Project Parameter Estimation

Part II: Short introduction to Kriging

Jemil Avers Butt
Prof. Dr. Andreas Wieser

Administrative information

Involved people from GSEG

Jemil Avers Butt

ETH Zurich
Jemil Avers Butt
Institute of Geodesy and Photogrammetry
HIL D 45.1
Stefano-Franscini-Platz 5
8093 Zurich
Switzerland

Phone: +41 44 633 34 84
E-Mail: jemil.butt@geod.baug.ethz.ch

←

- Administrative questions
- Technical questions
- Programming project help

Prof. Dr. Andreas Wieser

ETH Zurich
Prof. Dr. Andreas Wieser
Institut of Geodesy and Photogrammetry
HIL D 47.2
Stefano-Franscini-Platz 5
8093 Zurich
Switzerland

Phone: +41 44 633 05 55
Fax: +41 44 633 11 01
E-Mail: andreas.wieser@geod.baug.ethz.ch

←

- Final assessment of your performances

Administrative information

Course overview

06.11.2018	Lecture	Stochastic preliminaries
	Homework	Read an introductory paper
13.11.2018	Attendance	Discussion of paper; kriging crash course
	Group Homework	Conducting experiments and interpreting them
20.11.2018	Attendance	Applications and extensions of Kriging; introduction to programming project
	Group Homework	Project: Interpolating a DEM, Signal Separation, ...
27.11.2018	Attendance	Help with programming project
04.12.2018	Attendance	Presentations, state of the art

Administrative information

Goals of this course

Goals

- Concepts of spatial statistics
- How to diagnose spatial relations
- Firm grip on Kriging
- Being able to actually use it

Sources

- Power Point slides
- Cressie N (1990) The origins of Kriging
- Chiles J P & Delfiner P (2012) Geostatistics

Measurement

- Active Participation, well founded opinions (9 pts)
- Summary and discussion (6 pts)
- Written thoughts on experimenting with stochastic processes (6 pts)
- Programming Project (Documentation, correctness of code, interpretation of results) (9 pts)

Administrative information

Homeworks and project

Why?

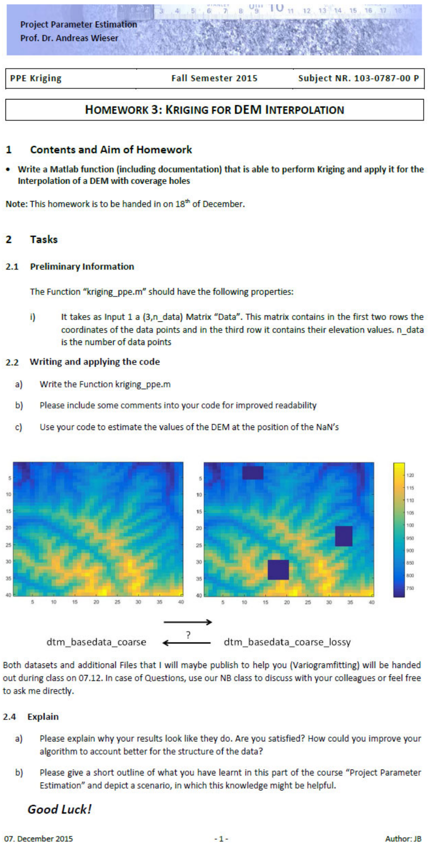
- **Homeworks and project instead of examination**
- **Useful possibility for practical experience**
- **How and when to apply which methods is important**

What?

- **Reading papers**
- **Generating your own spatially dependent data sets, analyzing and interpreting them**
- **Programming project to implement what you have learned**

How?

- **Alone or in groups, discussion in class afterwards**
- **Project handed in for grading**



Project Parameter Estimation
Prof. Dr. Andreas Wieser

PPE Kriging	Fall Semester 2015	Subject NR. 103-0787-00 P

HOMEWORK 3: KRIGING FOR DEM INTERPOLATION

1 Contents and Aim of Homework

- **Write a Matlab function (including documentation) that is able to perform Kriging and apply it for the Interpolation of a DEM with coverage holes**

Note: This homework is to be handed in on 18th of December.

2 Tasks

2.1 Preliminary Information

The Function "kriging_ppe.m" should have the following properties:

i) It takes as Input 1 a (3,n_data) Matrix "Data". This matrix contains in the first two rows the coordinates of the data points and in the third row it contains their elevation values. n_data is the number of data points

2.2 Writing and applying the code

a) Write the Function kriging_ppe.m

b) Please include some comments into your code for improved readability

c) Use your code to estimate the values of the DEM at the position of the NaN's

Both datasets and additional Files that I will maybe publish to help you (Variogramfitting) will be handed out during class on 07.12. In case of Questions, use our NB class to discuss with your colleagues or feel free to ask me directly.

2.4 Explain

a) Please explain why your results look like they do. Are you satisfied? How could you improve your algorithm to account better for the structure of the data?

b) Please give a short outline of what you have learnt in this part of the course "Project Parameter Estimation" and depict a scenario, in which this knowledge might be helpful.

Good Luck!

