Supplementary Material:

Bayesian Approach using Experts' knowledge

Paul-Marie Grollemund [1], Christophe Abraham [2], Meïli Baragatti [2]

[1] Department of Mathematics, Imperial College London, London, United Kingdom

[2] MISTEA UMR 729, Montpellier SupAgro, INRA, CNRS, Univ Montpellier, Montpellier, France

---

Address for correspondence: Paul-Marie Grollemund, 6M30 Huxley Building South Kensington Campus, Department of Mathematics, Imperial College London, London, SW7 2AZ, United Kingdom.

E-mail: `pgrollem@ic.ac.uk`.

Phone: .

Fax: .

Abstract: +++ !!! +++ Abstracts must be given. +++ !!! +++

Key words: +++ !!! +++ Keywords must be given. +++ !!! +++

A Posterior expectation of $(\boldsymbol{b}|\boldsymbol{\mathcal{I}})$

Let $A = \text{RSS} + \sum_{e=1}^{E} \text{RSS}_e + \mu^2 v_0^{-1} + \boldsymbol{b}^T \boldsymbol{\Sigma}(\boldsymbol{\mathcal{I}})^{-1}\boldsymbol{b}$, which is the exponential term of the posterior density (without the factor $1/2\sigma^2$). We can rewrite A:

$$\left(n + \sum_{e=1}^{E}\sum_{i=1}^{n_e} w_i^e + v_0^{-1}\right)\left[\mu - \frac{\sum_{i=1}^{n}(y_i - \boldsymbol{x_i}(\boldsymbol{\mathcal{I}})^T\boldsymbol{b}) + \sum_{e=1}^{E}\sum_{i=1}^{n_e} w_i^e(y_i^e - \boldsymbol{x_i^e}(\boldsymbol{\mathcal{I}})^T\boldsymbol{b})}{n + \sum_{e=1}^{E}\sum_{i=1}^{n_e} w_i^e + v_0^{-1}}\right]^2 + B,$$

where

$$B = \boldsymbol{b}^T\boldsymbol{\Sigma}(\boldsymbol{\mathcal{I}})^{-1}\boldsymbol{b} + \sum_{i=1}^{n}(y_i - \boldsymbol{x_i}(\boldsymbol{\mathcal{I}})^T\boldsymbol{b})^2 + \sum_{e=1}^{E}\sum_{i=1}^{n_e} w_i^e(y_i^e - \boldsymbol{x_i^e}(\boldsymbol{\mathcal{I}})^T\boldsymbol{b})^2 - \\ \frac{\left(\sum_{i=1}^{n}(y_i - \boldsymbol{x_i}(\boldsymbol{\mathcal{I}})^T\boldsymbol{b}) + \sum_{e=1}^{E}\sum_{i=1}^{n_e} w_i^e(y_i^e - \boldsymbol{x_i^e}(\boldsymbol{\mathcal{I}})^T\boldsymbol{b})\right)^2}{n + \sum_{e=1}^{E}\sum_{i=1}^{n_e} w_i^e + v_0^{-1}}.$$

Then, the posterior distribution of $\boldsymbol{b}|\boldsymbol{\mathcal{I}}$ is obtained by integrating out μ and σ^2:

$$\iint \pi(\mu, \boldsymbol{b}, \sigma^2, \boldsymbol{\mathcal{I}}|\boldsymbol{y}, \boldsymbol{y_1}, \ldots, \boldsymbol{y_E}; w)\mathrm{d}\mu\mathrm{d}\sigma^2 \propto \int \left(\sigma^2\right)^{-\frac{1}{2}(n+\sum_{e=1}^{E}\sum_{i=1}^{n_e} w_i^e + K)-1} \exp\left\{-\frac{B}{2\sigma^2}\right\}\mathrm{d}\sigma^2 \\ \propto B^{-\frac{1}{2}(n+\sum_{e=1}^{E}\sum_{i=1}^{n_e} w_i^e + K)}.$$

which is proportional to a Student density.

