pumas

ACoP12 Poster #: STPM247

pumas AI

Nonlinear mixed effects model based optimal design of experiments using mathematical programming in Pumas

Mohamed Tarek[1], Joakim Nyberg[2], Vijay Ivaturi[1,3], Andreas Noack[1,4]

[1]PumasAI Inc., ; [2]Pharmetheus AB, Uppsala, Sweden, [3]University of Maryland Baltimore, Baltimore, MD; [4]JuliaComputing Inc.

Introduction

A clinical trial is a critical stage of every drug development process before the drug can get approved for public use. Clinical trials for new drugs are generally both expensive and risky. Each clinical trial or experiment in general is designed with a goal in mind. The most common goal of clinical trials is to maximize the information learned about the model's parameters, e.g. by minimizing the expected standard errors.

There are a number of degrees of freedom that an experiment designer can control when designing an experiment. For simplicity, some may often be fixed using heuristics and best practices while others are allowed to change at any one time. This leads to different types of optimization problems. In this study, the optimal design task of interest is sample time optimization, where every aspect of the experiment is assumed to be known except when the sample measurements are going be conducted.

Mathematical programming [1] is a powerful family of techniques that includes: the formulation of complex decision problems using mathematical objectives and constraints, as well as a set of powerful and generic optimization algorithms that can be used to solve the decision problems formulated.

In this study, state-of-the-art mathematical programming is used to optimally design experiments based on a nonlinear mixed effects model (NLME) in Pumas [2].

Methods

In this study, 2 mathematical programming formulations for optimal design of sample times in clinical trials based on NLME models in Pumas are presented. The formulations presented encode the objective as well as a number of practical, logistic constraints, such as a minimum duration between 2 consecutive samples and multiple disjoint time windows. The formulations are solved using nonlinear programming (NLP) and mixed integer nonlinear programming (MINLP) algorithms in the IPOPT [3] and Juniper [4] software packages. Distributed parallelism using Pumas Enterprise powered by JuliaHub is then used to demonstrate the scalability of the MINLP formulation and algorithm used. D-optimality is demonstrated and compared against the popular optimal design package PopED [5]. The optimal design user interface in Pumas 2.0 provides a natural and user-friendly workflow for model definition, fitting and optimal design of experiments. Work shifts and days of hospital personnel can also be directly incorporated into the design by inputting the time windows using dates and times rather than floating point numbers.

The following logistic constraints are all supported in Pumas 2.0:

- 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 2 consecutive observations because depending on how the measurement is taken it may not be possible to take 2 measurements within a few seconds of each other for example.
- Time window constraints, e.g. the working hours of the clinical staff.
- Maximum number of measurements per time window because there may not be enough clinical staff or measuring devices to take more than a certain number of measurements per hour.
- Minimum number of measurements per time window to ensure the subjects are observed at least once every so long.

Methods

Formulation I: fixed number of observations per time window

$$
\begin{aligned}
\underset{t}{\text{maximize}} \quad & \det\left(\sum_{i=1}^{I} N_i F_i(t)\right) && && \text{(1a)}\\
\text{subject to} \quad & t_{i,j,w} + O \le t_{i,j+1,w}, && i \in [1,I], j \in [1, J_w - 1], w \in [1,W], && \text{(1b)}\\
& t_{i,j,w} \ge S_w, && i \in [1,I], j \in [1, J_w], w \in [1,W], && \text{(1c)}\\
& t_{i,j,w} \le E_w, && i \in [1,I], j \in [1, J_w], w \in [1,W] && \text{(1d)}
\end{aligned}
$$

Let:

1. $t_{i,j,w}$ be the time of the j^{th} observation of the i^{th} subject template in the w^{th} time window for $i \in [1, I], j \in [1, J_w], w \in [1, W]$ where I is the number of subject templates, J_w is the total number of observations per subject template for time window w and W is the number of time windows.
2. S_w be the start time of time window $w \in [1, W]$.
3. E_w be the end time of time window $w \in [1, W]$.
4. N_i be the number of replicas of subject template i for $i \in [1, I]$
5. O be the minimum duration difference between any 2 consecutive observations.
6. $F_i(t)$ be the expected FIM of subject template i given all the observations which is a function of $t = \{t_{i,j,w}\}$, the model, parameter values, the subject's covariates and dosing regimen.

