Bayesian Linear Regression on Functional Data

defended the Novembre 22th 2017 by Paul-Marie Grollemund

undervision of Christophe Abraham, Pierre Pudlo and Meili Baragatti

Abstract : The linear regression model is a common tool for a statistician. If a covariable is a curve, we tackle a high-dimensional issue. In this case, sparse models lead to successful inference, for instance by expanding the functional covariate on a smaller dimensional space.
In this thesis, we propose a Bayesian approach, named Bliss, to fit the functional linear regression model. The Bliss model supposes, through the prior, that the coefficient function is a step function. From the posterior, we propose several estimators to be used depending on the context: an estimator of the support and two estimators of the coefficient function: a smooth one and a stepwise one. To illustrate this, we explain the black PÃl'rigord truffle yield with the rainfall during the truffle life cycle. The Bliss method succeeds in selecting two relevant periods for truffle development.
As another feature of the Bayesian paradigm, the prior distribution enables the integration of preliminary judgments in the statistical inference. For instance, the biologists' knowledge about the truffles growth is relevant to inform the Bliss model. To this end, we propose two modifications of the Bliss model to take into account preliminary judgments. First, we indirectly collect preliminary judgments using pseudo data provided by experts. The prior distribution proposed corresponds to the posterior distribution given the experts' pseudo data. Futhermore, the effect of each expert and their correlations are controlled with weighting. Secondly, we collect experts' judgments about the most influential periods effecting the truffle yield and if the effect is positive or negative. The prior distribution proposed relies on a penalization of coefficient functions which do not conform to these judgments.
Lastly, the asymptotic behavior of the Bliss method is studied. We validate the proposed approach by showing the posterior consistency of the Bliss model. Using model-specific assumptions, efficient proof of the Wald theorem is given. The main difficulty is the misspecification of the model since the true coefficient function is surely not a step function. We show that the posterior distribution contracts on a step function which is the Kullback-Leibler projection of the true coefficient function on a set of step functions. This step function is derived from the true parameter and the design.
Keywords : Bayesian statistics, functional linear regression, high-dimensional statistics, sparsity, elicitation, prior information, pseudo data, penalization, asymptotic properties, posterior consistency, misspecification