Introduction to Metaheuristics with Applications to Clinical Research

Will Gertsch

Outline

- Intro to metaheuristics
  - Why use metaheuristics?
  - Examples of metaheuristics
  - Practical use
- Application 1: optimal design
  - Brief review of optimal design for dose-response
  - Demo of R Shiny application
- Application 2: subgroup analysis
  - My class project
  - Introduction to mixture models
  - Logistic normal mixture model for heterogeneous treatment effects
  - Fitting the model using metaheuristics

Intro to Metaheuristics

For some optimization problems, traditional algorithms such as simplex, Newton-Raphson, quasi-Newton, EM, MM, etc., may not be able to arrive at an optimal solution. Possible issues:

- Non-convexity => local optima
- Complicated gradient/Hessian structures
- Non-continuous objective function
- Mixed variable types => combinatorial optimization

Metaheuristics

Metaheuristic algorithms have several advantages over traditional algorithms:

- Stochasticity helps the algorithm avoid local optima
- Very few assumptions made about the problem => can be easily applied to solve almost any problem
- Algorithms are very simple conceptually.

Trade-offs include

- No guarantee of convergence
- Can be slower to converge

In real world situations, these downsides can be less important.

Examples of metaheuristics

Two major types of metaheuristic algorithms:

- Evolutionary algorithms
- Swarm-based (nature-inspired) algorithms

Examples:

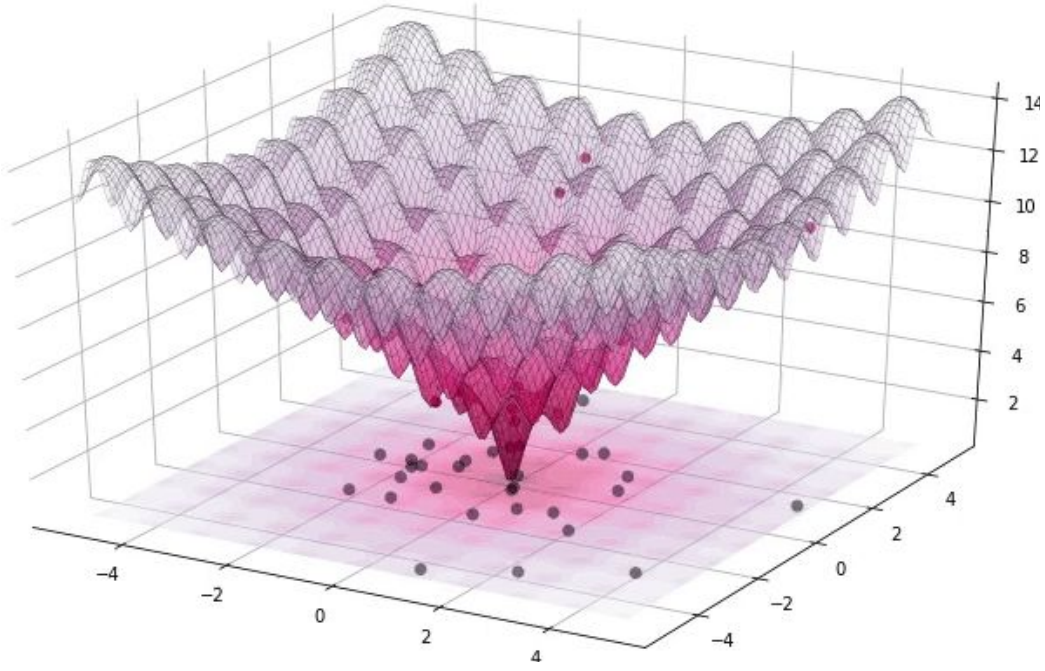
- Particle Swarm Optimization (PSO)
- Grey Wolf Optimizer (GWO)
- Genetic Algorithms (GA)
- Differential Evolution (DE)

Example algorithm - Differential Evolution

Differential evolution (Storn, Price 1997) is a evolutionary algorithm for global optimization. The algorithm works by randomly generating a population of solution vectors. The algorithm then iterates for each solution X in the population:

- Pick 3 solutions a, b, c from the population
- Compute Y = a + F * (b - c) where F is a tuning parameter.
- With probability CR, update X_i <- Y_i
- If the new value of X is better, replace X with the new value.

