Semi-Modular Inference

Sitong Liu

February 25, 2022

1 Introduction

Semi-modular inference (SMI), proposed by Carmona and Nicholls (2020), is a modification of Bayesian inference when dealing with model misspecification. Compared with cut model, which completely eliminates the feedback from poorly specified modules, SMI moderates the influence of misspecified modules. It has been shown that SMI may do better than cut model and traditional Bayesian inference in some scenarios (Carmona and Nicholls, 2020). In this report, we will analyze the Biased data under semi-modular inference scheme.

2 Data and Model

In this section, we will introduce the Biased data that we use to analyze in this report.

The Biased data is a simple synthetic example where the "misspecification" comes from the poorly chosen prior. Suppose we have two datasets informing an unknown parameter φ. The first is a "reliable" small sample $Z = (Z_1, \dots, Z_m)$, $Z_i \sim N(\varphi, \sigma_z^2)$, iid for $i = 1, \dots, m$ distribution, with σ_z unknown; the second is a larger sample $Y = (Y_1, \dots, Y_n)$, $Y_i \sim N(\varphi + \theta, \sigma_y^2)$, iid for $i = 1, \dots, n$, with σ_y unknown. The "bias" θ is unknown.

This model was used by Liu et al. (2009), Jacob et al. (2017), and Carmona and Nicholls (2020). As assigned by Carmona and Nicholls (2020), we set m = 25, n = 50, the true generative parameters $\varphi^* = 0$, $\theta^* = 1$, and $\sigma_z = 2$, $\sigma_y = 1$; assign a constant prior for φ, and a conjugate prior $N(0, \sigma_\theta^2)$ for θ, where $\sigma_\theta = 0.33$. Note that σ_θ is assigned a different value from Carmona and Nicholls (2020) to generate a more extreme result.

3 Semi-Modular Inference

In this section, we will introduce the SMI distribution of the Biased data introduced in Section 2.

We introduce an auxiliary parameter $\tilde{\theta}$, which has the same distribution as that of θ. The **η-smi posterior** is

$$p_{smi,\eta}(\varphi, \theta, \tilde{\theta}|Z, Y) = p_{pow,\eta}(\varphi, \tilde{\theta}|Z, Y)p(\theta|Y, \varphi),$$

where $p_{pow,\eta}(\varphi, \tilde{\theta}|Z, Y)$ is the power posterior

$$p_{pow,\eta}(\varphi, \tilde{\theta}|Z, Y) \propto p(Z|\varphi)p(Y|\varphi, \tilde{\theta})^\eta p(\varphi, \tilde{\theta}).$$

The power $\eta \in [0, 1]$ is the degree of influence, controlling the contribution of the suspect module, i.e. $p(Y|\varphi, \theta)$ in this example, in the inference. η is selected beforehand and the selecting criteria will be introduced later.

In this example, since it involves Gaussian distribution, the η-smi posterior can be written in closed forms:

$$\begin{aligned} p_{pow,\eta}(\varphi,\tilde{\theta}|Z,Y) &\propto p(Z|\varphi)p(Y|\varphi,\tilde{\theta})^{\eta}p(\varphi,\tilde{\theta}) \\ &\propto \exp\{-\frac{1}{2}[\frac{m}{\sigma_z^2}(\bar{z}-\varphi)^2+\frac{n}{\sigma_y^2}(\bar{y}-(\varphi+\tilde{\theta}))^2\eta+\frac{1}{\sigma_{\tilde{\theta}}^2}\tilde{\theta}^2]\}; \end{aligned}$$

$$\begin{aligned} p(\theta|Y,\varphi) &\propto p(Y|\theta,\varphi)p(\theta|\varphi) \\ &\propto \exp\{-\frac{1}{2}[\frac{n}{\sigma_y^2}(\bar{y}-(\theta+\varphi))^2+\frac{1}{\sigma_\theta^2}\theta^2]\}; \end{aligned}$$