Formulation II: total number of observations for all time windows

$$
\begin{aligned}
\underset{t,\,x}{\text{maximize}} \quad & \det\left(\sum_{i=1}^{I} N_i F_i(t)\right) && && \text{(2a)}\\
\text{subject to} \quad & \sum_{w=1}^{W} x_{i,j,w} = 1, && \forall i \in [1,I], j \in [1,J], && \text{(2b)}\\
& \sum_{i=1}^{I}\sum_{j=1}^{J} x_{i,j,w} \le MAX_w, && \forall w \in [1,W], && \text{(2c)}\\
& \sum_{i=1}^{I}\sum_{j=1}^{J} x_{i,j,w} \ge MIN_w, && \forall w \in [1,W], && \text{(2d)}\\
& t_{i,j} + O \le t_{i,j+1}, && \forall i \in [1,I], j \in [1, J-1], && \text{(2e)}\\
& t_{i,j} \le \sum_{w=1}^{W} E_w x_{i,j,w}, && \forall i \in [1,I], j \in [1,J], && \text{(2f)}\\
& t_{i,j} \ge \sum_{w=1}^{W} S_w x_{i,j,w}, && \forall i \in [1,I], j \in [1,J], && \text{(2g)}\\
& x_{i,j,w} \le \sum_{w'=w}^{W} x_{i,j',w'}, && \forall i \in [1,I], j \in [1,J], j' \in [j+1, J], w \in [1,W], && \text{(2h)}\\
& x_{i,j,w} \le \sum_{w'=1}^{w} x_{i,j',w'}, && \forall i \in [1,I], j \in [1,J], j' \in [1, j-1], w \in [1,W], && \text{(2i)}\\
& x_{i,j,w} \in \{0,1\} && \forall i \in [1,I], j \in [1,J], w \in [1,W], && \text{(2j)}\\
& L \le t_{i,j} \le U && \forall i \in [1,I], j \in [1,J] && \text{(2k)}
\end{aligned}
$$

Methods

Let:

1. $t_{i,j}$ be the time of the j^{th} observation of the i^{th} subject template for $i \in [1, I], j \in [1, J]$ where I is the number of subject templates and J is the total number of observations per subject template.
2. $x_{i,j,w}$ be 1 iff time $t_{i,j}$ is in the w^{th} time window and 0 otherwise, for $i \in [1, I], j \in [1, J], w \in [1, W]$.
3. S_w be the start time of time window $w \in [1, W]$.
4. E_w be the end time of time window $w \in [1, W]$.
5. N_i be the number of replicas of subject template i for $i \in [1, I]$
6. MIN_w and MAX_w be the maximum number of observations in time window w for all the subject templates combined.
7. L and U be the minimum and maximum times of all the time windows.
8. O be the minimum duration difference between any 2 consecutive observations.
9. $F_i(t)$ be the expected FIM of subject template i given all the observations which is a function of $t = \{t_{i,j}\}$, the model, parameter values, the subject's covariates and dosing regimen.

Experiments and results

Pumas vs PopED benchmark

A benchmark of Pumas and the well-known software package PopED [6] for optimal design will be performed. The following are the specifications of the test problem used:

- Model: IVGTT insulin model [6,7]
- # of unique sample times: 10
- Total number of sample times (with repeats): 29
- Objective: logdet of the expected FIM
- # of unique subjects: 1
- # of subjects (with repeats): 42
- Time window: 0..240

Table 1: log determinant of the expected FIM using PopED and Pumas for D-optimal design of sample times in the IVGTT model.

	log determinant of expected FIM	
	PopED	Pumas
Initial value	133.5	133.5
After 1 hour	138.6	163.1

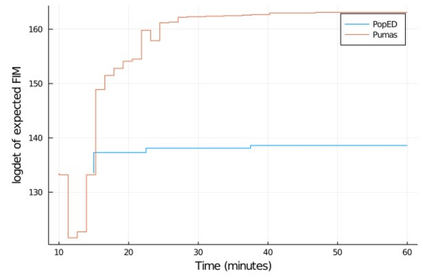

Figure 1: trace plot of the objective function value of PopED vs Pumas for 1 hour.