DE visualization



Application areas for metaheuristics

- Design of global clinical trials
- Estimation and design for PK-PD studies
- Engineering: helping to design complicated components
- Computer science: training neural networks, classification
- Parameter estimation of statistical models

In this presentation:

- Optimal design of dose-response experiments
- Subgroup analysis using mixture models

Tips for using metaheuristics - my personal experience

- No single algorithm is the best for every problem (No Free Lunch theorem).
- Be smart about choosing variable bounds.
- PSO is one of the best options when the search space is large.
- Constraint and invalid solution handling can make or break certain problems.
- Metaheuristics are best used when:
  - The problem is intractable by the usual methods.
  - The problem has local optima
  - Checking the results of another algorithm
  - Advanced problems (multi-objective, mixed variable types, multilevel, etc.)

Software packages

R:

- metaheuristicOpt
- EmiR

MATLAB:

- PLATEMO

Julia:

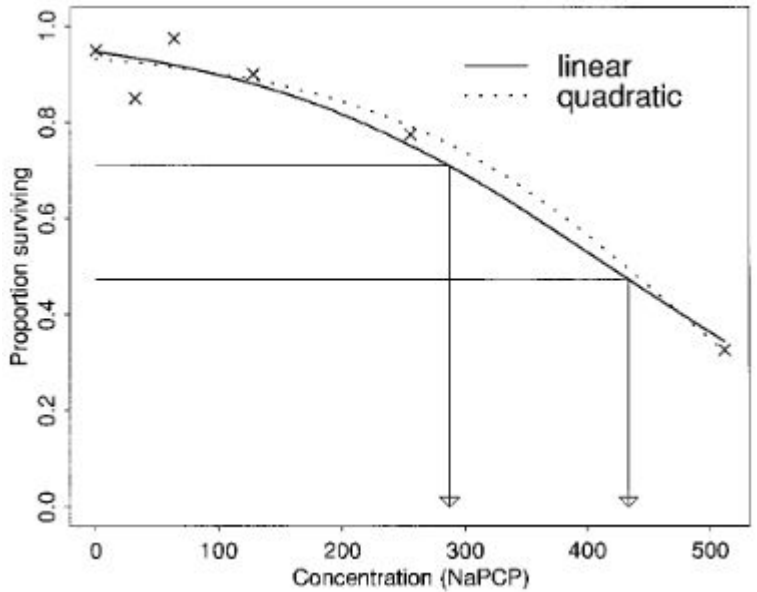
- Metaheuristics.jl

Application 1: optimal design

Background: review of binary dose-response studies

A dose-response study seeks to characterize the relationship between the dosage of a drug/substance some event of interest such as toxicity.

- Optimal design of these experiments can reduce costs and improve the quality of the analysis.
- We need optimally choose doses and how many subjects are assigned to each dose.



Optimal design for a fractional polynomial logistic model

If the event of interest is $y_i \sim \text{Bernoulli}(p_i)$, a common model is

$$\log\left(\frac{p_i}{1-p_i}\right) = \eta_i$$

where η is some linear predictor in x. A flexible form of η is the family of fractional polynomial functions (Royston & Altman 1994), which include standard polynomials and Box-Cox transformations.

We are interested in:

- D-optimal designs
- Dual objective designs for D-optimality and estimating a percentile.

R Shiny app and code

The information matrix for FP logistic model is complex to derive and optimize using standard methods, so we use metaheuristics to find the optimal designs.

My R Shiny app:

- Web app, code
- Algorithms from metaheuristicOpt package
- Uses equivalence theorem to confirm optimality.
- More details in background section of the webapp.

Application 2: subgroup analysis

Introduction to subgroup analysis

- In most clinical trials, the effect of the treatment is considered fixed at the same value for all trial participants.
- EX: For a two-arm RCT, we compute the difference in the mean outcomes between treatment groups as an estimate of the true (singular) treatment effect between groups.
- However, the treatment may be better or worse for certain subgroups of the trial population.
- A subgroup analysis investigates treatment effect heterogeneity explained by partitions of the patient population.

Classical subgroup analysis

If some known sub-population (ex: sex, race, etc.) is expected to have a different response to the treatment compared to the global population, a separate analysis of the treatment is often performed. Possible problems with this approach include:

- Need to adjust for multiplicity
- High potential for “data-dredging”
- Study needs to be adequately powered
- Doesn’t help discover new groups with better treatment response

Data-driven subgroup analysis

An alternative approach is search for a latent subgroup that has certain properties. Some uses include: (Lipkovich 2016)

- Salvaging a failed phase III study by identifying a subgroup with substantial treatment benefit
- Identifying “super-responders” in a successful trial
- Identifying optimal treatment regimes (OTR) for precision medicine
- Regulatory label restrictions due to treatment effect inconsistency or safety issues in a specific group