07. December 2015 - 1 - Author: JB

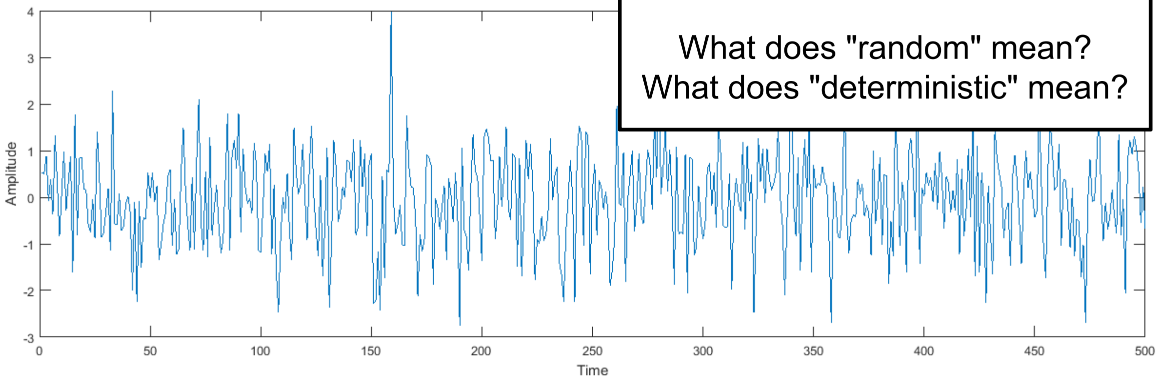
Exposition and introduction

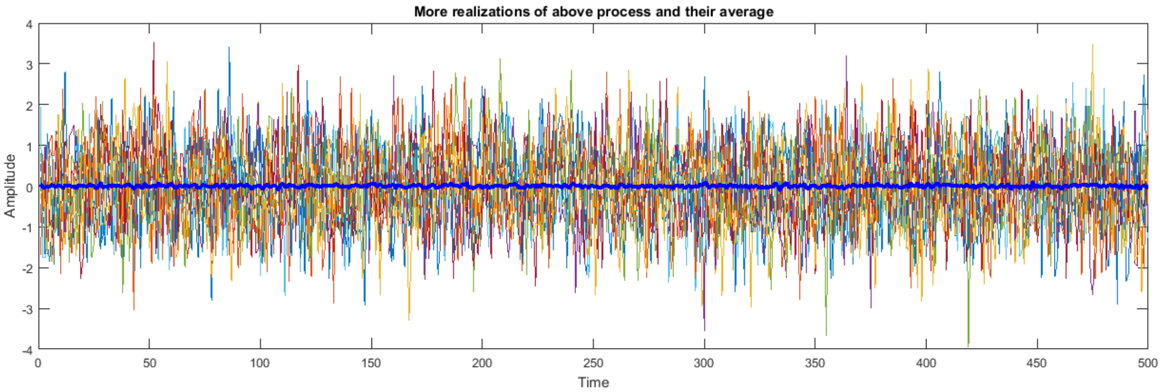
- Rethink your notion of "Randomness"
- Know conceptual underpinnings of Kriging

Determinism vs Randomness

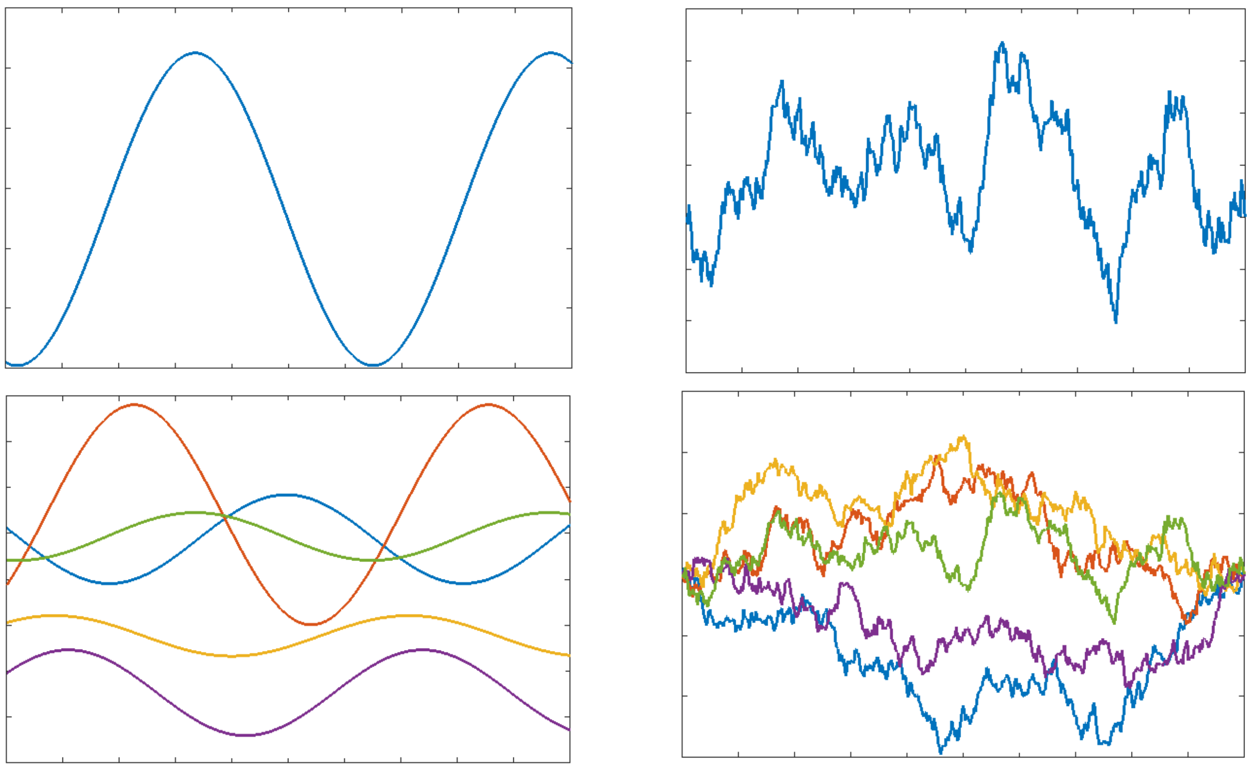
Please discuss briefly:

What does "random" mean?
What does "deterministic" mean?

