Selected topics in statistics
Spatial Statistics
Course specifications

Lecturer: François Bachoc, PhD

Practical information

Time and Place

Time: 12am-2pm on Monday

- October 14, 21, 28,
- November 4, 11, 18, 25
- December 2, 9, 16
- January 13, 20, 27

Place:

- For **lectures** and **exams**, on 1.OG, in Seminarraum 4.
- For **lab sessions**, on 3.OG, in PC Seminarraum 07.

Office hours

On Friday: 10am - 11am.

Evaluation

The evaluation has three components:

- *Homeworks* (30 %). Homeworks are given at the end of a class, and are due for the following class. The homeworks can be send on paper version at the beginning of the following class, or by email (ensure that I acknowledge the receipt) before the following class.
- **Mid-term exam** (35 %). On site paper exam. No course documents and no calculators are allowed.
- **Final exam** (35 %). On site paper exam. No course documents and no calculators are allowed.

Prerequisite

Background in probability and statistics (e.g. cdf, pdf, conditional distribution, parameter estimation, Gaussian vectors). Some notions in programming and some first notions in the use of a numerical software like R or Matlab.

An introduction to the course

In this course, we study the field of **spatial statistics** called **geostatistics**. Geostatistics consists in observing output quantities $Z(x_1), ..., Z(x_n)$, where $x_1, ..., x_n$ are input quantities belonging to a domain $\mathcal{D}$. In practical applications, $x \in \mathcal{D}$ can represent, for instance, a point in a two-dimensional map, a position in the three-dimensional atmosphere, or a set of numerical conditions specifying a computer-based simulation of a physical system. In the three examples above, the quantity $Z(x)$ can represent, for instance, the amount of precipitation received in a month, the concentration of a pollutant, and the result of the computer-based simulation.

In this course, we will focus on the case where Z is modeled as a trajectory of a random process. This is known as a **Kriging** model. We will also especially treat the case when the random process Z is a **Gaussian process**.

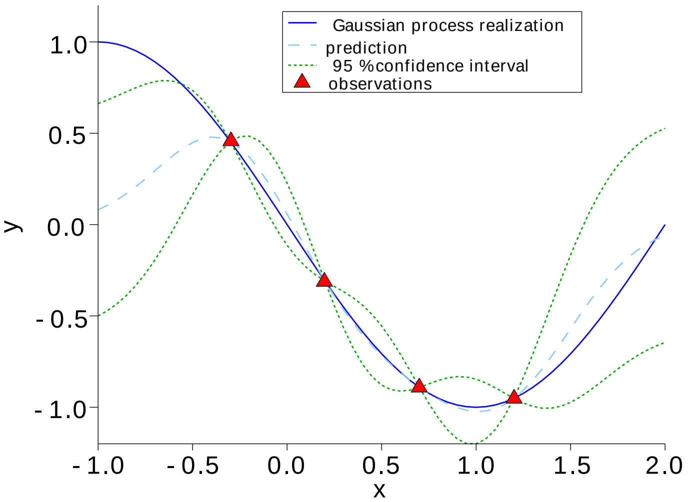

Figure 1: A Gaussian process trajectory on the domain $\mathcal{D} = [-1, 2]$ is observed at four observation points. Thanks to Gaussian process prediction formulas, we can compute an approximation of the unknown trajectory, and an associated 95% confidence band.

Our goal will essentially be, form the observations $Z(x_1), ..., Z(x_n)$, to predict the rest of the trajectory $Z : \mathcal{D} \to \mathbb{R}$. This consists either in **prediction**, meaning approximating the realization $Z(x_{new})$ for any $x_{new} \in D$, or **conditional simulation**, meaning generating trajectories $Z : \mathcal{D} \to \mathbb{R}$ that are conditioned by the observations $Z(x_1), ..., Z(x_n)$. Examples of prediction and conditional simulation are given in figures 1 and 2.

Prediction and conditional simulation necessitate to estimate the distribution of the random process Z. In the case where this process is Gaussian, the main problem is to estimate its **covariance function**. We will focus on the **Maximum Likelihood** estimation method for this problem.

