Identifying important prior hyperparameters in Bayesian inverse problems with efficient variance-based global sensitivity analysis

John Darges

Department of Mathematics
North Carolina State University

April 24, 2023

Supported by NSF through awards DMS-1745654 and DMS 1953271.

Inverse problem

Consider the problem

$$\arg\min_{\boldsymbol{\theta}} J(\boldsymbol{\theta}), \quad J(\boldsymbol{\theta}) := \|\mathbf{B}\boldsymbol{y}(\boldsymbol{\theta}) - \mathbf{d}\|^2 \tag{1}$$

where $\boldsymbol{\theta}$ are unknown parameters, $\boldsymbol{y}$ is the solution to

$$\begin{cases} \boldsymbol{y}' = f(\boldsymbol{y};\boldsymbol{\theta}) \\ \boldsymbol{y}(t_0) = \boldsymbol{y}_0 \end{cases}, \quad \boldsymbol{y} \in \mathbb{R}^d \tag{2}$$

Observation operator $\mathbf{B}$ at each time selects only some of the responses from $\boldsymbol{y}(t_i)$ corresponding to data available in array $\mathbf{d}$

Bayesian context

Assume data measurement procedure involves i.i.d. random noise

$$y_i(\boldsymbol{\theta}) - d_i \sim \pi_{\text{noise}}$$

Noise model gives likelihood $\pi_{\text{like}}(\boldsymbol{\theta}) = \pi(\mathbf{d}|\boldsymbol{\theta})$

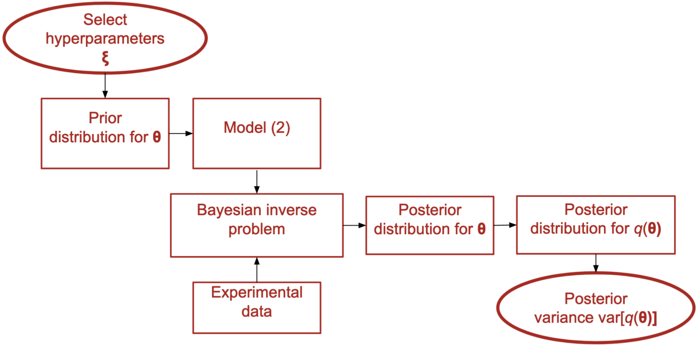
Incorporate prior beliefs/assumptions about $\boldsymbol{\theta}$ in prior distribution $\pi_{\text{prior}}(\boldsymbol{\theta})$

$$\textbf{Bayes' rule: } \pi_{\text{post}}(\boldsymbol{\theta}) = \frac{\pi_{\text{like}}(\boldsymbol{\theta})\pi_{\text{prior}}(\boldsymbol{\theta})}{\int_{\mathbb{R}^n} \pi_{\text{like}}(\boldsymbol{\theta})\pi_{\text{prior}}(\boldsymbol{\theta})d\boldsymbol{\theta}}$$

Posterior distribution in general constructed using Markov Chain Monte Carlo

GSA for Bayesian inverse problem

For Bayesian inference: Select a parameterized prior $\pi_{\text{prior}}(\boldsymbol{\theta}; \boldsymbol{\xi})$ with hyperparameters $\boldsymbol{\xi}$
Consider $F(\boldsymbol{\xi})$, which maps hyperparameters to a statistic of the posterior for QoI $q(\boldsymbol{\theta})$



Goal: Study sensitivity of $F(\boldsymbol{\xi})$ to prior hyperparameters

Variance-based global sensitivity analysis

- Consider a model $y = f(\boldsymbol{x})$ where $y \in \mathbb{R}$, and $\boldsymbol{x} \sim \pi(\boldsymbol{x})$ has independently distributed entries
- Sobol' indices are invaluable tools for GSA which measure the contribution of each input to variance in model output:

$$S_k := \frac{\mathrm{var}[f_k(x_k)]}{\mathrm{var}[f(\boldsymbol{x})]}, \quad S_k^T := 1 - \frac{\mathrm{var}[\mathbb{E}(f(\boldsymbol{x})|x_j,\ j \neq k)]}{\mathrm{var}[f(\boldsymbol{x})]}$$

