Bayesian Notes for building the geostatistical MANOVA-KNN pipeline

Roxana Tesileanu

roxana.te@web.de
INCDS, Romania

November 2017

Contents

1 Introduction

From Gelman et al. 2014

BAYESIAN INFERENCE is the process of fitting a probability model to a set of data and SUMMARIZING THE RESULT BY A PROBABILITY DISTRIBUTION ON:

1. THE PARAMETERS OF THE MODEL and on
2. THE UNOBSERVED QUANTITIES SUCH AS PREDICTIONS FOR NEW OBSERVATIONS.

=> make inferences from data using probability models for quantities we observe and for quantities we wish to learn. **THE ESSENTIAL CHARACTERISTIC OF BAYESIAN MODELS IS THEIR EXPLICIT USE OF PROBABILITY FOR QUATIFYING UNCERTAINTY IN INFERENCES BASED ON STATISTICAL DATA ANALYSIS**. This is the main idea of the MANOVA-KNN pipeline, to analyse errors and reduce them. I've already sketched a geometric approach. It needs to be better backed up by probability theory. Which is what I expect to find in these Bayesian texts.

Steps of Bayesian Data Analysis:

1. setting up A FULL PROBABILITY MODEL - a JOINT PROBABILITY DISTRIBUTION FOR ALL OBSERVABLE AND UNOBSERVABLE QUANTITIES IN A PROBLEM.
2. CONDITIONING ON OBSERVED DATA - calculating and interpreting the appropriate POSTERIOR DISTRIBUTION.
3.

Note

This document is "under construction". It contains older notes of mine on Bayesian data analysis. Some were used in technical reports of mine (see https://w
ww.researchgate.net/publication/317549069_poisson_model) and also new sections aiming at creating the background necessary for the implementation of the MANOVA-KNN pipeline in geostatistics using the idea of **posterior predictive checks** (Introduction and Deduction in Bayesian Data Analysis, Andrew Gelman, 2011) [1]. For this purpose, I will have to work through books building up my skills, fortunately I was given a hint (and a copy) by a friend on "Bayesian Data Analysis for Social Sciences" by Simon Jackman (Wiley, 2009) [2] and "Bayesian Data Analysis" by Andrew Gelman, John B. Carlin, Hal S. Stern, David B. Dunson, Aki Vehtari and Donald B. Rubin (CRC,

2014) [3]. Please download the current version from my GitHub profile under the multivariate_analyses project repository: https://github.com/RoxanaTesileanu/multivariate_analyses/blob/master/literature_analysis/geospatial_scala/bayesian_notes_geosp.pdf.

The statistical plots in this document were generated in Scala using the JavaPlot package developed by Panayotis Katsaloulis [4]. You can find the scala source files used for generating them under the link: https://github.com/RoxanaTesileanu/multivariate_analyses/tree/master/DeepLearning/src/main/scala/com/mai/scalaPlot.

The present document was edited using Latex [5] (https://www.latex-project.org/). The source .tex file of the present document is also available in the multivariate_analyses repository on my GitHub profile. Special thanks to Gustavo Mezzetti for the Latex halloweenmath package: http://mirrors.concertpass.com/tex-archive/macros/latex/contrib/halloweenmath/halloweenmath-man.pdf!

References

[1] A. Gelman, "Introduction and Deduction in Bayesian Data Analysis," *Rationality, Markets and Morals Journal*, vol. 2, pp. 67–78, 2011.

[2] S. Jackman, *Bayesian analyses for social sciences*, ser. Wiley Series in Probability and Statistics. John Wiley & Sons, 2009.

[3] A. Gelman, J. B. Carlin, H. S. Stern, Dunson, David B., A. Vehtari, and D. B. Rubin, *Bayesian Data Analysis*, 3rd ed. CRC Press, 2014.

[4] P. Katsaloulis, "JavaPlot," 2017. [Online]. Available: http://javaplot.panayotis.com/

[5] C. Vellage, "LaTeX-Tutorial.com." [Online]. Available: https://www.latex-tutorial.com/