Introduction

Bayesian experimental design (BED) is to choose designs that maximize the information gathering. For implicit models, where the likelihood is intractable, but sampling is possible, conventional BED methods have difficulties in efficiently estimating the posterior distribution and maximizing the mutual information (MI) between data and parameters.

Recent work using gradient ascent to maximize a lower bound on the mutual information is usually inaccessible for implicit models.

• We propose a general unified framework that leverages stochastic approximate gradient without the requirement or assumption of pathwise gradients for implicit models.

• We introduce a smoothed MI lower bound to conduct robust MI estimation and optimization, which allows the variance of the design and posterior distribution to be much smaller than existing approaches.

• We show the superior performance of the approach through several experiments and demonstrate that the approach enables the optimization to be performed by stochastic gradient ascent algorithm and thus well suited to considerable high dimensional design problems.

Bayesian experimental design (BED)

BED framework aims at choosing an experimental design $\xi$ to maximize the information gained about some parameters of interest $\theta$ from the outcome $y$ of the experiment.

Expected information gain: $I(\xi) = E_p[y|x][\mathcal{I}[p(\theta)] - \mathcal{I}[p(\theta|y, \xi)]]$

$EIG$ can be interpreted as a mutual information between $\theta$ and $y$.

$\mathcal{I}(\theta, \xi) = E_p[y|x][\log p(y|\theta, \xi) / p(y|x)]$

$\xi$ estimate is the core

Bayesian optimal design is defined as $\xi^* = \arg\max_\xi I(\xi)$

Conventional approach: nested Monte Carlo (NMC) estimator,

$I_{NMC}(\xi) = \frac{1}{N} \sum_{i=1}^{N} \log \frac{p(y|\theta_i, \xi)}{\sum_{j=1}^{N} p(y|\theta_j, \xi)}$

Computationally intensive!

Stochastic Approximate Gradient Ascent BED

Belghazi et al. (2018) proposed to estimate MI by gradient ascent over neural networks and argued that the lower bound can be tightened by optimizing the neural network parameters, $\text{MINE}_{\text{BED}}(\xi, \psi) = E_p(y|x)[\mathcal{I}_\theta(\theta, y)] - \log E_p(y|x)e^{\mathcal{I}_\theta(\theta, y)}$

BED problem can be formulated by maximizing the overall objective $\xi^* = \arg\max_\xi I_{\text{NMC}}(\xi, \psi)$. 

Smoothed MI estimator: reduce the variance with clip operation $I_{\text{SMILE}}(\xi, \psi) = E_p(y|x)[\mathcal{I}_\theta(\theta, y)] - \log E_p(y|x)[\text{clip}(e^{\mathcal{I}_\theta(\theta, y)}, e^{-\tau}, e^\tau)]$

Stochastic gradient approximate: evolution strategies $f_\alpha(\xi) = E_{\xi \sim \mathcal{N}(\theta_0, \sigma)}[f(\xi + \sigma \epsilon)]$

$V_\alpha(\xi) = \frac{1}{\sigma^2} E_{\xi \sim \mathcal{N}(\theta_0, \sigma)}[f(\xi + \sigma \epsilon) - f(\xi)]$

$\xi_1 = \xi + V_{\alpha T}^\top f_\alpha(\xi)$

Guided ES algorithm

$\Sigma = (\alpha/n) \cdot I_0 + (1-\alpha)/k - U U^\top T^2$

SAGABED for implicit models

• Unified framework vs. two-stage framework

• Scalability, portability, and parallelization

• Robust estimation with low variance

Experiments: toy examples

Noisy linear regression

Pharmacokinetic model

Conclusion and future work

Develop a general unified framework that utilizes the stochastic approximate gradient for BED with implicit models. Our approach allows to scaled to substantial high dimensional design problems. The future work will focus on the extension of our proposed framework to sequential Bayesian optimization design (SBED).