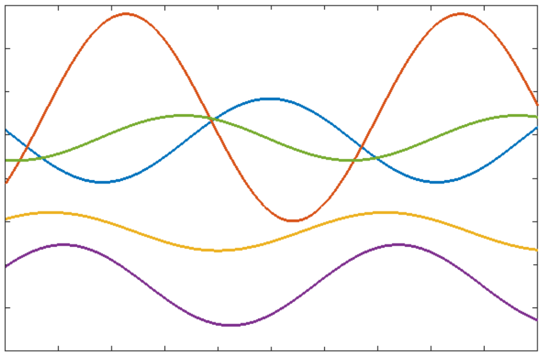


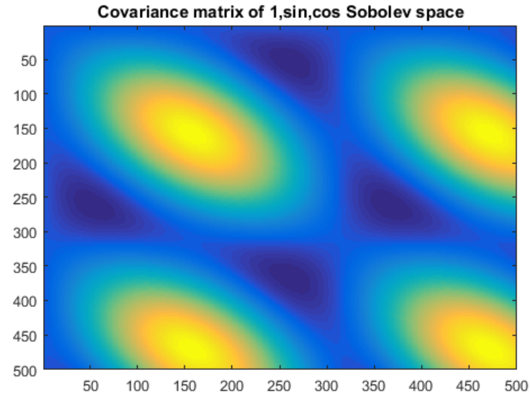


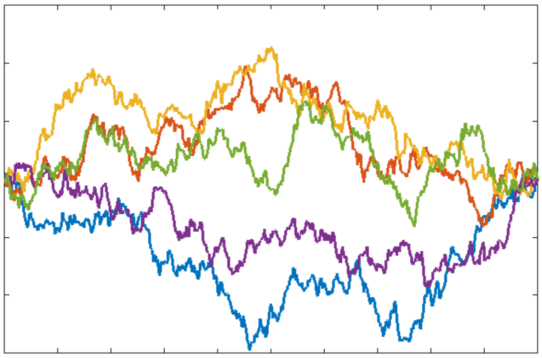
Determinism vs Randomness

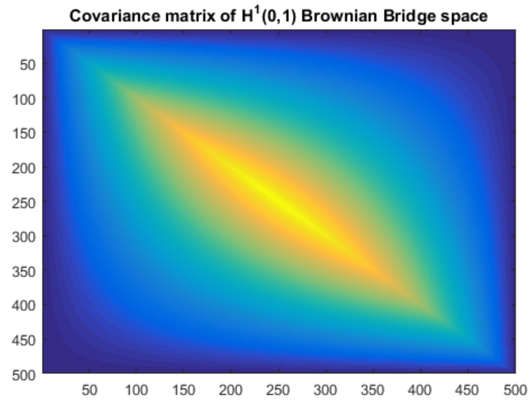


Determinism vs Randomness

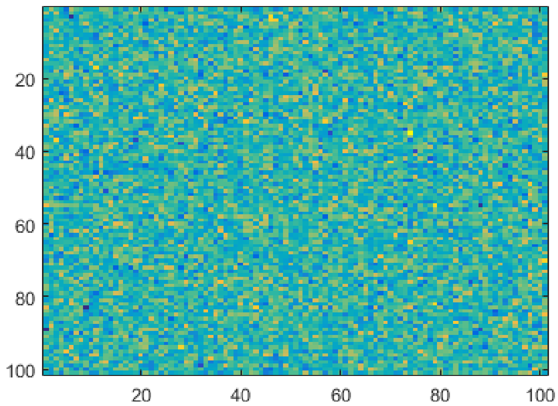


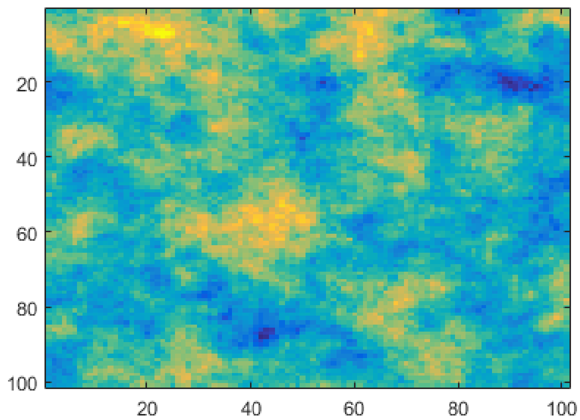


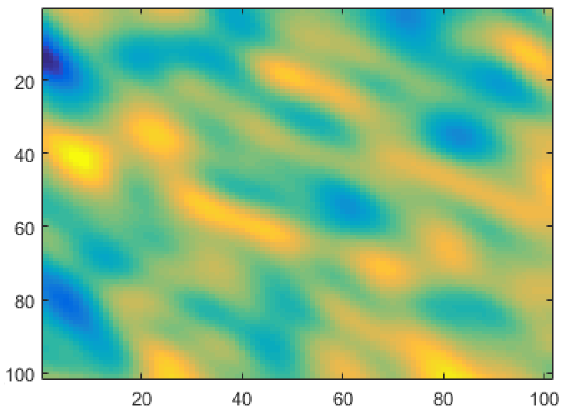




Determinism vs Randomness







Again, please discuss briefly:

Has your opinion about "randomness" changed?

Explain in which sense the images are random

Determinism vs Randomness

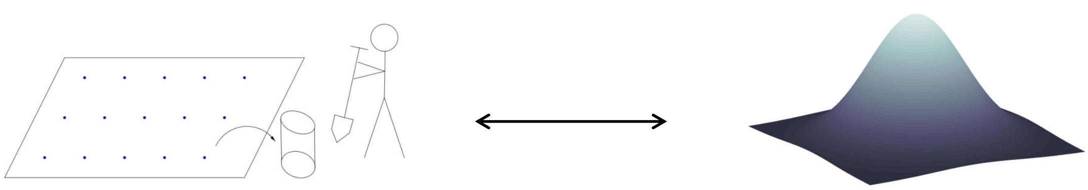
Discussion: Definition/ Intuition concerning randomness

Introduction to geostatistics

<u>Geo</u>	<u>statistics</u>
Origins	Theoretical justification
First applications	Reason for wide applicability
Widest use	New Developments

Why it is probably useful for you

- **Refreshing your knowledge on fundamental statistical principles**
- **Sensibilisation for spatial structure in your data.**
- **Getting new ideas on how to infer parameters from your data**
- **Having code that you can actually use later on**



Spatial relationships in statistics

- Sparsely sampled geological data
- Some core samples are given. Mining economically feasible? Ore content estimation needed
- At that time mathematical geology was quite new
- Also computational power only increased in 1953
- Rules of thumb and intuition were common

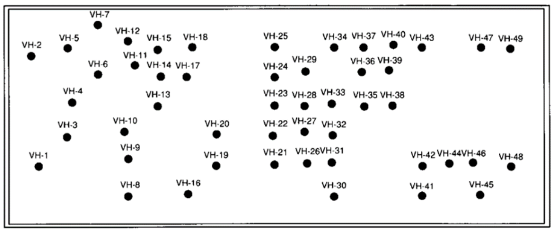


Image: Christakos(1992)

- Relation between samples and population?
- Spatial structure unignorable -> need for formalization
- High financial risks: justification of interpolation methods?