- $f_k(x_k) := \int f(\boldsymbol{x}) d\boldsymbol{x}_{-k} - \mathbb{E}(f\boldsymbol{x})$, where $d\boldsymbol{x}_{-k}$ denotes integrating over all inputs **except** x_k
- First order Sobol' index S_k measures influence of x_k **outside of** interactions
- Total Sobol' index S_k^T measures influence of x_k **including** interactions with other inputs

Related Work

- Robust Bayesian analysis - tools to determine if posterior is robust to different priors in inference problems
- Hyper-differential sensitivity analysis (HDSA) has been used for Bayesian inverse problems to study measures of posterior uncertainty[1]
- Derivative-based global sensitivity measures (DGSM) has been used to study the sensitivity of information gain to uncertain model parameters [2]
- Variance-based GSA of function similar to $F(\xi)$. Emulated by Gaussian process, training data computed by MCMC [3]

[1] I. Sunseri, A. Alexanderian, J. Hart, B. van Bloemen Waanders. Hyper-differential sensitivity analysis for nonlinear Bayesian inverse problems. 2022.

[2] A. Chowdhary, A. Alexanderian. Sensitivity Analysis of the Information Gain in Infinite-Dimensional Bayesian Linear Inverse Problems. 2023.

[3] I. Vernon, J. P. Gosling. A Bayesian Computer Model Analysis of Robust Bayesian Analyses. 2022.

Method

- Most statistics or measures of uncertainty require estimating an integral
- Evaluating $F(\xi)$ using MCMC for different ξ is expensive
- **Note:** π_{like} does not depend on ξ
- **Question:** Can we re-use likelihood evaluations for different ξ's?
- Yes, but we have to be careful about what distribution we integrate over!

Importance Sampling

Consider integrating the following by Monte Carlo integration

$$F_{\mathrm{mean}}(\boldsymbol{\xi}) = \int_{\mathbb{R}^n} q(\boldsymbol{\theta})\pi_{\mathrm{post}}(\boldsymbol{\theta};\boldsymbol{\xi})d\boldsymbol{\theta}$$

Sampling uniformly over $\mathbb{R}^n$ will not work

Importance Sampling: Choose an auxiliary distribution π_{IS} to sample from

$$\int_{\mathbb{R}^n} q(\boldsymbol{\theta})\pi_{\mathrm{post}}(\boldsymbol{\theta};\boldsymbol{\xi})d\boldsymbol{\theta} = \int_{\mathbb{R}^n} q(\boldsymbol{\theta})\frac{\pi_{\mathrm{post}}(\boldsymbol{\theta};\boldsymbol{\xi})}{\pi_{\mathrm{IS}}(\boldsymbol{\theta})}\pi_{\mathrm{IS}}(d\boldsymbol{\theta}),$$

π_{IS} should be "close" to π_{post} for this to work

How to choose the auxiliary distribution

- For this scheme to work, π_{IS} should be "close" to $\pi_{\mathrm{post}}(\boldsymbol{\xi})$ for all $\boldsymbol{\xi}$
- Take $\pi_{\mathrm{IS}} \propto \pi_{\mathrm{like}}\pi_{\mathrm{prIS}}$ where π_{prIS} is same class distribution as priors
- Building IS sample set requires one MCMC run
- Find hyperparameters for π_{prIS} by minimizing the total KL-divergence[4]:

$$\boldsymbol{\xi}^* = \arg\min_{\boldsymbol{\xi}} \sum_{i=1}^{M} \int_{\mathbb{R}^n} \log\left(\frac{\pi_{\mathrm{IS}}(\boldsymbol{\theta};\boldsymbol{\xi})}{\pi_{\mathrm{post}}(\boldsymbol{\theta};\boldsymbol{\xi}_i)}\right)\pi_{\mathrm{IS}}(\boldsymbol{\theta};\boldsymbol{\xi})d\boldsymbol{\theta}$$

[4] J. Zhang, M.D. Shields. On the quantification and efficient propagation of imprecise probabilities resulting from small datasets. 2017.