$$\begin{aligned} p_{smi,\eta}(\varphi,\theta,\tilde{\theta}|Z,Y) &= p_{pow,\eta}(\varphi,\tilde{\theta}|Z,Y)p(\theta|Y,\varphi) \\ &\propto \exp\{\frac{1}{2}[\frac{m}{\sigma_z^2}(\bar{z}-\varphi)^2+\frac{n}{\sigma_y^2}(\bar{y}-(\varphi+\tilde{\theta}))^2\eta+\frac{1}{\sigma_{\tilde{\theta}}^2}\tilde{\theta}^2+\frac{n}{\sigma_y^2}(\bar{y}-(\theta+\varphi))^2+\frac{1}{\sigma_\theta^2}\theta^2]\} \\ &\propto \exp\{\frac{1}{2}[\varphi^2(\frac{m}{\sigma_z^2}+\frac{n}{\sigma_y^2}(1+\eta))-2\varphi(\frac{m}{\sigma_z^2}\bar{z}+\frac{n}{\sigma_y^2}\bar{y}(1+\eta))+\theta^2(\frac{n}{\sigma_y^2}+\frac{1}{\sigma_\theta^2})-2\theta(\frac{n}{\sigma_y^2}\bar{y})+ \\ &\quad \tilde{\theta}^2(\frac{n}{\sigma_y^2}\eta+\frac{1}{\sigma_{\tilde{\theta}}^2})-2\tilde{\theta}(\frac{n}{\sigma_y^2}\bar{y}\eta)+2\varphi\theta\frac{n}{\sigma_y^2}+2\varphi\tilde{\theta}\frac{n}{\sigma_y^2}\eta]\}. \end{aligned}$$

where $\bar{Z}=\frac{1}{m}\sum_{i=1}^m Z_i$, $\bar{Y}=\frac{1}{n}\sum_{i=1}^n Y_i$.

Hence, the η-smi posterior is

$$p_{smi,\eta}(\varphi,\theta,\tilde{\theta}|Z,Y)=N((\varphi,\theta,\tilde{\theta});\mu,\Sigma), \quad (\varphi,\theta,\tilde{\theta})\in\Re^3$$

with

$$\Sigma=\begin{bmatrix} (\frac{m}{\sigma_z^2}+\frac{n}{\sigma_y^2}(1+\eta))^{-1} & \frac{\sigma_y^2}{n} & \frac{\sigma_y^2}{n\eta} \\ \frac{\sigma_y^2}{n} & (\frac{n}{\sigma_y^2}+\frac{1}{\sigma_\theta^2})^{-1} & 0 \\ \frac{\sigma_y^2}{n\eta} & 0 & (\frac{n}{\sigma_y^2}\eta+\frac{1}{\sigma_\theta^2})^{-1} \end{bmatrix}, \text{ and } \mu=\begin{bmatrix} \frac{\frac{m}{\sigma_z^2}\bar{Z}+\frac{n}{\sigma_y^2}\bar{Y}(1+\eta)}{\frac{m}{\sigma_z^2}+\frac{n}{\sigma_y^2}(1+\eta)} \\ \frac{\frac{n}{\sigma_y^2}\bar{Y}}{\frac{n}{\sigma_y^2}+\frac{1}{\sigma_\theta^2}} \\ \frac{\frac{n}{\sigma_y^2}\bar{Y}\eta}{\frac{n}{\sigma_y^2}\eta+\frac{1}{\sigma_\theta^2}} \end{bmatrix}.$$

The η-smi posterior for the original parameters is just the marginal, by integrating over $\tilde{\theta}$:

$$\begin{aligned} p_{smi,\eta}(\varphi,\theta|Z,Y) &= \int p_{smi,\eta}(\varphi,\theta,\tilde{\theta}|Z,Y)d\tilde{\theta} && (1) \\ &= p_\eta(\varphi|Z,Y)p(\theta|Y,\varphi), && (2) \end{aligned}$$

where the posterior for θ given φ is

$$p(\theta|Y,\varphi)=N(\theta;\mu_{\theta|Y,\varphi},\sigma^2_{\theta|Y,\varphi}), \tag{3}$$

with $\mu_{\theta|Y,\varphi}=\rho(\bar{Y}-\varphi)$, $\sigma^2_{\theta|Y,\varphi}=(1-\rho)\sigma_\theta^2$, $\rho=\frac{n\sigma_\theta^2}{n\sigma_\theta^2+\sigma_y^2}$.