http://ironoretrader.biz

Kriging

- Also called: Gaussian process regression, Kolmogorov-Wiener prediction

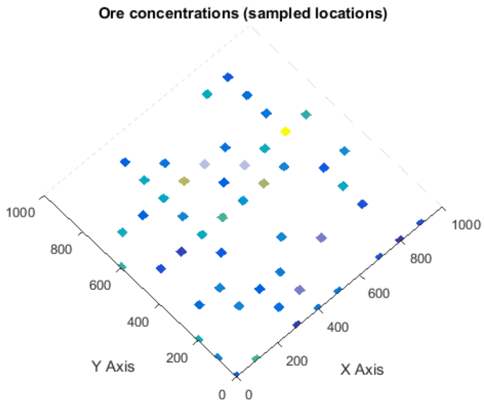
- First crude applications by Krige (1951)
- Practical and problem-oriented approach

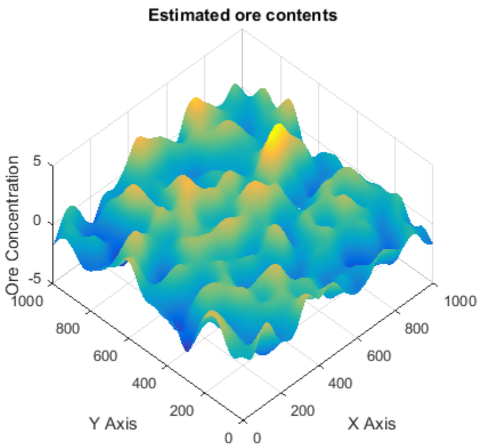
- Theoretical foundations by Matheron (1962)
- Regional Variables, Random Fields
- Close relation to regression

  ... but important differences

Interpolating datasets
- Mining, meteorology, environmental sciences, epidemiology, ecology

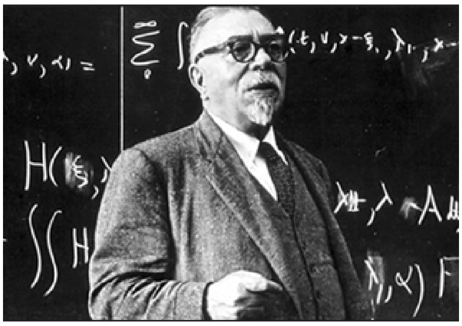
Surrogate modelling
- Computer science, numerical simulations



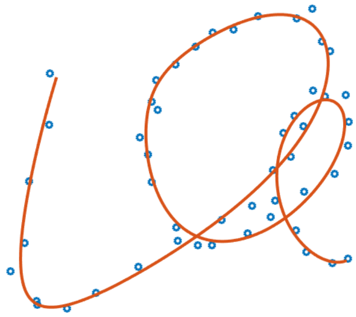


Origins of Kriging

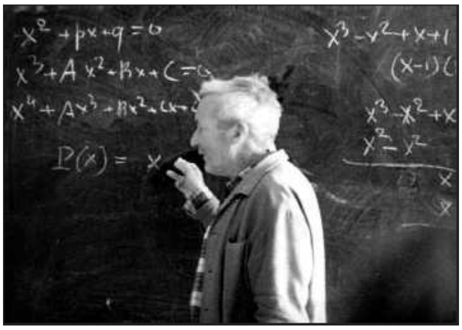
http://www.norbertwiener.umd.edu



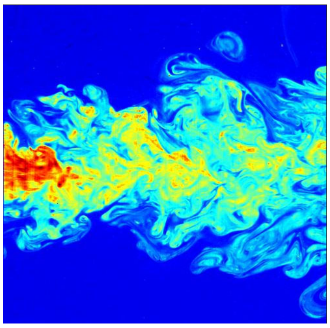
http://www.wikipedia.com

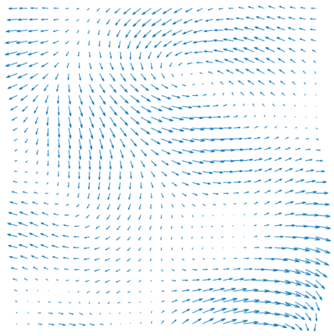


http://www.kolmogorov.com



http://www.wikipedia.com





Kriging

- **Estimation should be right on the average**
- **Estimation should not deviate to much from the truth**
- **Estimation should respect spatial correlations found in the data**

- **We want accuracy and precision**
- **So we get a minimization problem with an equality constraint -> Method of Lagrange multipliers**

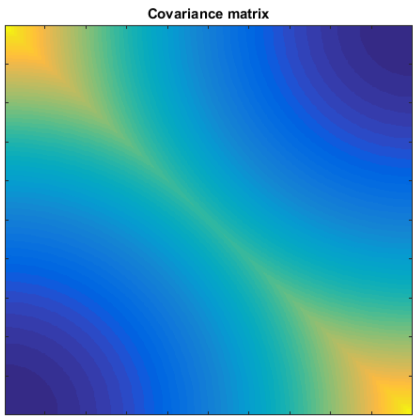
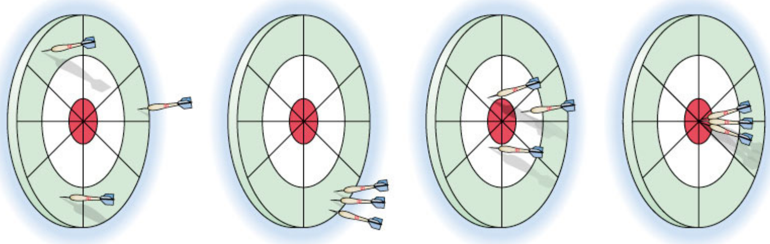
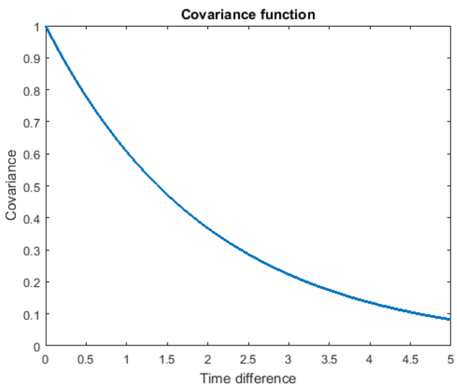


