior parameters:

$$\begin{aligned} [Y(\vec{s})|Z(\vec{s}_0)] &= \int \int [Y(\vec{s}), \sigma_r^2, \tau | Z(\vec{s}_0)] d\sigma_r^2 d\tau \\ &= \int \int [Y(\vec{s})|Z(\vec{s}_0), \sigma_r^2, \tau] \cdot [\sigma_r^2, \tau | Z(\vec{s}_0)] d\sigma_r^2 d\tau \end{aligned}$$

As we don't know the posterior distribution for the parameters σ_r^2, τ , we use Bayes Theorem combined with proportionality to evaluate it in integral ??, i.e.

$$\begin{aligned} [\sigma_r^2, \tau | Z(\vec{s}_0)] \cdot [Z(\vec{s}_0)] &= [Z(\vec{s}_0) | \sigma_r^2, \tau] \cdot [\sigma_r^2, \tau] \\ \implies [\sigma_r^2, \tau | Z(\vec{s}_0)] &\propto [Z(\vec{s}_0) | \sigma_r^2, \tau] \cdot [\sigma_r^2, \tau] \end{aligned}$$

As $[\sigma_r^2, \tau]$ is an assigned prior, we have given the priors σ_r^2 and τ independence with respect to each other. In addition, we need to scale the distribution to one for it to be a proper pdf. Combining this with a discretization of the priors, we get:

$$[\sigma_r^2, \tau | Z(\vec{s}_0)] = \frac{[Z(\vec{s}_0) | \sigma_r^2, \tau] \cdot [\sigma_r^2] \cdot [\tau]}{\int [Z(\vec{s}_0) | \sigma_r^2, \tau] \cdot [\sigma_r^2] \cdot [\tau]}$$

1.3.4 Case main

- sampling pattern, relation to case motivation

1.3.5 Case discussion

- model
- motivation
- results
- critique

1.3.6 Case code

1.4 Plan

- Skrive ned prosessmodel og påfølgende fordelinger på en fornuftig og oversiktlig måte
- Skrive om korrelasjonsfunksjoner og hvilken som er brukt her
- Bevis for at to gaussiske fordelinger gir en ny gaussisk fordeling ved kondisjonell fordeling