Importance sampling-Monte Carlo estimator

Estimate $F_{\mathrm{mean}}(\boldsymbol{\xi})$ by Monte Carlo integration:

$$\int_{\mathbb{R}^n} q(\boldsymbol{\theta}) \frac{\pi_{\mathrm{post}}(\boldsymbol{\theta};\boldsymbol{\xi})}{\pi_{\mathrm{IS}}(\boldsymbol{\theta})} \pi_{\mathrm{IS}}(d\boldsymbol{\theta}) = \frac{1}{\int_{\mathbb{R}^n} \frac{\pi_{\mathrm{pr}}(\boldsymbol{\theta};\boldsymbol{\xi})}{\pi_{\mathrm{prIS}}(\boldsymbol{\theta})} \pi_{\mathrm{prIS}}(d\boldsymbol{\theta})} \int_{\mathbb{R}^n} q(\boldsymbol{\theta}) \frac{\pi_{\mathrm{pr}}(\boldsymbol{\theta};\boldsymbol{\xi})}{\pi_{\mathrm{prIS}}(\boldsymbol{\theta})} \pi_{\mathrm{prIS}}(d\boldsymbol{\theta})$$

$$\approx \frac{1}{\sum_{j=1}^{N} \frac{\pi_{\mathrm{pr}}(\boldsymbol{\theta}_j;\boldsymbol{\xi})}{\pi_{\mathrm{prIS}}(\boldsymbol{\theta}_j)}} \sum_{j=1}^{N} q(\boldsymbol{\theta}_j) \frac{\pi_{\mathrm{pr}}(\boldsymbol{\theta}_j;\boldsymbol{\xi})}{\pi_{\mathrm{prIS}}(\boldsymbol{\theta}_j)}, \quad \boldsymbol{\theta}_j \sim \pi_{\mathrm{IS}}$$

When we change $\boldsymbol{\xi}$, now we only need to re-evaluate the prior distribution

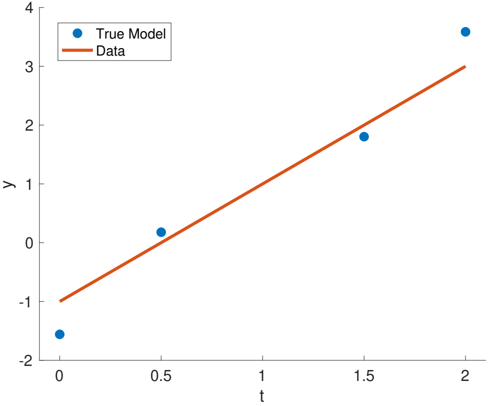
Proposed Method

1. Build ISMC sample set $\{\boldsymbol{\theta}_j\}_{j=1}^N$ by sampling from π_{IS} using MCMC
2. Evaluate the QoI $q(\boldsymbol{\theta})$ at the sample points
3. For hyperparameter samples $\{\boldsymbol{\xi}_i\}_{i=1}^M$, evaluate π_{pr} at samples $\{\boldsymbol{\theta}_j\}_{j=1}^N$
4. Estimate by Monte Carlo integration $\{F(\boldsymbol{\theta}_j)\}_{j=1}^N$
5. Use $\{F(\boldsymbol{\theta}_j)\}_{j=1}^N$ to estimate Sobol' indices (surrogate-assisted or sampling)

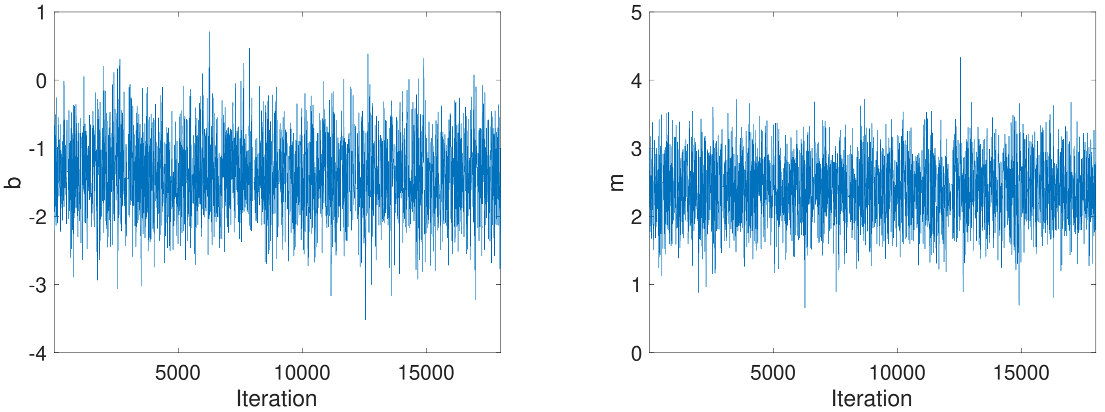
Assuming the likelihood is expensive to evaluate, the costliest step is the first one