Image:extensionengine.com



(a) Low accuracy Low precision
(b) Low accuracy High precision
(c) High accuracy Low precision
(d) High accuracy High precision



Stochastic preliminaries

- Methods of statistical inference
- Typology of inference tasks
- Measures of stochastic dependence

Central tools of statistics

Descriptive and inductive statistics

- **Descriptive: Describe characteristics of (sample) distribution**
- **Inductive: Infer real distribution based on sample distribution**
- **Two characteristics are generally considered most important**

First Moment $\mu_1 = \int_{-\infty}^{\infty} x f(x)dx = E[X] = \text{Expected value}$

Second Moment $\mu_2 = \int_{-\infty}^{\infty} (x-\mu_1)^2 f(x)dx = E[(X-E[X])^2] = \text{Variance}$

$= E[X^2] - E[X]^2$

E Linear operator: $E[aX + bY + c] = aE[X] + bE[Y] + c$

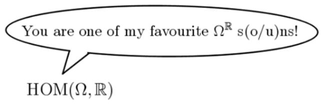
E[X] just a number: $E[E[X]] = E[X]$

Generalizations to several random variables possible (e.g. covariance)

Covariance $= \int_{-\infty}^{\infty}\int_{-\infty}^{\infty} (x-\mu_x)(y-\mu_y) f(x,y)dxdy = E[(X-E[X])(Y-E[Y])]$

Lets start small: μ_1, μ_2 **from finite sample?**

Central tools of statistics

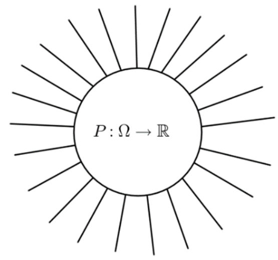


Parameter estimation and maximum likelihood

- **For a sample find the parameters for a distribution that makes the sample most probable.**
- **Use the probability** $P(X_1 = x_1 \wedge X_2 = x_2 \ldots \wedge X_n = x_n)$

$$\mathcal{L}(\theta, x_1, x_2, \ldots x_n) = \prod_{i=1}^{n} f(x_i, \theta) = f(x_1, \theta) * f(x_2, \theta) * \ldots * f(x_n, \theta)$$

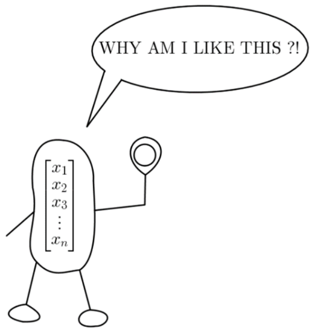
$$\left\{\hat{\theta}\right\} := \left\{\underset{\theta \in \Theta}{\arg\max}\, \mathcal{L}(\theta, x_1, x_2, \ldots x_n)\right\}$$



- **The value of this parameter is called:**

  Maximum likelihood estimation for θ
- **Discuss this formula. Why product? Implications?**

  How and when would you use this formula to solve an estimation problem?
- **Since logarithmic function is increasing, we can also maximize** $\ln(\mathcal{L}(\theta, x_1, x_2, \ldots x_n))$ **instead.**
- **This function is called log-likelihood function.**



Central tools of statistics

Maximum likelihood, Least Squares

Central tools of statistics

Interpolation and regression

- **As seen on the last slide: Parameter estimation can be seen as optimization problem**
- **Same kinds of solutions as typical regression problems with least squares approach**

- **Conversely fitting by least squares is just maximization of a likelihood function**
- **We have gotten quite technical:**

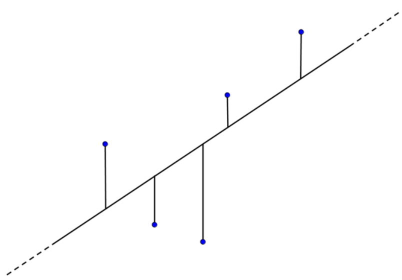
  - were only able to estimate mean of normally distributed random variable

  - can we do something more useful?

- **Other kinds of regression and interpolation problems can also be best understood by their stochastical meaning**

Central tools of statistics

Interpolation and regression examples



- **Typical regression: fit a line $p(x) = \theta_1 * x + \theta_2$ into data v_{x_i}.**

$$\text{Find } \begin{bmatrix} \theta_1 \\ \theta_2 \end{bmatrix} \quad s.t. \quad R \; = \; \sum_{i=1}^{n} \left[p(x_i) - v_{x_i} \right]^2 \;\rightarrow\; min$$

$$\Rightarrow \begin{cases} \dfrac{\partial}{\partial \theta_1} R = 2\sum \left[\theta_1 x_i + \theta_2 - y_i \right] x_i = 0 \\ \dfrac{\partial}{\partial \theta_2} R = 2\sum \left[\theta_1 x_i + \theta_2 - y_i \right] \;\; = 0 \end{cases} \quad \Leftrightarrow \quad \begin{bmatrix} \sum x_i^2 & \sum x_i \\ \sum x_i & n \end{bmatrix} \begin{bmatrix} \theta_1 \\ \theta_2 \end{bmatrix} = \begin{bmatrix} \sum x_i y_i \\ \sum y_i \end{bmatrix}$$

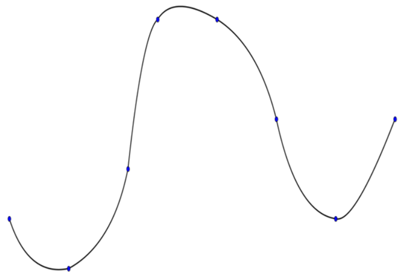
$$\Leftrightarrow \quad \begin{bmatrix} \theta_1 \\ \theta_2 \end{bmatrix} = \begin{bmatrix} \sum x_i^2 & \sum x_i \\ \sum x_i & n \end{bmatrix}^{-1} \begin{bmatrix} \sum x_i y_i \\ \sum y_i \end{bmatrix}$$

- **Classical solution: least squares estimation**
  = Maximum likelihood estimation. Better?
  See top right picture for weaknesses.

