Model-based optimal design of experiments with Pumas: `OptimalDesign.jl`

Jose Storopoli and Mohamed Tarek `[jose.storopoli|mohamed]@pumas.ai`
PumasAI

Outline

1. Motivation
2. Fisher Information Matrix – FIM
3. Optimal Design Objective
4. Optimal Design in Pumas
5. Selected Literature

Model-based

- M – parametric model.
- $\theta \in \Theta$ – model parameters.
- (M, Θ) – hypothesis.

Why Optimal Design?

- Which model M best describes the drug's effect?
- Which parameters θ are the best estimates?
- How to quantify the uncertainty in those θ?

Which model M best describes the drug's effect?

Not committing to a particular model M and instead collecting data that would let us learn the parameter values of multiple models $M_1, M_2, \ldots, M_n$ simultaneously.

Which parameters θ are the best estimates?

Not committing to particular parameter values θ_i for each model M_i and instead using a set of values $\Theta_i : \theta_i \in \Theta_i$, e.g. a discrete set of specific parameter values or a continuous set.

pumas

How to quantify the uncertainty in those θ?

Using the expected Fisher information matrix (FIM) instead of the observed one to estimate the expected standard errors at each parameter value $\theta_i \in \Theta_i$.

Fisher Information Matrix – FIM

- **Fisher information** is a way of measuring the amount of information that an observable random variable X carries about an unknown parameter θ upon which the probability of X depends.
- Formally, it is the expected value of the observed information, which in turn is the **negative of the second derivative of the loglikelihood**.
- if θ is *not* a scalar, then the information is expressed as a matrix, FIM, with the second derivative becoming the **Hessian matrix**:

$$-\operatorname{E}_{p(x|\theta)}\left[\mathbf{H}_{\log p(x|\theta)}\right]$$

pumas

Fisher Information Matrix Properties

- $N \times N$ positive semidefinite.
- symmetric, if second partial derivatives are all continuous.

Optimal Design Objective

There are a number of possible objectives that correspond to maximizing the information learned in the optimal experiment design:

- **A-optimality**:
  - **minimizing** the **trace** of the inverse of the expected FIM.
  - minimizing the sum of the expected standard errors.
- **D-optimality**:
  - **maximizing** the **determinant** of the expected FIM.
  - maximizing the product of the eigenvalues of the expected FIM.
  - which indirectly minimizes the expected standard errors.
- **T-optimality**:
  - **maximizing** the **trace** of the expected FIM.
  - maximizing the sum of the eigenvalues of the expected FIM.
  - which is also roughly correlated to minimizing the sum of the expected standard errors but giving more weight to the parameters with smaller standard errors.

Optimal Design in Pumas

- There are a number of degrees of freedom that an experiment designer can control when designing an experiment.
- This leads to different types of optimization problems, or so-called optimal design tasks.
- Currently[1], the main optimal design task supported in Pumas is the **sample time optimization**.

[1] Pumas version `2.2`

Sample Time Optimization

The following are assumed to be **fixed**:

- **model and parameter values**, e.g. from a similar study.
- **subjects' covariates**, a.k.a. subject templates, e.g. typical values in the target population.
- **number of replicas of each subject template**, using best practices in randomized sampling.
- **dosing regimen of each subject**, e.g. from initial simulations to avoid toxicity.
- **number of observations per subject template**.

The **only degree of freedom allowed to change is which times the observations are to be made for each subject template**.

Sample Time Optimization

The same optimal design task may be **repeated for different values of the other fixed degrees of freedom**, e.g. different dosing regimens or number of subjects/observations, to find a satisfactory design.

This parametric study is a form of naive **search-based, bi-level optimization**[2].

[2] Bilevel optimization is a special kind of optimization where one problem is embedded (nested) within another. The outer optimization task is commonly referred to as the upper-level optimization task, and the inner optimization task is commonly referred to as the lower-level optimization task.

Sample Time Optimization – Constraints

Some possible **constraints** to consider when doing sample time optimization are:

- **lower and upper bounds on the sample times**, e.g. the start and end date of the data collection part of the study.
- **minimum offset between any two consecutive observations**.
- **time window constraints**, e.g. the working hours of the clinical staff.
- **maximum and/or minimum number of measurements per time window**.

Selected Literature I