Example: Fitting Noisy Data to a Line

- Fit data to $y = mx + b$
- Estimate $\boldsymbol{\theta} = (m, b)$
- $b = -1,\ m = 2$
- Noise is i.i.d. normally distributed with $\sigma^2 = 1$
- Likelihood and prior are Gaussian $\implies$ posterior is Gaussian and can be analytically computed



Parameter estimation with MCMC



Problem Setup

- Prior is Gaussian with $\boldsymbol{\xi} = (\mu, \Gamma)$
- Nominal $\mu = \begin{bmatrix} 2.4 \\ -1.4 \end{bmatrix}$
- Nominal $\Gamma = \begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix}$
- We let the hyperparameters vary by $\pm 50\%$ of the respective nominal value

- Consider two different QoIs
- Linear: $q(\boldsymbol{\theta}) = \boldsymbol{\theta}^\top \begin{bmatrix} 5 \\ 1 \end{bmatrix}$
- Nonlinear: $q(\boldsymbol{\theta}) = \boldsymbol{\theta}^\top \boldsymbol{\theta}$
- We are interested in both the posterior means and variances
- These can be analytically computed for both QoIs

ISMC total convergence for linear QoI

Do estimates of $F(\xi)$ converge on average for different choices of ξ?

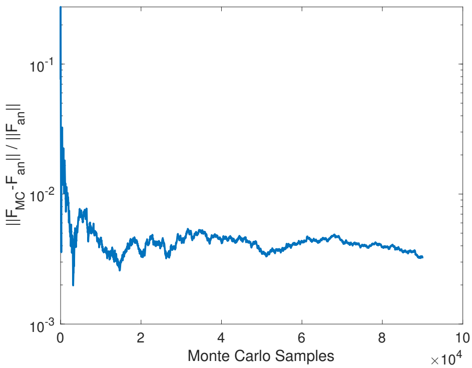


Figure: $F_{\mathrm{mean}}(\xi)$ for the linear QoI

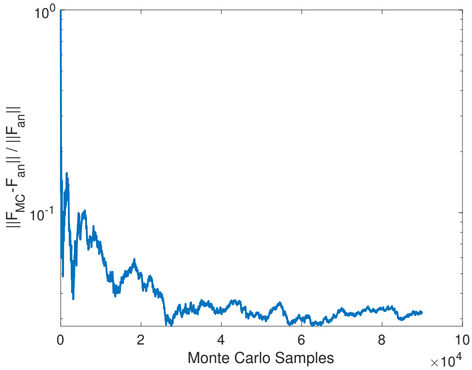


Figure: $F_{\mathrm{var}}(\xi)$ for the linear QoI

ISMC GSA for linear QoI

We use a polynomial chaos surrogate model and compare to true indices

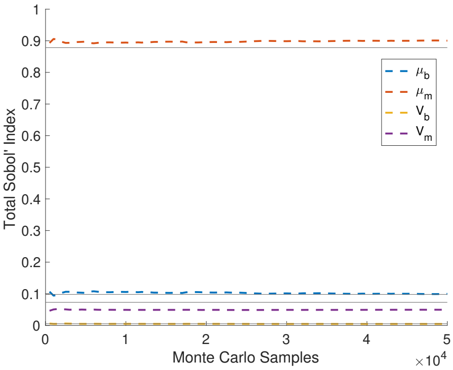


Figure: Total indices of $F_{\mathrm{mean}}(\xi)$ for the linear QoI

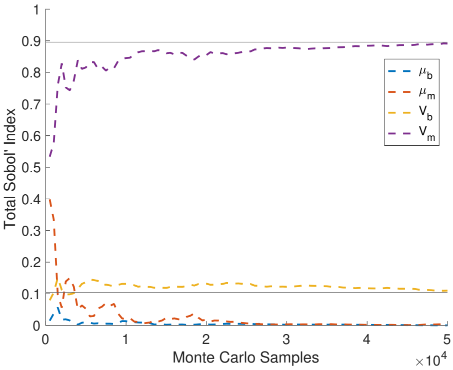


Figure: Total indices of $F_{\mathrm{var}}(\xi)$ for the linear QoI

ISMC total convergence for nonlinear QoI

Do estimates of $F(\xi)$ converge on average for different choices of ξ?