As the observed data and the pseudo data are centered according to their relative weights, we obtain that for given intervals $\boldsymbol{\mathcal{I}} = (\mathcal{I}_1, \ldots, \mathcal{I}_K)$:

$$\sum_{i=1}^{n} y_i = 0 \text{ and } \sum_{i=1}^{n} x_i(\boldsymbol{\mathcal{I}}) = 0$$
$$\sum_{i=1}^{n_e} w_i^e y_i^e = 0 \text{ and } \sum_{i=1}^{n_e} w_i^e x_i^e(\boldsymbol{\mathcal{I}}) = 0.$$

Hence, we obtain:

$$B = \boldsymbol{b}^T\boldsymbol{\Sigma}(\boldsymbol{\mathcal{I}})^{-1}\boldsymbol{b} + \sum_{i=1}^{n}(y_i - \boldsymbol{x_i}(\boldsymbol{\mathcal{I}})^T\boldsymbol{b})^2 + \sum_{e=1}^{E}\sum_{i=1}^{n_e} w_i^e(y_i^e - \boldsymbol{x_i^e}(\boldsymbol{\mathcal{I}})^T\boldsymbol{b})^2$$
$$= \left(\boldsymbol{b} - \boldsymbol{M_w}^{-1}\left(\boldsymbol{x}(\boldsymbol{\mathcal{I}})\boldsymbol{y} + \sum_{e=1}^{E}\boldsymbol{x^e}(\boldsymbol{\mathcal{I}})\boldsymbol{W^e}\boldsymbol{y^e}\right)\right)^T \boldsymbol{M_w}\left(\boldsymbol{b} - \boldsymbol{M_w}^{-1}\left(\boldsymbol{x}(\boldsymbol{\mathcal{I}})\boldsymbol{y} + \sum_{e=1}^{E}\boldsymbol{x^e}(\boldsymbol{\mathcal{I}})\boldsymbol{W^e}\boldsymbol{y^e}\right)\right)$$

where h does not depend on $\boldsymbol{b}$. So, we identify the expectation: $\boldsymbol{M_w}^{-1}(\boldsymbol{x(\mathcal{I})y} + \sum_{e=1}^{E} \boldsymbol{x^e(\mathcal{I})W^e y^e})$.

B Details of implementation

The implementation of the methods is available on the following webpage:

https://github.com/pmgrollemund/blissElicitation.

2.1 Full conditional distributions

Below are given the full conditional distributions required for the sampler algorithm described in Section **??** or in Section 2.2 with more details. First, we consider the case of the first approach for which the prior is based on elicited pseudo data, described in Section **??**. Let $\boldsymbol{D}$ be the observed data set and $\boldsymbol{D^E} = (\boldsymbol{D_1}, \ldots, \boldsymbol{D_E})$ be the pseudo data set of E experts.