- **Coordinate system invariant solutions?**

$$\begin{tikzcd} \textit{Problem} \arrow[r, "\text{Coordinate}", "\text{transform}"'] \arrow[d, "\text{minimize}"', "\text{something}"] & \textit{Problem} \arrow[d, "\text{minimize}"', "\text{something}"] \\ \textit{Solution} \arrow[r, "\text{Coordinate}", "\text{transform}"'] & \textit{Solution} \end{tikzcd}$$

Central tools of statistics



Interpolation and regression examples

- Typical IDW Interpolation:For each point x, estimate its value v_x as a linear combination of the measured values v_{x_i} in form of a weighted average.
- To find the weights, the minimization of a functional is necessary (just like in regression)

$$\text{Find } v_x \ s.t. \ R = \sum_{i=1}^{n} \frac{\left[v_x - v_{x_i}\right]^2}{d(x,x_i)^2} \rightarrow min \qquad v_x = \sum_{i=1}^{n} w_i v_{x_i}$$

- This is just like in regression: But this time points far away matter less.

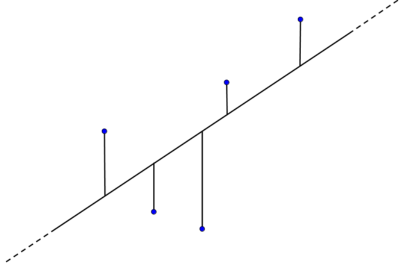
$$\frac{\partial}{\partial v_x} R = 2\sum_{i=1}^{n} \frac{(v_x - v_{x_i})}{d(x,x_i)^2} = 0 \Leftrightarrow v_x \sum_{i=1}^{n} \frac{1}{d(x,x_i)^2} = \sum_{i=1}^{n} \frac{v_{x_i}}{d(x,x_i)^2}$$

- Different goal, same procedure : Both related to maximum likelihood and normal distribution

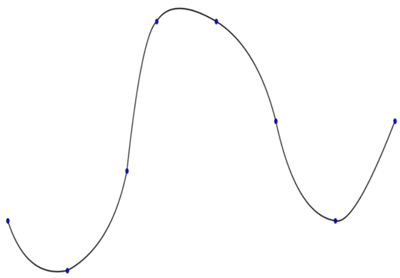
$$\Leftrightarrow v_x = \frac{\sum_{i=1}^{n} \frac{v_{x_i}}{d(x,x_i)^2}}{\sum_{i=1}^{n} \frac{1}{d(x,x_i)^2}}$$

Central tools of statistics

Regression vs **interpolation**

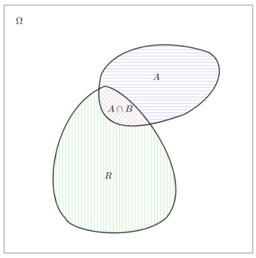


"Find line through points"



"Find some function through points"

Central tools of statistics



Important theorems

- **What happens when things are not independent?**

  Bayes theorem: $P(A\mid B)=\dfrac{P(B\mid A)P(A)}{P(B)}$

- **Given the observation B, A is this probable**
- **You have installed a warning system. It detects the presence of a dangerous condition and issues an alarm. It measures once a day. Tests in your laboratory lead to the following conclusions:**

  1. **P(alarm) given presence of risk** **=** 0.5 **(bad system!)**
  2. **P(alarm) given no presence of risk** **=** 10^{-5}
  3. **P(risk)** **=** 10^{-3}

- **One morning it issues an alarm. What is the probability of the dangerous condition being reality?**

$$P(R\mid A)=\frac{P(A\mid R)P(R)}{P(A)} = \frac{P(A\mid R)P(R)}{P(A\mid R)P(R)+P(A\mid R^c)P(R^c)}$$

$$=\frac{1}{1+\dfrac{P(A\mid R^c)P(R^c)}{P(A\mid R)P(R)}} = \frac{5*10^{-4}}{5.0999*10^{-4}} \approx 98\%$$

Central tools of statistics

Proof: Independence => Uncorrelatedness

Central tools of statistics

Conditional probabilities: Discussion

- Card game!
- Famous US-american football star O.J. Simpson was accused of killing his wife in 1994. It was one of the most high profile cases in the last century. During the trial O.J. Simpsons lawyer claimed that the known history of violence of Simpson against his wife is irrelevant to the case. He used the argument:

0.1% of violent husbands kill wife -> being violent doesn’t make O.J.’s guilt more likely

- Statistician Irving J. Good replied to this in a letter. What do you think he wrote?
- Lesson: Never trust your gut feeling (when you are lost in Probability Space)!

Random variables, random functions, random fields

- Definition of RV's
- Calculation rules for Expectation, Covariance
- Conceptual idea of more abstract stochastic objects

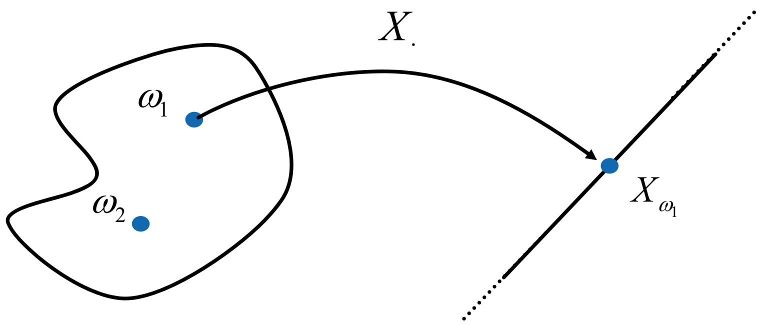
Algebra of Random Variables

Definition and explanation

Random Variable: $X_{\cdot} : \Omega \ni \omega \mapsto X_{\omega} \in \mathbb{R}$

Ω : Sample space (of outcomes)

$\mathbb{R}$: The real numbers

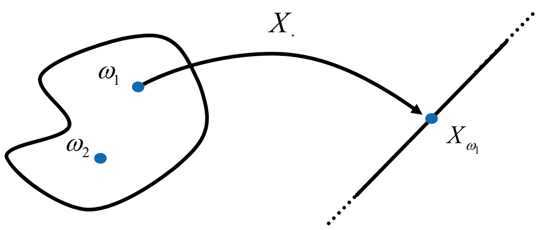


ω_1 = Temperature, time of day, mood, atmospheric turbulence,...

X_{ω_1} = Body temperature in °C
(A number !)