Tentative course programm

- **Class 1 (lecture):** General introduction, stochastic processes. *Homework.*
- **Class 2 (lecture):** Stochastic processes. *Homework.*
- **Class 3 (lecture):** Gaussian processes. *Homework.*
- **Class 4 (lecture):** A focus on the covariance function. *Homework.*
- **Class 5 (lecture):** Classical covariance functions. *Homework.*
- **Class 6 (lecture):** Gaussian conditioning, prediction. *No homework.*
- **Class 7: mid-term exam**. *No homework.*
- **Class 8 (lab-session):** Unconditional simulation, prediction, conditional simulation. *Homework: finish the lab-session and write a report.*
- **Class 9 (lecture):** Covariance function estimation, Maximum Likelihood. *Homework.*
- **Class 10 (lab-session):** Maximum Likelihood. *Homework: finish the lab-session and write a report.*
- **Class 11 (lecture):** Universal Kriging. *Homework.*
- **Class 12 (lecture):** Some additional topics. *No homework.*
- **Class 13: final exam**. *No homework.*

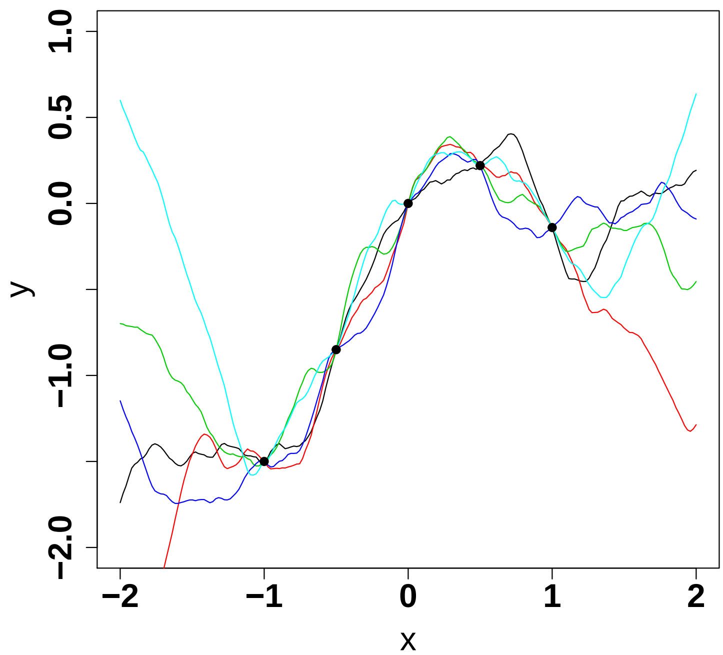

Figure 2: Plot of realizations of a Gaussian process on $[-2, 2]$, that are conditioned by five observations. These realizations are obtained thanks to Gaussian process conditioning formulas.

Some further references

- Books:

[SWN03] On Kriging models for computer experiments.

[RW06] On Gaussian process modeling.

[Ste99] On asymptotic theory for Kriging models.

- Articles:

[SWMW89] On Kriging models for computer experiments.

[JSW98] On the use of Kriging models for global optimization.

[BGL+12] On the use of Kriging models for estimating a probability of failure.

[BBP+07] On the use of Kriging models for computer model validation (from experimental data).

- The Gaussian process website: http://www.gaussianprocess.org/.
- The R packages KrigInv (Kriging-based inversion), MuFiCokriging (Kriging for multi-fidelity computer models), DiceKriging (Kriging methods) and DiceOptim (Kriging-based inversion).

References

[BBP+07] M. J. Bayarri, J.O. Berger, R. Paulo, J. Sacks, J.A. Cafeo, J. Cavendish, C.H. Lin, and J. Tu. A framework for validation of computer models. *Technometrics*, 49(2):138–154, 2007.

[BGL+12] J Bect, D Ginsbourger, L Li, V Picheny, and E Vazquez. Sequential design of computer experiments for the estimation of a probability of failure. *Statistics and Computing*, 22:773–793, 2012.

[JSW98] D.R. Jones, M. Schonlau, and W.J. Welch. Efficient global optimization of expensive black box functions. *Journal of Global Optimization*, 13:455–492, 1998.

[RW06] C.E. Rasmussen and C.K.I. Williams. *Gaussian Processes for Machine Learning*. The MIT Press, Cambridge, 2006.

[Ste99] M.L Stein. *Interpolation of Spatial Data: Some Theory for Kriging.* Springer, New York, 1999.

[SWMW89] J. Sacks, W.J. Welch, T.J. Mitchell, and H.P. Wynn. Design and analysis of computer experiments. *Statistical Science*, 4:409–423, 1989.

[SWN03] T.J Santner, B.J Williams, and W.I Notz. *The Design and Analysis of Computer Experiments.* Springer, New York, 2003.