Methods for identifying latent subgroups

- Treatment selection based on overall population benefit (Song and Pepe 2004)
- Tree/forest based methods (Chen et. al. 2007)
- “Virtual twins” (Foster et al 2011)
- Parametric scoring (Cai et al 2011, Zhao et al 2013)
- Thresholding (Wang, Li, Li, Wong 2019)
- “Depth Importance” (Chen, Zhang 2022) (Biostat. Dept. seminar Spring 2022)
- **Mixture models (Shen, He 2015)**

Mixture model approach is “less aggressive” than other methods and can assess the statistical significance of the identified subgroup.

Mixture models

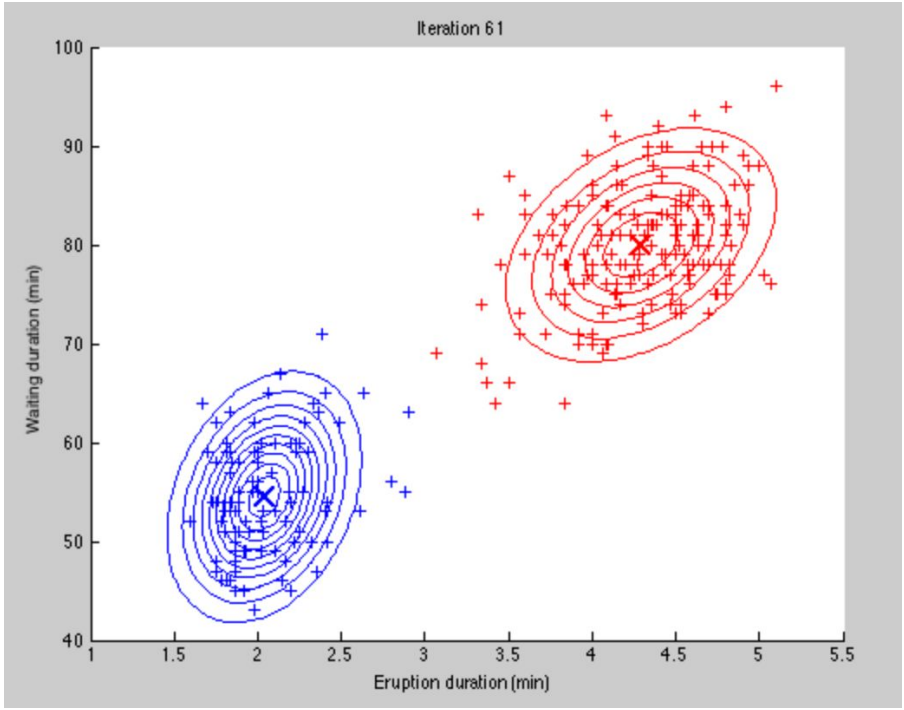
A mixture model assumes that the data are sampled from a population with g sub-populations. The probability density function is

$$f(y|w,\theta) = \sum_{j=1}^{g} w_j f_j(y|\theta_j)$$

where w_j is the probability of sampling from sub-population j with density function f_j and is usually unknown. The density functions may all come from the same family or come from different families of distributions.

Example: Old Faithful data

- Old Faithful is a geyser in Yellowstone National Park that erupts every 44min to 2 hours.
- Eruption duration and time in between eruptions can be modeled using a mixture of 2 bivariate Gaussians.



Estimation of mixture models

Estimation of mixture models is most commonly done by maximum likelihood. The log-likelihood for a general mixture model is

$$\ell(w, \theta|y) = \sum_{i=1}^{n} \log \sum_{j=1}^{g} w_j f_j(y_i|\theta_j)$$

Note that the log of the summation makes it more difficult to derive the MLE analytically.

Estimation - EM algorithm

The expectation-maximization (EM) algorithm is the most common way of maximizing the likelihood for a mixture model.

- Latent group membership is assumed to be missing data.
- Expectation step estimates the posterior probability of an observation belonging to each group given the current parameter estimates
- Maximization step uses the posterior probabilities to find new values of the other parameters that maximize the complete data likelihood.
- Each iteration is guaranteed to increase the value of the likelihood function.

Estimation - metaheuristics

EM estimation drawbacks:

- Slower convergence than Newton-type algorithms.
- May converge to local optima => huge problem for classification
- Starting values matter.
- Have to manually derive the algorithm.

Alternatively, we could use metaheuristic algorithms

- Optimize log-likelihood directly
- Can avoid convergence to local optima
- Algorithms can outperform EM in certain cases.