The marginal posterior for φ is

$$p_\eta(\varphi|Z,Y)=N(\varphi;\mu_\eta,\sigma_\eta^2), \tag{4}$$

with $\mu_\eta=\lambda\bar{Z}+(1-\lambda)\bar{Y}$, $\sigma_\eta^2=\frac{\lambda\sigma_z^2}{m}$, $\lambda=\frac{m/\sigma_z^2}{m/\sigma_z^2+n/(\frac{\sigma_y^2}{\eta}+n\sigma_\theta^2)}$.

The derivation is as follows:

$$
\begin{aligned}
p_\eta(\varphi|Z,Y) &= \int p_{pow,\eta}(\varphi,\tilde{\theta}|Z,Y)d\tilde{\theta} \\
&\propto \int \exp\{-\frac{1}{2}[\frac{m}{\sigma_z^2}(\bar{z}-\varphi)^2 + \frac{n}{\sigma_y^2}(\bar{y}-(\varphi+\tilde{\theta}))^2\eta + \frac{1}{\sigma_{\tilde{\theta}}^2}\tilde{\theta}^2]\}d\tilde{\theta} \\
&\propto \exp\{-\frac{1}{2}[\frac{m}{\sigma_z^2}(\bar{z}-\varphi)^2 + \frac{n\eta}{\sigma_y^2}(\bar{y}-\varphi)^2]\} \int \exp\{-\frac{1}{2}[\frac{n\eta}{\sigma_y^2}\tilde{\theta}^2 - 2\frac{n\eta}{\sigma_y^2}(\bar{y}-\varphi)\tilde{\theta} + \frac{1}{\sigma_{\tilde{\theta}^2}}\tilde{\theta}^2]\}d\tilde{\theta} \\
&\propto \exp\{-\frac{1}{2}[\frac{m}{\sigma_z^2}(\bar{z}-\varphi)^2 + \frac{n\eta}{\sigma_y^2}(\bar{y}-\varphi)^2 - \frac{\frac{n^2\eta^2}{\sigma_y^4}(\bar{y}-\varphi)^2}{\frac{n\eta}{\sigma_y^2}+\frac{1}{\sigma_{\tilde{\theta}^2}}}]\} \\
&\qquad \int \exp\{-\frac{1}{2}(\frac{n\eta}{\sigma_y^2}+\frac{1}{\sigma_{\tilde{\theta}^2}})[\tilde{\theta}^2 - \frac{2\frac{n\eta}{\sigma_y^2}(\bar{y}-\varphi)}{\frac{n\eta}{\sigma_y^2}+\frac{1}{\sigma_{\tilde{\theta}^2}}}\tilde{\theta} + \left(\frac{\frac{n\eta}{\sigma_y^2}(\bar{y}-\varphi)}{\frac{n\eta}{\sigma_y^2}+\frac{1}{\sigma_{\tilde{\theta}^2}}}\right)^2]\}d\tilde{\theta} \\
&\propto \exp\{-\frac{1}{2}[\frac{m}{\sigma_z^2}(\bar{z}-\varphi)^2 + \frac{n\eta}{\sigma_y^2}(\bar{y}-\varphi)^2 - \frac{\frac{n^2\eta^2}{\sigma_y^4}(\bar{y}-\varphi)^2}{\frac{n\eta}{\sigma_y^2}+\frac{1}{\sigma_{\tilde{\theta}^2}}}] \int \exp\{-\frac{\left(\tilde{\theta} - \frac{\frac{n\eta}{\sigma_y^2}(\bar{y}-\varphi)}{\frac{n\eta}{\sigma_y^2}+\frac{1}{\sigma_{\tilde{\theta}^2}}}\right)^2}{2\frac{1}{\frac{n\eta}{\sigma_y^2}+\frac{1}{\sigma_{\tilde{\theta}^2}}}}\}d\tilde{\theta} \\
&\propto \exp\{-\frac{1}{2}[\varphi^2(\frac{m}{\sigma_z^2}+\frac{n\eta}{\sigma_y^2}-\frac{n^2\eta^2\sigma_{\tilde{\theta}^2}}{n\eta\sigma_y^2\sigma_{\tilde{\theta}^2}+\sigma_y^4}) - 2\varphi(\frac{m}{\sigma_z^2}\bar{z}+\frac{n\eta}{\sigma_y^2}\bar{y}-\frac{n^2\eta^2\sigma_{\tilde{\theta}^2}}{n\eta\sigma_y^2\sigma_{\tilde{\theta}^2}+\sigma_y^4}\bar{y})]\} \\
&\propto \exp\{-\frac{\left(\varphi - (\frac{m/\sigma_z^2}{m/\sigma_z^2+n/(\frac{\sigma_y^2}{\eta}+n\sigma_\theta^2)}\bar{z} + \frac{n/(\frac{\sigma_y^2}{\eta}+n\sigma_\theta^2)}{m/\sigma_z^2+n/(\frac{\sigma_y^2}{\eta}+n\sigma_\theta^2)}\bar{y})\right)^2}{2\frac{1}{m/\sigma_z^2+n/(\frac{\sigma_y^2}{\eta}+n\sigma_\theta^2)}}\} \\
&= N(\varphi;\lambda\bar{Z}+(1-\lambda)\bar{Y},\frac{\lambda\sigma_z^2}{m}).
\end{aligned}
$$

