Elicitation of Experts’ Knowledge for Functional Linear Regression

Grollemund, P.-M., Abraham, C., Pudlo, P. and Baragatti, M.
paul-marie.grollemund@umontpellier.fr

Abstract
We present an approach to elicit experts’ knowledge about the Bliss model [1] which is a parcimonious Bayesian Functional Linear Regression model. We derive an informative prior from elicited information and we define weights to tune prior information contribution on the estimators.

Sparse Step function
\[
\beta_0(t) = \sum_{k=1}^{K} b_k 1_{[a_k, b_k]}(t) \tag{1}
\]

Bliss
Bayesian Functional Linear Regression with Sparse Step functions aggregates "weak learners" in a Bayesian way.

Data: \( y_i, x_i(\cdot) \) where \( x_i(\cdot) \in L^2([0, 1]) \)

Model: see [2, 5]
\[
y_i|x_i(\cdot), \theta \sim \mathcal{N}\left( \mu + \int_0^1 x_i(t) \beta_0(t) \, dt, \sigma^2 \right) \tag{2}
\]
where \( \beta_0(t) \) is given in (1).

Noninformative Prior \( \pi_0(\cdot) \) and MCMC: [1]

Bayesian Estimator with \( L^2 \)-loss:
\[
\hat{\beta}(t) = \int \pi_0(\theta|D) \beta_0(t) \, d\theta \tag{3}
\]
where \( \pi_0(\cdot|D) \) is the posterior and \( \hat{\beta}_0 \) as in (1).

Elicitation
Aim:
- Collect experts’ knowledge [4] about the coefficient function
Various experts provide: \( D^e = (y^e_i, x^e_i(\cdot), \xi^e_i) \) for \( i = 1, \ldots, n_e \)
  - pseudo data: \( y^e_i \) and \( x^e_i(\cdot) \)
  - certainty: \( \xi^e_i \) for each pseudo observation

Informative prior
Informative prior is defined as a fractional posterior [3] for which:
- the initial prior is the vaguely noninformative prior \( \pi_0(\cdot) \) and
- the pseudo data model is (2)
\[
\pi_D(\theta) = \pi(\theta|D_1, \ldots, D^e; w) \propto \pi_0(\theta) \prod_{i=1}^{n_e} p(D^e|\theta; w) \quad \text{where} \quad p(D^e|\theta; w) = \prod_{i=1}^{n_e} p(y^e_i|x^e_i(\cdot), \theta)^{w^e_i}
\]
Interpretation: a sequential learning approach
- from initial prior, learn from pseudo data to derive an informative prior
- from informative prior, learn from observed data to derive a posterior

Tuning Weights
Naive approach: \( w^e = \xi^e_0 \)
- Overconfidence Bias, see [4]
Deriving weights from important properties:
- Experts’ interactions (\( \xi^e_{i,j} \in [0, 1] \) is the dependence between expert \( \epsilon \) and expert \( f )
- Pseudo data weight \( \leq \) observed data weight
\[
w^e_i = \xi^e_i \times \text{(interaction)} \times \text{(lower ps. data weight)} = \xi^e_i \times \frac{1}{1 + \sum_{f \neq \epsilon} r_{\epsilon,f} \times n_e \times E}
\]

Illustrations

Prior

Posterior

The marginal posterior distributions of \( \beta_0(t) \) for each \( t \) are represented using heat maps. Red (resp. white) colour is used to represent high (resp. low) posterior densities.

Prior expectation for each expert
Prior expectation
---
Posterior expectation
---
Posterior expectation without prior information
---

References