Algebra of Random Variables

Discussion



$L^2(\Omega)$ = Space of RV's with finite Variance

$\mathrm{L}^2(\Omega) = \left\{ X : \Omega \to \mathbb{R} \mid \sigma_X^2 < \infty \right\}$

$1. X, Y \in L^2(\Omega) \Rightarrow \alpha X + \beta Y \in L^2(\Omega)$

$2. X, Y \in L^2(\Omega) \Rightarrow \quad XY \quad \in L^2(\Omega)$

What do you think of:

- ☐ The number 5 can be interpreted as a RV with variance 0
- ☐ Choosing a number $x \in \mathbb{R}$ at random is a RV
- ☐ The rule for choosing a number $x \in \mathbb{R}$ at random is a RV
- ☐ A RV is a number that is not yet known but becomes known after observation

Why so nitpicky?!

Because:

Algebra of Random Variables

Calculation rules for $E[\cdot]$ and $\sigma(\cdot,\cdot)$

Let $E[\cdot]$ be the "Expectation operator":

$$E[\cdot]: L^2(\Omega) \ni X \mapsto E[X] \in \mathbb{R}$$

$$E[X] = \int_\Omega X_\omega P(d\omega) = \int_\mathbb{R} x f_X(x) dx$$

Additional info:
$E[\cdot]$ is a linear functional, since it belongs to the dual space $L^2(\Omega)^*$

$\sigma(\cdot,\cdot)$ defines an inner product of RV making them into a Hilbert space

$E[\cdot]$ is linear operator:

$$E\left[\sum_{k=1}^{n} \alpha_k X_k + a\right] = \sum_{k=1}^{n} \alpha_k E[X_k] + a \qquad X_k \in L^2(\Omega) \quad \alpha_k, a \in \mathbb{R}$$

Let $\sigma(\cdot,\cdot)$ be the "Covariance operator":

$$\sigma(\cdot,\cdot): L^2(\Omega) \times L^2(\Omega) \ni (X,Y) \mapsto \sigma(X,Y) \in \mathbb{R}$$

$$\sigma(X,Y) = \int_\Omega X_\omega Y_\omega P(d\omega) - E[X]E[Y] = \int_\mathbb{R}\int_\mathbb{R} \left(x - E[X]\right)\left(y - E[Y]\right) f_{XY}(x,y) dx dy$$

$$= E[XY] - E[X]E[Y]$$

$\sigma(\cdot,\cdot)$ is bilinear operator:

$$\sigma\left(\sum_{i=1}^{m} \alpha_i X_i, \sum_{j=1}^{n} \beta_j Y_j\right) = \sum_{i=1}^{m}\sum_{j=1}^{n} \alpha_i \beta_j \sigma(X_i, Y_j) \qquad X_k, Y_k \in L^2(\Omega) \quad \alpha_k, \beta_k \in \mathbb{R}$$

Stochastic Processes, Random Fields

Definition Stochastic Process: $X_\cdot(\cdot): \Omega \times T \ni (\omega, t) \mapsto X_\omega(t) \in \mathbb{R}$

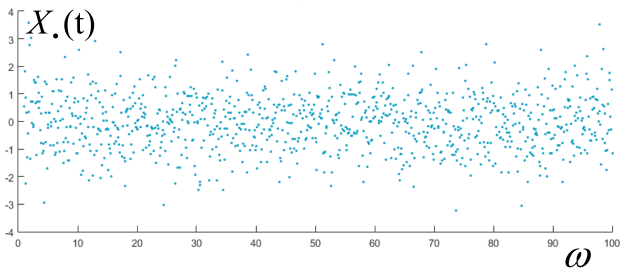
Ω Probability space, $T \subseteq \mathbb{R}$ index set, $X_\cdot(t) \in L^2(\Omega)$ RV $\forall t \in T$

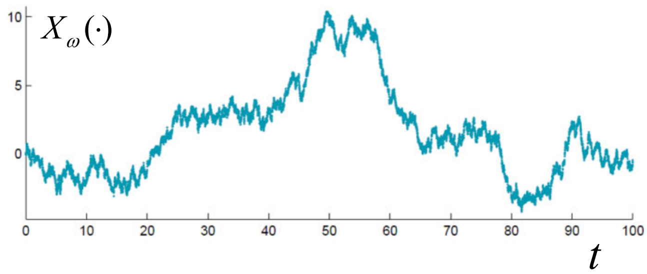
Interpretation I:
Sequence of RV's

$X_\cdot(\cdot): T \ni t \mapsto X_\cdot(t) \in L^2(\Omega)$

Interpretation II:
Random function

$X_\cdot(\cdot): \Omega \ni \omega \mapsto X_\omega(\cdot) \in F(T)$





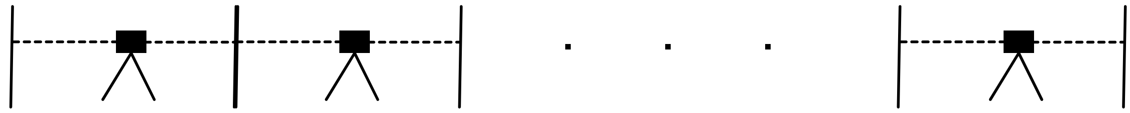
Random Field: Same thing but index set T has dim >=2

Stochastic Processes, Random Fields

Levelling error, Wiener Process

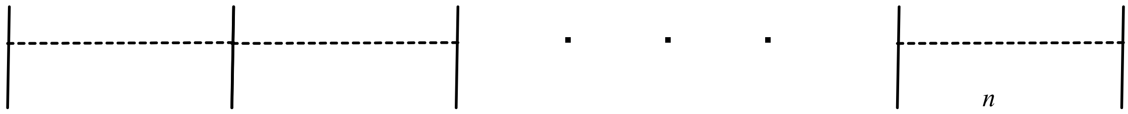
$X_k \sim N(0, \sigma_0^2)$
all uncorrelated

Individual error



X_1 X_2 . . . X_n

Total error



X_1 $X_1 + X_2$ $\sum_{k=1}^{n} X_k$

W_1 W_2 W_n

Stochastic Processes, Random Fields

Levelling error, Wiener Process

$X_k \sim N(0, \sigma_0^2)$ all uncorrelated; little errors during measurement k

$W_t = \sum_{k=1}^{t} X_k$ is called the (discrete) Wiener process; a stochastic process

i) Explain, how this is a stoch. proc.