4 Choosing the Influence Parameter

In this section, we will introduce how to select the influence parameter η according to the *expected log pointwise predictive density (elpd)*:

$$elpd(\eta) = \int\int p^*(z,y)\cdot\log p_{smi,\eta}(z,y|Z,Y)dzdy,$$

where p^* is the distribution of the true data-generating process and

$$p_{smi,\eta}(z,y|Z,Y) = \int\int p(z,y|\varphi,\theta)\cdot p_{smi,\eta}(\varphi,\theta|Z,Y)d\varphi d\theta,$$

is a candidate posterior predictive distribution, indexed by η (Carmona and Nicholls, 2020).

Let (z,y) be a new sample from the smi-posterior, and write

$$
\begin{aligned}
z &= \mu_\eta + \tilde{\epsilon}\sigma_\eta + \epsilon_i\sigma_z \\
y &= \mu_{\theta|Y,\varphi} + \tilde{\epsilon}'\sigma_{\theta|Y,\varphi} + \mu_\eta + \tilde{\epsilon}\sigma_\eta + \epsilon_j\sigma_y,
\end{aligned}
$$

where $\tilde{\epsilon},\epsilon_i,\tilde{\epsilon}',\epsilon_j$ have standard normal distribution and are mutually independent.

Then we have

$$
\begin{aligned}
E(z) &= E(\mu_\eta + \tilde{\epsilon}\sigma_\eta + \epsilon_i\sigma_z) \\
&= \mu_\eta; \\
E(y) &= E(\mu_{\theta|Y,\varphi} + \tilde{\epsilon}'\sigma_{\theta|Y,\varphi} + \mu_\eta + \tilde{\epsilon}\sigma_\eta + \epsilon_j\sigma_y) \\
&= E(\mu_{\theta|Y,\varphi} + \mu_\eta) \\
&= \rho\bar{Y} + (1-\rho)\mu_\eta; \\
Var(z) &= Var(\mu_\eta + \tilde{\epsilon}\sigma_\eta + \epsilon_i\sigma_z) \\
&= \sigma_\eta^2 + \sigma_z^2; \\
Var(y) &= Var(\mu_{\theta|Y,\varphi} + \tilde{\epsilon}'\sigma_{\theta|Y,\varphi} + \mu_\eta + \tilde{\epsilon}\sigma_\eta + \epsilon_j\sigma_y) \\
&= (1-\rho)^2\sigma_\eta^2 + \sigma_{\theta|Y,\varphi}^2 + \sigma_y^2; \\
Cov(z,y) &= E[(z - E(z))(y - E(y))] \\
&= (1-\rho)\sigma_\eta^2.
\end{aligned}
$$