Distributed parallelism in Pumas

Another experiment design task with multiple time windows and some logistic constraints was solved. The following are the specifications of the optimal design task:

- Model: two compartment PK model
- # of unique sample times per subject: 15
- # of unique subjects: 5
- # of subjects (with repeats): 15 (each unique subject repeated 3 times)
- Minimum offset between 2 consecutive samples per subject: 5 minutes
- Model time unit: 2 hours
- Time windows: 9:00 to 12:00 and 13:00 to 16:00 over 2 consecutive days
- Objective: logdet of the expected FIM

The problem was solved using the distributed parallelism feature of Pumas powered by JuliaHub using different number of processes. All the runs were given a time limit of 15 minutes. The following table highlights the effectiveness of distributed parallelism to obtain better designs. The #branches is the number of branches explored in the branch and bound algorithm in the mixed integer nonlinear programming algorithm. More branches typically leads to improved designs. More processes The scaling of the number of branches with the number of processes used is nearly linear. The same random number generation seed was used for all the experiments, therefore the variability can only be attributed to the number of processes and how they are used in the algorithm which is completely deterministic. The optimal schedule (rounded to the nearest minutes) of the 8 processes run was:

1. **Day 1**: 13:00
2. **Day 2**: 9:00, 9:05, 9:10, 9:15, 9:20, 9:25, 9:30, 9:35, 15:35, 15:40, 15:45, 15:50, 15:55, 16:00

Notice how the minimum 5 minute gap as well as the time windows are respected in the optimal design.

Table 2: the optimal objective values and number of branches explored by Pumas after 15 minutes of optimization.

	# processes		
	1	4	8
Optimal logdet of the expected FIM	14.85	15.19	16.89
#branches explored	55	173	376

Conclusions and Discussions

- Two novel mathematical programming (MP) formulations of the sample time optimization problem were proposed.
- State-of-the-art MP algorithms were used to find optimal experiment designs using automatic differentiation.
- The proposed approach offers better support for logistic constraints than other tools.
- Even on a simple problem with a single time window and no logistic constraints, Pumas is significantly faster than the well known software package PopED for the test problem used.
- Distributed parallelism in Pumas can be used to accelerate the finding of good designs that respect realistic logistic constraints.

References

1. Bradley, Hax, and Magnanti (1997). Applied Mathematical Programming. Addison-Wesley.
2. Chris Rackauckas, et. al. (2020). *doi*: https://doi.org/10.1101/2020.11.28.402297
3. Wächter, A et. al. (2002). IPOPT: An interior point algorithm for large-scale nonlinear optimization.
4. Ole Kröger et. al. (2018). Juniper: An Open-Source Nonlinear Branch-and-Bound Solver in Julia. *Integration of Constraint Programming, Artificial Intelligence, and Operations Research (p377-386). Springer International Publishing*
5. Nyberg J et. al. (2012). PopED: An extended, parallelized, nonlinear mixed effects models optimal design tool. *Computer Methods and Programs in Biomedicine*, 108.
6. Silber HE, Jauslin PM, Frey N, Gieschke R, Simonsson US, Karlsson MO. An integrated model for glucose and insulin regulation in healthy volunteers and type 2 diabetic patients following intravenous glucose provocations. J Clin Pharmacol. 2007 Sep;47(9):1159-71. doi: 10.1177/0091270007304457. PMID: 17766701.
7. Silber HE, Nyberg J, Hooker AC, Karlsson MO. Optimization of the intravenous glucose tolerance test in T2DM patients using optimal experimental design. J Pharmacokinet Pharmacodyn. 2009 Jun;36(3):281-95. doi: 10.1007/s10928-009-9123-y. Epub 2009 Jun 25. PMID: 19554431.