Application: A hypothetical clinical trial

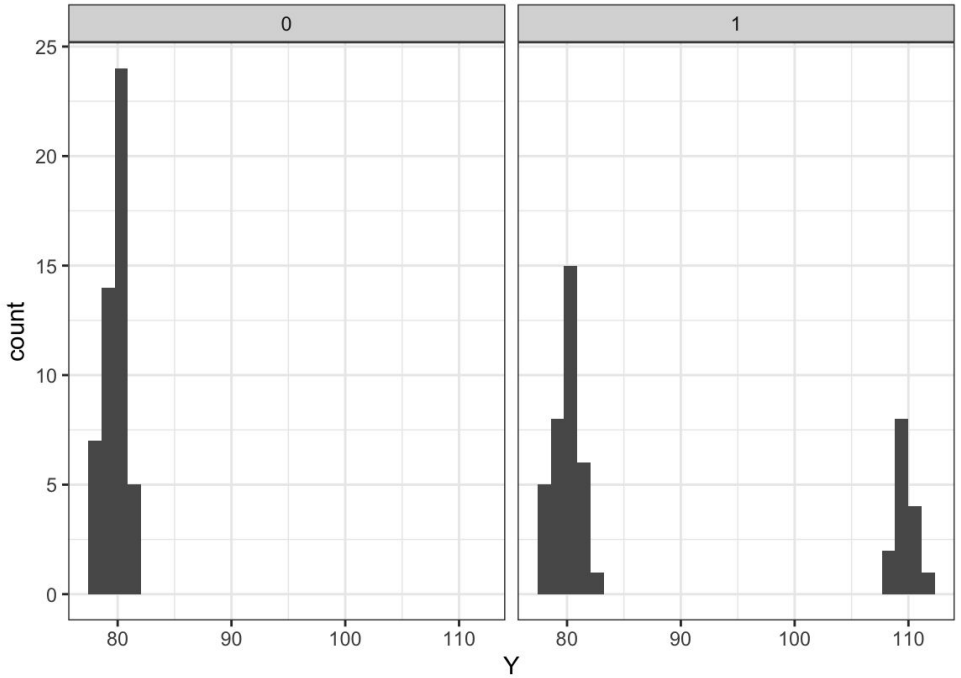
A pharmaceutical company runs a phase III clinical trial to evaluate if their new drug is effective. An effective drug will increase biomarker Y.

	Placebo (n = 50)	Treatment (n = 50)	Overall (n = 100)
Y: mean (sd)	79.8 (0.916)	88.9 (13.9)	84.3 (10.8)
Sex (male): n (%)	28 (56)	30 (60)	58 (58)

- Welch's 95% CI (5.2, 13.1) indicates a significant treatment effect.
- A Chi-squared test indicates sex is balanced across arms (p = 0.86).
- Seems like the drug is effective?

Application: a hypothetical clinical trial

- After review, it is clear that the drug only works in a subgroup of patients.
- Tests of all covariate interactions fail to identify the subgroup.
- Culprit is an interaction with unobserved variable, geneX.
- We need a joint model for the latent factor and observed data.



Modeling the heterogeneity of the treatment effect

Let Y_i be the observed outcome and let δ_i be an unobserved indicator of whether the patient has geneX. We define a model

$$Y_i = Z_i'(\beta_1 + \delta_i \beta_2) + \epsilon_i$$
$$P(\delta_i = 1|\gamma) = \text{ilogit}(X_i'\gamma)$$
$$\epsilon_i \sim N(0, \sigma^2)$$

- β_1 : intercept, treatment effects for the worse performing subgroup.
- β_2 : differences for subgroup that experiences higher treatment effect. => interaction effects
- γ : parameters for prediction of latent class membership from baseline characteristics.
- ilogit: inverse logistic function $\text{ilogit}(x) = \frac{\exp(x)}{1 + \exp(x)}$

Likelihood and estimation

This model is mixture of normal distributions where the weights are determined by a logistic regression of the unobserved factor on baseline covariates. The likelihood is

$$L(\theta) = \prod_{i=1}^{n} \left[\text{ilogit}(X_i'\gamma)\phi(y_i|Z_i'(\beta_1+\beta_2),\sigma^2) + (1-\text{ilogit}(X_i'\gamma))\phi(y_i|Z_i'\beta_1,\sigma^2) \right]$$

where ϕ is the normal PDF.