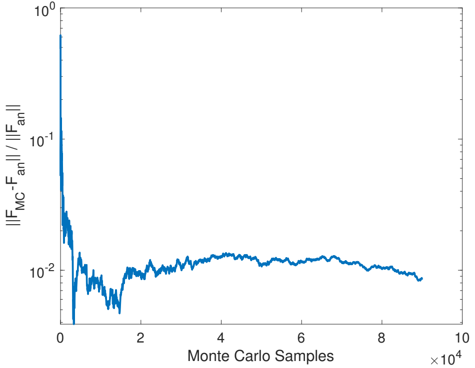


Figure: $F_{\mathrm{mean}}(\xi)$ for the nonlinear QoI

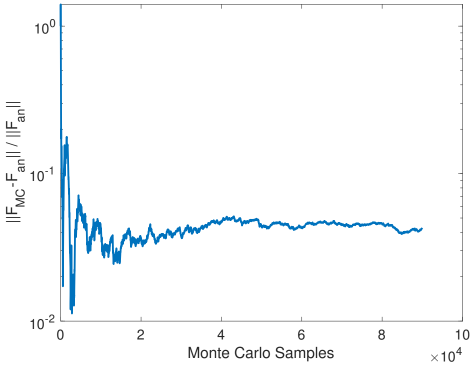


Figure: $F_{\mathrm{var}}(\xi)$ for the nonlinear QoI

ISMC GSA for nonlinear QoI

We use a polynomial chaos surrogate model and compare to true indices

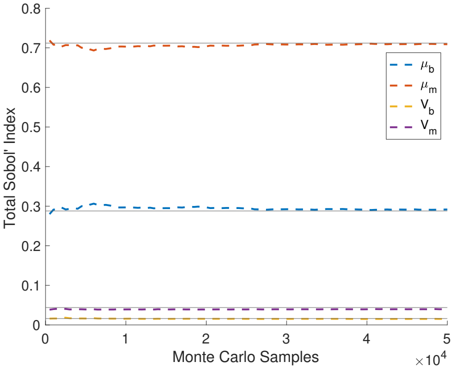


Figure: Total indices of $F_{\mathrm{mean}}(\xi)$ for the nonlinear QoI

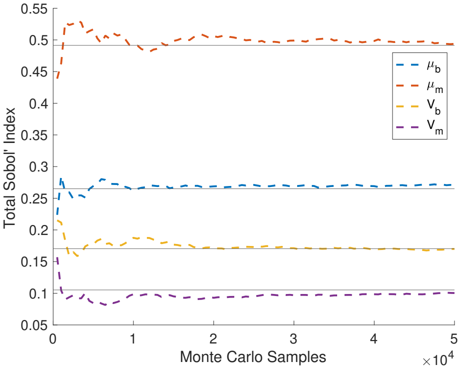


Figure: Total indices of $F_{\mathrm{var}}(\xi)$ for the nonlinear QoI

Nonlinear Bayesian inverse problems

Applying the method to nonlinear problems introduces more challenges:

- Model is a "black-box"
- Nothing to compare our results against
- Efficiently sampling from π_{IS} by MCMC could be difficult
- Importance sampling could fail for some priors

References

1. I. Sunseri, A. Alexanderian, J. Hart, B. van Bloemen Waanders. Hyper-differential sensitivity analysis for nonlinear Bayesian inverse problems. 2022.
2. A. Chowdhary, A. Alexanderian. Sensitivity Analysis of the Information Gain in Infinite-Dimensional Bayesian Linear Inverse Problems. 2023.
3. I. Vernon, J. P. Gosling. A Bayesian Computer Model Analysis of Robust Bayesian Analyses. 2022.
4. J. Zhang, M.D. Shields. On the quantification and efficient propagation of imprecise probabilities resulting from small datasets. 2017.
5. S. Tokdar, R. Kass, Importance sampling: A review. 2010.
6. J. O. Berger, D. R. Insua, F. Ruggeri. Bayesian Robustness. 2000.