Discussion of

"Screening the Discrepancy Function of a Computer Model"

by Pierre Barbillon, Anabel Forte and Rui Paulo

Marilena Barbieri
Università Roma Tre

O'Bayes 2022: Objective Bayes Methodology Conference
UC Santa Cruz - September 6-10, 2022

Computer model Discrepancy (bias) function

$$\boldsymbol{y}(\boldsymbol{x}_i) = \boldsymbol{f}(\boldsymbol{x}_i, \boldsymbol{\theta}) + \boldsymbol{\delta}(\boldsymbol{x}_i) + \boldsymbol{\epsilon}_i \qquad i = 1, \ldots, n$$

Field data

$$\epsilon_i \sim N(0, \sigma_0^2)$$

$\boldsymbol{x}_i = (x_{1i}, \ldots, x_{pi})^T$ input variables

$$\delta(\cdot)|\sigma^2, \rho \sim GaSP(0, \sigma^2 c(\cdot, \cdot|\rho))$$

Separable correlation kernel: $c(\boldsymbol{x}_i, \boldsymbol{x}_j|\rho) = \prod_{i=1}^{p} c(x_{li}, x_{lj}|\rho_l)$

$c(x_{li}, x_{lj}|\rho_l) = \rho_l^{2^a |x_{li} - x_{lj}|^a}$ where $0 < a \leq 2$ is a fixed constant

$\rho_l \in [0, 1]$ If $\rho_l = 1$: x_l is "inhert"

? Which of the input variables is active in the discrepancy function?

Sampling distribution of the field data

$$\boldsymbol{y}|\boldsymbol{\rho},\sigma^2,\sigma_0^2,\boldsymbol{\theta}),f(\boldsymbol{x}) \sim N_n\left(f(\boldsymbol{\theta}),\sigma^2 R+\sigma_0^2 I_n\right)$$

$$R=[c(\boldsymbol{x}_i,\boldsymbol{x}_j|\boldsymbol{\rho})]_{i,j=1,\ldots,n}$$

Screening as a variable selection problem: models are indexed using

$$\boldsymbol{\gamma}=(\gamma_1,\ldots,\gamma_p)^T \in \{0,1\}^p$$

Under model M_l : $\gamma_l = 0$ ⟺ $\rho_l = 1$

Prior

$$\pi(\sigma^2, \sigma_0^2, \boldsymbol{\theta}, \boldsymbol{\rho}, \boldsymbol{\gamma}) = \pi(\sigma^2, \sigma_0^2, \boldsymbol{\theta})\pi(\boldsymbol{\rho}|\boldsymbol{\gamma})\pi(\boldsymbol{\gamma})$$

$$\pi(\boldsymbol{\rho}|\boldsymbol{\gamma}) = \prod_{l=1}^{p} \left[\gamma_l \, I_{(0,1)}(\rho_l) + (1-\gamma_l)\, B(\rho_l|\alpha_l, 1)\right]$$

Proposal:

- ✓ run an MCMC under $\boldsymbol{\gamma} = \mathbf{1}$
- ✓ obtain $\pi(\boldsymbol{\gamma}|\boldsymbol{y})$ using the importance sampling approximation

$$B_{\boldsymbol{\gamma}} \approx \frac{1}{M}\sum_{r=1}^{M} \pi(\boldsymbol{\rho}^{(r)}|\boldsymbol{\gamma})$$

a sample from the posterior distribution of $\boldsymbol{\rho}$ under $\boldsymbol{\gamma} = \mathbf{1}$

Posterior Inclusion Probabilities Screening (PIPS)

$$\pi(x_l|\boldsymbol{y}) = \sum_{\boldsymbol{\gamma}:\gamma_l=1} \pi(\boldsymbol{\gamma}|\boldsymbol{y})$$

? Use of the proposal in variable selection

? Which strategy do you suggest when p is large?

Proposal:

- ✓ run an MCMC under $\gamma = \mathbf{1}$
- ✓ obtain $\pi(\gamma|\boldsymbol{y})$ using the importance sampling approximation

$$B_{\boldsymbol{\gamma}} \approx \frac{1}{M} \sum_{r=1}^{M} \pi(\boldsymbol{\rho}^{(r)}|\boldsymbol{\gamma})$$

a sample from the posterior distribution of $\boldsymbol{\rho}$ under $\gamma = \mathbf{1}$

Posterior Inclusion Probabilities Screening (PIPS)

$$\pi(x_l|\boldsymbol{y}) = \sum_{\boldsymbol{\gamma}:\gamma_l=1} \pi(\gamma|\boldsymbol{y})$$

$$\delta(\cdot)|\sigma^2, \rho \sim GaSP(0, \sigma^2 c(\cdot, \cdot|\rho))$$

Separable correlation kernel: $$c(\boldsymbol{x}_i, \boldsymbol{x}_j|\rho) = \prod_{i=1}^{p} c(x_{li}, x_{lj}|\rho_l)$$

$$c(x_{li}, x_{lj}|\rho_l) = \rho_l^{2^a|x_{li}-x_{lj}|^a}$$

where $0 < a \leq 2$ is a fixed constant

? $a = 1.9$

Thank you!