Hence, the predictive posterior is

$$p_{smi,\eta}(z, y|Z, Y) = N((z, y)^T; \mu_{zy}, \Sigma_{zy}), \tag{5}$$

with $\mu_{zy} = \begin{bmatrix} \mu_\eta \\ (1-\rho)\mu_\eta + \rho\bar{Y} \end{bmatrix}$, $\quad \Sigma_{zy} = \begin{bmatrix} \sigma_\eta^2 + \sigma_z^2 & (1-\rho)\sigma_\eta^2 \\ (1-\rho)\sigma_\eta^2 & (1-\rho)^2\sigma_\eta^2 + \sigma_{\theta|Y,\varphi}^2 + \sigma_y^2 \end{bmatrix}$.

Then elpd in this setting is

$$
\begin{aligned}
elpd(\eta) &= E_{p^*}(\log p_{smi,\eta}(z, y|Z, Y)) \\
&= -\log(2\pi) - \frac{1}{2}\log(\det(\Sigma_{zy})) - \frac{1}{2}E_{p^*}(((z, y) - \mu_{zy}^T)\Sigma_{zy}^{-1}((z, y)^T - \mu_{zy})) \\
&= -\log(2\pi) - \frac{1}{2}\log(\det(\Sigma_{zy})) - \frac{1}{2}(tr(\Sigma_{zy}^{-1}\Sigma^*) + (\mu - \mu_{zy}^T)\Sigma_{zy}^{-1}(\mu^* - \mu_{zy})),
\end{aligned}
$$

where $\mu^* = (\varphi^*, \theta^* + \varphi^*)^T$, and $\Sigma^* = \begin{bmatrix} \sigma_z^2 & 0 \\ 0 & \sigma_y^2 \end{bmatrix}$.

Here, we use the fact that

$$E_{p^*}(((z, y) - \mu_{zy}^T)\Sigma_{zy}^{-1}((z, y)^T - \mu_{zy})) = (tr(\Sigma_{zy}^{-1}\Sigma^*) + (\mu^{*T} - \mu_{zy}^T)\Sigma_{zy}^{-1}(\mu^* - \mu_{zy})),$$

which follows directly from Lemma 1.

Lemma 1 *If υ is a random vector, and A a commensurate matrix then*

$$E(\upsilon^T A\upsilon) = E(\upsilon)^T AE(\upsilon) + tr(Avar(\upsilon))$$

Proof:

$$
\begin{aligned}
E(\upsilon^T A\upsilon) &= E(tr(\upsilon^T A\upsilon)) \\
&= tr(AE(\upsilon\upsilon^T)) \\
&= tr(A(var(\upsilon) + E(\upsilon)E(\upsilon)^T)) \\
&= tr(E(\upsilon)^T AE(\upsilon)) + tr(Avar(\upsilon)) \\
&= E(\upsilon)^T AE(\upsilon) + tr(Avar(\upsilon)). \quad [EOP]
\end{aligned}
$$