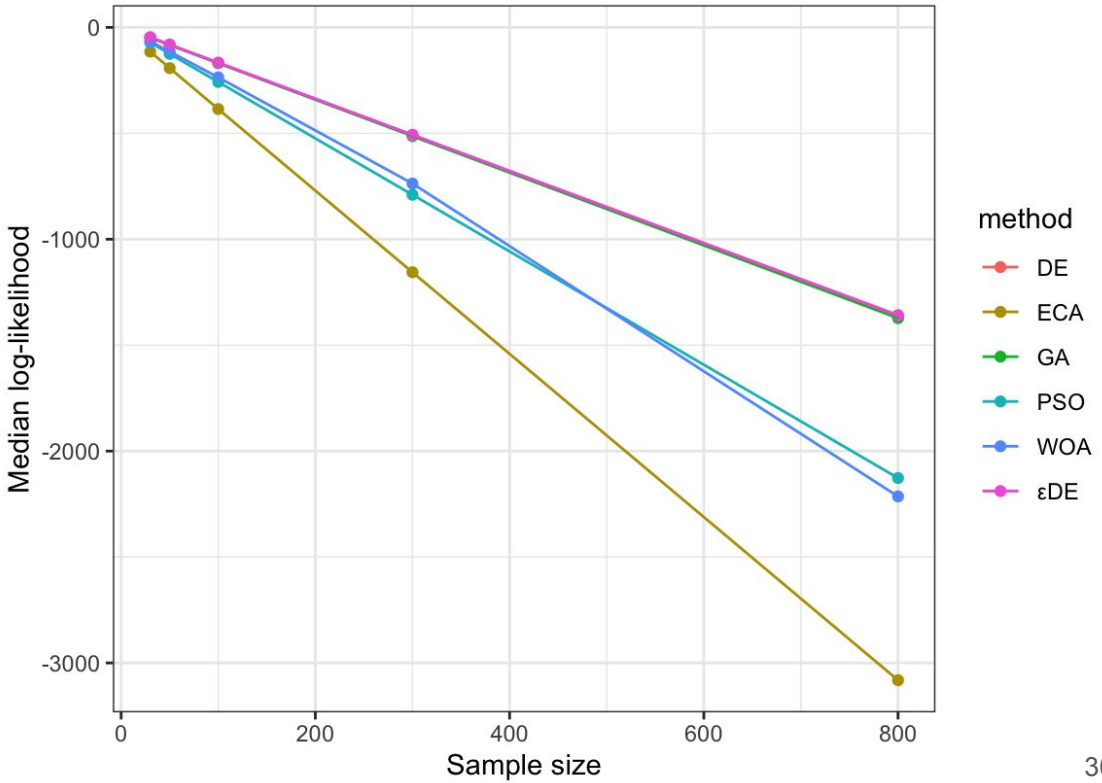
- Implementation of an EM algorithm requires an additional IRWLS step.
- Alternatively, we can use metaheuristics to directly optimize the log-likelihood.
- I fit the model using 7 different metaheuristics.

Results - comparison of algorithms at default settings

Algorithm	Loglik	β_{11}	β_{12}	β_{21}	β_{2_2}	γ_1	γ_2	σ
DE	-164.65	79.7	0.22	0.22	29.71	-1.39	1.2	0.94
εDE	-164.65	79.7	0.22	0.22	29.71	-1.39	1.2	0.94
GA	-166.37	79.55	0.3	0.9	29.29	-1.67	0.99	0.96
WOA	-183.94	79.51	0.83	-4.29	33.75	-1.82	-0.15	1.04
PSO	-370.57	77.55	-24.63	2.22	33.75	10	5	10.81
ECA	-411.11	88.76	6.64	12.85	20.62	-3.46	2.69	10.68
SA	-1627.56	86.09	27.78	1.53	19.1	5.97	2.81	4.61
True values		80.0	0.0	0.0	30.0	-1.39	1.79	1.0