$$
\begin{aligned}
\mu|\boldsymbol{D},\boldsymbol{D^E},\boldsymbol{b},\sigma^2,\boldsymbol{m},\boldsymbol{\ell} &\sim \mathcal{N}\left(\frac{\boldsymbol{1_n}^T(\boldsymbol{y}-\boldsymbol{x}_{\boldsymbol{.}}(\boldsymbol{\mathcal{I}}_{\boldsymbol{.}})^T\boldsymbol{b})+\sum_{e=1}^E\boldsymbol{1_{n_e}}^T\boldsymbol{W}^e(\boldsymbol{y^e}-\boldsymbol{x^e_{.}}(\boldsymbol{\mathcal{I}}_{\boldsymbol{.}})^T\boldsymbol{b})}{n+v_0^{-1}+\sum_{e=1}^E n_e w^e},\right.\\
&\qquad\left.\frac{\sigma^2}{n+v_0^{-1}+\sum_{e=1}^E n_e w^e}\right)\\
\boldsymbol{b}|\boldsymbol{D},\boldsymbol{D^E},\mu,\sigma^2,\boldsymbol{m},\boldsymbol{\ell} &\sim \mathcal{N}_K\left(\hat{\boldsymbol{b}}_{\boldsymbol{1}}+\sum_{e=1}^E\hat{\boldsymbol{b}}_{\boldsymbol{2,e}}\,,\;\sigma^2\boldsymbol{M_w}^{-1}\right)\\
\sigma^2|\boldsymbol{D},\boldsymbol{D^E},\mu,\boldsymbol{b},\boldsymbol{m},\boldsymbol{\ell} &\sim \Gamma^{-1}\left(\frac{n+\sum_{e=1}^E n_e w^e+K+1}{2},\right.\\
&\qquad\left.\frac{\mathrm{RSS}+\sum_{e=1}^E\mathrm{RSS}_e}{2}+\frac{1}{2}(\mu,\boldsymbol{b})^T\underline{\mathrm{V}}^{-1}(\mu,\boldsymbol{b})\right),\\
\pi\left(m_k|\boldsymbol{D},\boldsymbol{D^E},\mu,\boldsymbol{b},\sigma^2,\boldsymbol{m_{-k}},\boldsymbol{\ell}\right) &\propto \exp\left(-\frac{\mathrm{RSS}+\sum_{e=1}^E\mathrm{RSS}_e}{2\sigma^2}\right)\times\pi_0(\boldsymbol{b}|\boldsymbol{m},\boldsymbol{\ell},\sigma^2)\\
\pi\left(\ell_k|\boldsymbol{D},\boldsymbol{D^E},\mu,\boldsymbol{b},\sigma^2,\boldsymbol{m},\boldsymbol{\ell_{-k}}\right) &\propto \exp\left(-\frac{\mathrm{RSS}+\sum_{e=1}^E\mathrm{RSS}_e}{2\sigma^2}\right)\times\pi_0(\ell_k)\times\pi_0(\boldsymbol{b}|\boldsymbol{m},\boldsymbol{\ell},\sigma^2)
\end{aligned}
$$

where $\boldsymbol{W^e}$ is a diagonal matrix whose the i^{th} element is w_i^e, $\mathrm{RSS}=\|\boldsymbol{y}-\mu\boldsymbol{1_n}-\boldsymbol{x}_{\boldsymbol{.}}(\boldsymbol{\mathcal{I}}_{\boldsymbol{.}})^T\boldsymbol{b}\|^2$, $\underline{\mathrm{V}}$ is given in the appendix of **?** and $\mathrm{RSS}_e=(\boldsymbol{y^e}-\mu\boldsymbol{1_{n_e}}-\boldsymbol{x^e_{.}}(\boldsymbol{\mathcal{I}}_{\boldsymbol{.}})^T\boldsymbol{b})^T\boldsymbol{W}^e(\boldsymbol{y^e}-\mu\boldsymbol{1_{n_e}}-\boldsymbol{x^e_{.}}(\boldsymbol{\mathcal{I}}_{\boldsymbol{.}})^T\boldsymbol{b})$.

Below are given the full conditional distributions concerning the approach based on the elicitation of features of the coefficient function, described in Section **??**.

$$
\begin{aligned}
\mu|\boldsymbol{D},\boldsymbol{b},\sigma^2,\boldsymbol{m},\boldsymbol{\ell} &\sim \mathcal{N}\left(\frac{\boldsymbol{1_n}^T(\boldsymbol{y}-\boldsymbol{x}_{\boldsymbol{.}}(\boldsymbol{\mathcal{I}}_{\boldsymbol{.}})^T\boldsymbol{b})}{n+v_0^{-1}},\frac{\sigma^2}{n+v_0^{-1}}\right)\\
\pi(\boldsymbol{b}|\boldsymbol{D},\mu,\sigma^2,\boldsymbol{m},\boldsymbol{\ell}) &\propto \exp\left\{-\left(\frac{\mathrm{RSS}+\boldsymbol{b}^T\boldsymbol{\Sigma}(\boldsymbol{\mathcal{I}})^{-1}\boldsymbol{b}}{2\sigma^2}+\tau\,\mathrm{dist}^2(s^\beta,\bar{s}^\beta_E;\bar{g}_E)\right)\right\}\\
\pi(\boldsymbol{m}|\boldsymbol{D},\mu,\boldsymbol{b},\sigma^2,\boldsymbol{\ell}) &\propto \exp\left\{-\left(\frac{\mathrm{RSS}}{2\sigma^2}+\tau\,\mathrm{dist}^2(s^\beta,\bar{s}^\beta_E;\bar{g}_E)\right)\right\}\\
\pi(\boldsymbol{\ell}|\boldsymbol{D},\mu,\boldsymbol{b},\sigma^2,\boldsymbol{m}) &\propto \exp\left\{-\left(\frac{\mathrm{RSS}}{2\sigma^2}+\tau\,\mathrm{dist}^2(s^\beta,\bar{s}^\beta_E;\bar{g}_E)\right)\right\}\pi_0(\boldsymbol{\ell})
\end{aligned}
$$