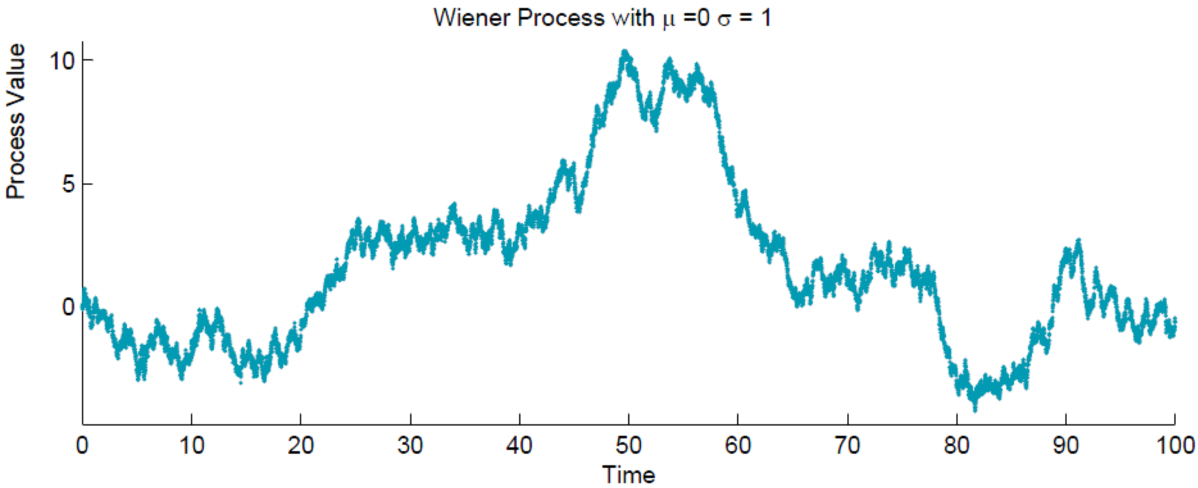
ii) Calculate E[W_t]

ii) Calculate $\sigma(W_t, W_t) = \sigma_t^2$. Do you know this from somewhere?

Stochastic dependence and independence

Stochastic processes

- How to estimate the strength of neighborhood-relations?
- Can we estimate global features of this stochastic process?

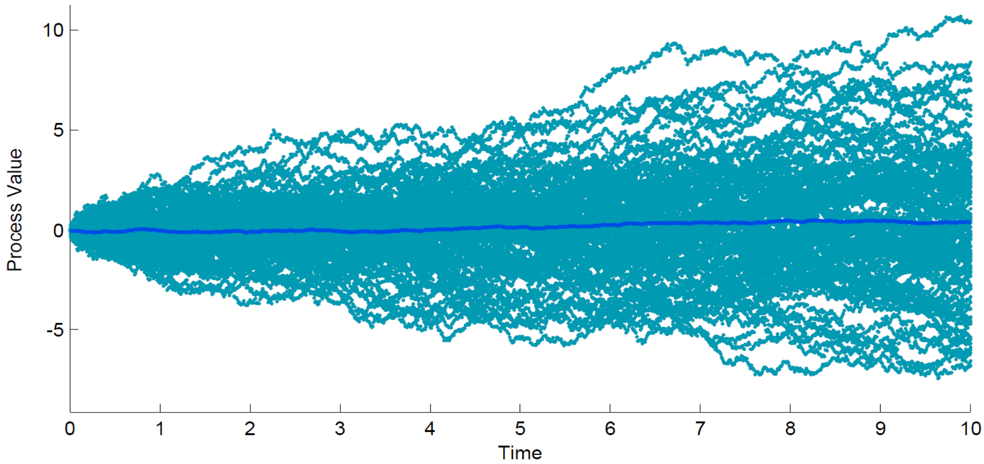


- Wiener process = integrated white noise. Expected value = 0. A lie? How to estimate?

Stochastic dependence and independence

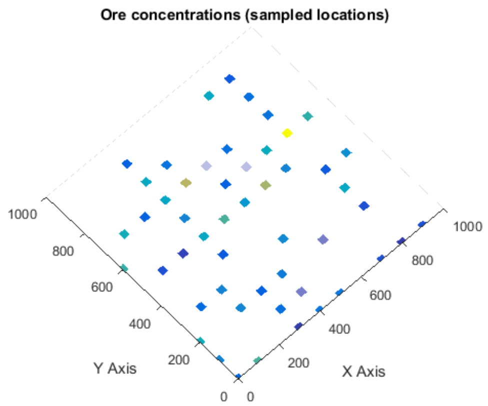
Stochastic processes

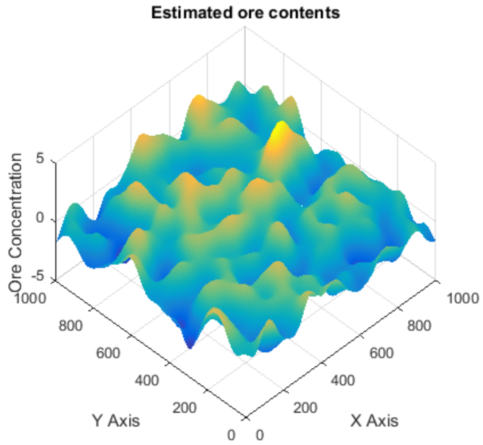
- We only have found a way to estimate the strength of neighborhood-relations
- Can we estimate global features (expected value) of this stochastic process?



- We need a new method for statistically sound estimation of stochastic processes.
- Luckily we already have all the tools. We just need to patch them together.
- Minimization of a functional to find coefficients <- Wiener-Kolmogorov filter theory

What does Kriging do?





Input
- Spatial data
  (= vals, coords)

Estimate $\mathrm{E}[X_{.}(\cdot)]$

Estimate $\sigma(X_{.}(s), X_{.}(t)) = C(s,t)$

Minimize $\mathrm{E}\left[(\hat{X}_{.}(s) - X_{.}(s))^2\right]$

$\hat{X}_{.}(s)$ optimal Estimator

Output
- Optimal estimation
- Estimation variances