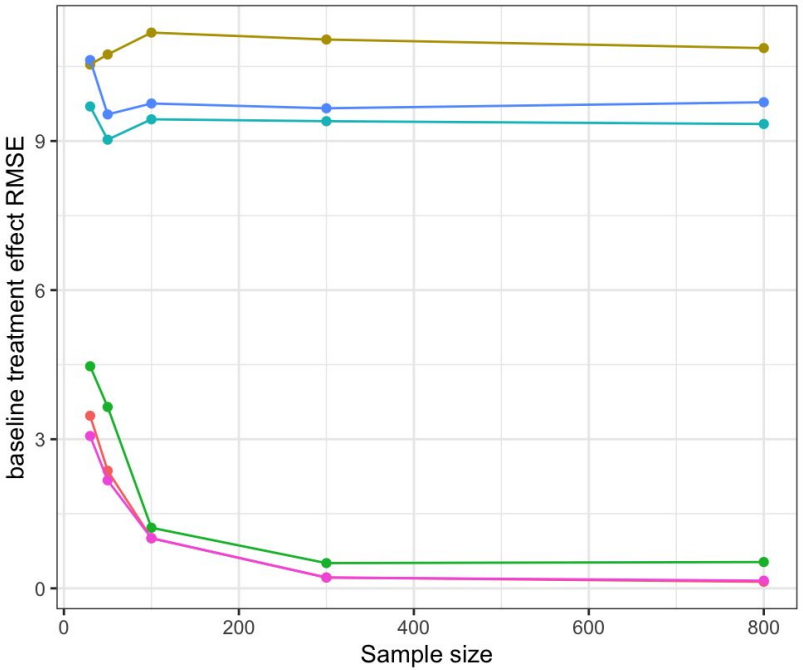
Simulation study - log-likelihood

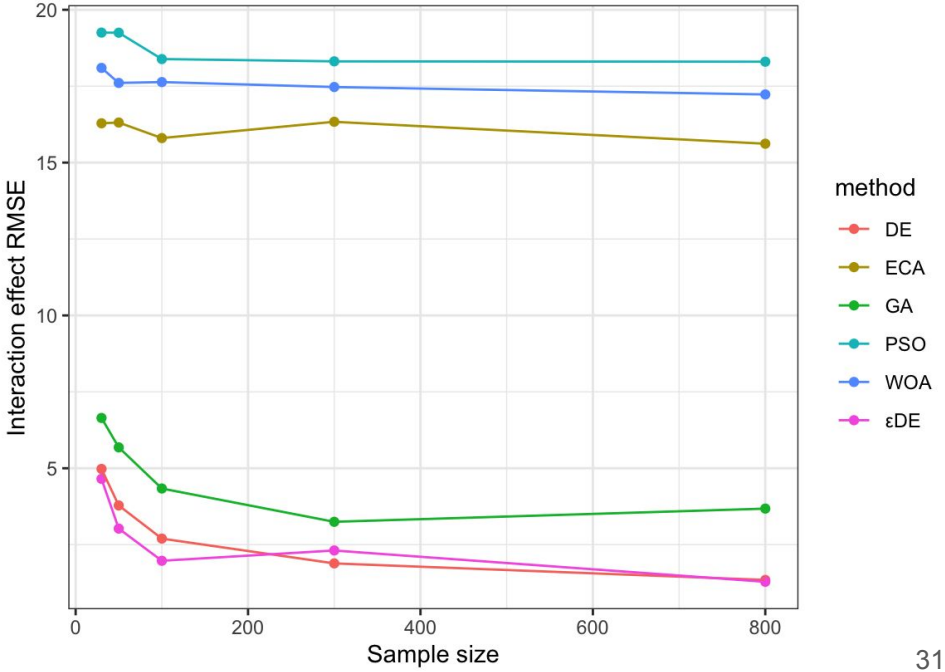
- A simulation study at different sample sizes shows that the evolutionary algorithms perform the best.
- SA is omitted for vastly inferior performance.
- PSO and WOA might have better performance with tuning.



Simulation study - important parameters

DE-based algorithms performed best for parameter estimation.





Future work (for my report)

- Extension to longitudinal data
- Bootstrap interval estimation and inference
- Comparison to EM algorithm
- Using data from my GSR

Conclusion and takeaways

- Metaheuristics are useful tool for optimization and can work better than standard algorithms in many situations.
- Metaheuristics can be applied to solve problems in clinical research such as optimal design and subgroup analysis.
- Can run many different algorithms to find one that is “good enough” for the problem at hand.
- ½ of statistics is optimization.

Code

My Github repos relevant to this presentation

- https://github.com/willgertsch/MetaEst => parameter estimation using metaheuristics
- https://github.com/willgertsch/ODpoly/releases/tag/v1.0 => optimal design app

Some references

- Shen, J., & He, X. (2015). Inference for subgroup analysis with a structured logistic-normal mixture model. Journal of the American Statistical Association
- Lipkovich, I., Dmitrienko, A., & B D'Agostino Sr, R. (2017). Tutorial in biostatistics: data‐driven subgroup identification and analysis in clinical trials. Statistics in medicine