Concerning distribution of σ^2, we refer the reader to the appendix of **?**.

2.2 Sampler algorithm

In the following, we present the *Metropolis-Within-Gibbs* algorithm for sampling from the posterior distribution as described in Section **??**.

Algorithm : *Sampling from the posterior distribution*

- Determine a starting point $\boldsymbol{\theta_0}$ and τ.

- Determine the value of ρ.

  $1-i$ Propose a value for ρ.

  $1-ii$ Sample from the posterior distribution using the steps $3-i$ to $3-v$.

  $1-iii$ If the acceptance rate of the Metropolis step is between 0.3 and 0.5, continue otherwise, go to step $1-i$ and propose a new ρ.

- Compute approximations of the posterior expectations $E_{\text{RSS}} = \mathbb{E}\,\frac{1}{\sigma^2}\text{RSS}$ and $E_{\text{dist}} = \mathbb{E}\ \text{dist}^2(s^\beta, \bar{s}^\beta_E; \bar{g}_E)$ thanks to the posterior sample obtained in $1-ii$.

- Determine the vector $\boldsymbol{\tau}$ from 0 to $E_{\text{RSS}}/E_{\text{dist}}$.

- For each value of $\boldsymbol{\tau}$, do:

  $2-i$ Sample $\boldsymbol{\theta_t}$ for $t = 1, \ldots, T$ by executing the steps $3-i$ to $3-v$.

  $2-ii$ Compute the terms $p_{\boldsymbol{\theta_t},i}(y_i)$ with the Importance Sample, and

  $2-iii$ derive the utility $u_{\text{IS-LOO}}(\tau)$.

- Choose τ among $\boldsymbol{\tau}$ with the greater utility $u_{\text{IS-LOO}}(\tau)$.

- Sample from the posterior distribution. For i from 1 to N:

$3-i$ Simulate $\mu_i \sim \mu|\boldsymbol{y}, \boldsymbol{b_{i-1}}, \sigma^2_{i-1}, \boldsymbol{m_{i-1}}, \boldsymbol{\ell_{i-1}}; \tau$,

$3-ii$ Simulate $\sigma^2_i \sim \sigma^2|\boldsymbol{y}, \mu_i, \boldsymbol{b_{i-1}}, \boldsymbol{m_{i-1}}, \boldsymbol{\ell_{i-1}}; \tau$,

$3-iii$ Simulate $\boldsymbol{m_i} \sim \boldsymbol{m}|\boldsymbol{y}, \boldsymbol{\mu_i}, \boldsymbol{b_{i-1}}, \sigma^2_i, \boldsymbol{\ell_{i-1}}; \tau$,

$3-iv$ Simulate $\boldsymbol{\ell_i} \sim \boldsymbol{\ell}|\boldsymbol{y}, \mu_i, \boldsymbol{b_{i-1}}, \sigma^2_i, \boldsymbol{m_i}; \tau$,

$3-v$ Update $\boldsymbol{b}$ using a Metropolis step:

- $3-v-a$ Generate the proposal $\boldsymbol{b'} \sim \mathcal{N}_K(\boldsymbol{b_{i-1}}, \rho \mathbf{Id}_{\boldsymbol{K}})$.
- $3-v-b$ Compute the acceptance ratio α.
- $3-v-c$ Simulate $u \sim \mathrm{Unif}(0,1)$.
- $3-v-d$ If $u < \alpha$ then $\boldsymbol{b_i} := \boldsymbol{b'}$ else $\boldsymbol{b_{i-1}} := \boldsymbol{b'}$.