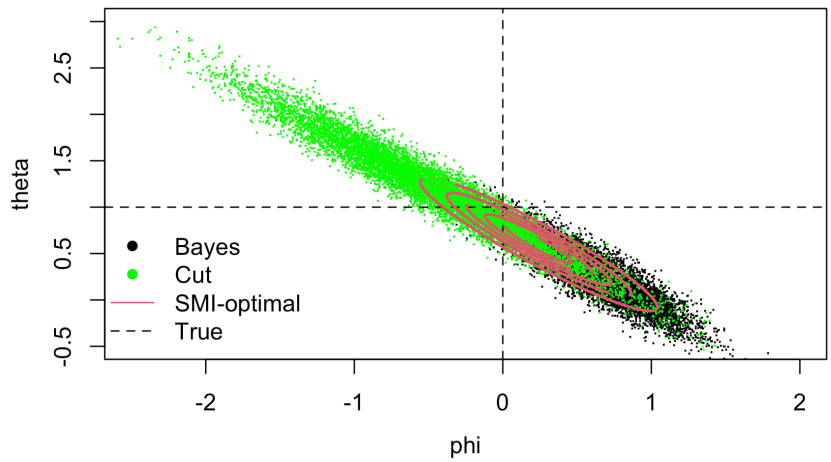

Figure 1: Candidate posterior distributions for normal biased data example. The points are samples from $p_{smi,\eta}(\varphi, \theta|Y, Z)$ for three values of $\eta = 0, \eta^*, 1$ yielding the Cut, optimal η-SMI, and Bayes posterior. The dotted lines show true parameter values.

5 Simulation

In this section, we will show the results following the formulas derived in previous section.

The Fig. 1 is the scatter plot of η-smi posterior at three values of $\eta = 0, \eta^*, 1$, which $\eta = 0$ is the cut model, η^* is the one that maximizes elpd(η), and $\eta = 1$ is the full Bayes model.

The Fig. 2 is the marginal η-SMI posteriors for φ (left) and θ (right) at $\eta = 0$ (Cut) and $\eta = 1$ (Bayes) together with the optimal η that maximizes the elpd.

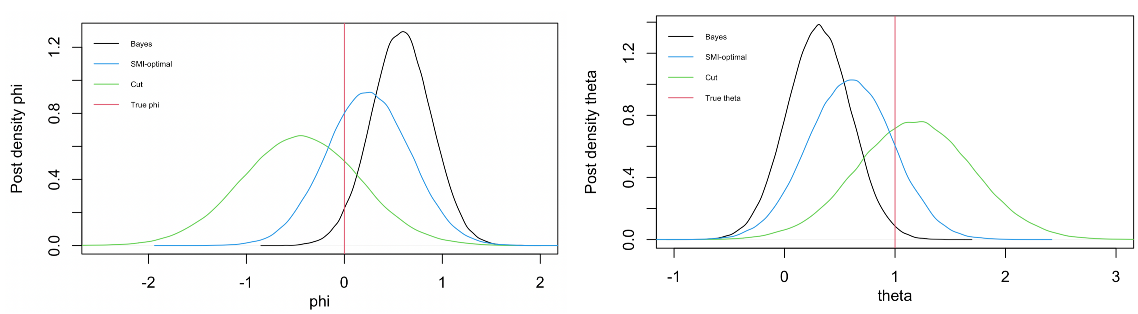

Figure 2: η-SMI posterior for φ (left) and θ (right) showing the Cut (green) and Bayes (black) and selected η^*-SMI posterior (blue) with the true parameter values indicated by a vertical line (red).

The Fig. 3 is the plot of elpd with varying $\eta \in [0, 1]$. The dotted line indicates the optimal η^*.

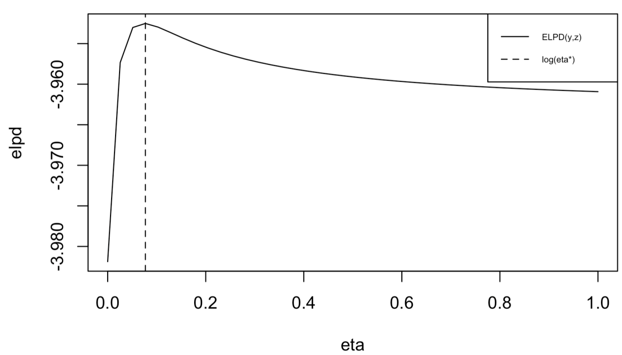

Figure 3: $elpd_{y,z}(Y, Z; \eta)$ as a function of η. the selected η^* is indicated by the vertical dashed line.