**Search for the optimal two-phase experimental designs using simulated annealing algorithm**

**Structure**

1. Introduction
   1. With a given set of design parameters, the treatment and animal allocations from the phase 1 experiments to the second phase experiments can be many.
   2. The aim is to find an allocation that is said to be *optimal*, then describe the different optimal criteria from John and Williams.
   3. Briefly introduce the simulated annealing algorithm to the reader, and it will be further discussed in detail in Section 2).
   4. Theoretical ANOVA table is used to study the properties of designs
      1. Can a valid test for the treatment differences be conducted?
      2. Can the variances of treatment effects be estimated?
      3. Degrees of freedom remain in the between animals stratum for conducting the test.
      4. Amount of treatment information remains, based the average efficiency factor computed from the harmonic means of the canonical efficiency factors.
   5. Aim is to develop a method for generating the optimal two-phase experimental designs focusing on the multiplex proteomics experiments. This chapter describes in generating the two-phase optimal designs for the following design parameters:
      1. Phase 1 experiment – completely randomised designs with 2-8 treatments, 2-10 biological replicates and 2 technical replicates,
      2. Phase 2 experiment – randomised block designs with 4/8 tags.
      3. Introduce an example to be used throughout this chapter.
2. Objective function
   1. Since the aim of the animals and treatments allocations to runs and tags is to minimise the level of confounding between animal and treatment effects with both runs and tag effects. Instead, we can find the design that has the most animal and treatment information within the runs and tags stratum. Hence, the second phase design is in fact the row-column design.
   2. The objective function is computing the weighted sum of the average efficiency factors of animals and treatments within runs and tags stratum. Explain why using the A-optimal criterion instead of MS-optimal, i.e. do not bother about the whether the animal component of the design is disconnected or completely confounded (i.e. losing degrees of freedom) with runs. Example?
   3. In addition, to avoid the treatment to be disconnected to the runs and tags, addition criterion on monitoring the degrees of freedom of the treatment effects is included in the objective function. Example?
3. Describe the Simulated annealing in detail and how it will be used for this experiments
   1. Generating the starting design - discusses the 2-by-2 grouping for designing experiment of 2 technical replicates, try to illustrate it with the tables or matrices.
   2. Temperature control – applied accelerated cooling method described by John and Whitaker (1993).
   3. Swapping method
      1. Discuss the swapping of treatment/animal pairs for the experiments with 2 technical replicates
      2. Two-stage swapping described by Williams and John (1996) was also applied by swapping the observations restricting to runs and tags alternatively. This will allow the reduction of the search space while applying the simulated annealing.
   4. How the simulated annealing algorithm described here is implemented by the optim function in R.
4. Illustrate some examples and table of results
   1. Show few examples with different set of design parameters with their theoretical ANOVA tables,
      1. 6 treatment 3 biological replicates and 4-plex experiments to show how the simulated annealing algorithm described here improves the design.
      2. Another set of design parameters where simulated annealing cannot improve the starting designs.
      3. Another design parameter with 8-plex experiments
   2. Results – table
5. Discussion

Print out the designs layout in the appendix.

**Introduction**

Studying the properties of different two-phase experiments can take a long time.

This is because for a given experiment with a given number of treatments, blocks and replicates, there can be many different ways of assigning treatments and blocks from the first phase experiment to the blocks in the second phase.

We refer to the numbers of treatments, blocks and replicates as the *design parameters*.

With a given set of the design parameters, we should be able to generate at least one design that is optimal with respect to some statistical criteria.

This design is known as the optimal design. There are a number of different optimality criteria available for selecting the “best” two-phase design for a given set of design parameters. These criteria are ….

We will focus on identifying designs that satisfy the A-optimal criterion.

The searching method is based on simulated annealing algorithm ….

Finding the optimal designs for two-phase experiments

The aim is to develop a method for generating the optimal two-phase experimental design focusing on the multiplex proteomics experiments. In particular, we will consider experiments using the iTRAQ isobaric labelling technology which enables the simultaneous assaying of up to eight proteomic samples.

**Studying the patterns of the designs with 2 treatment groups with 4/8-plex experiments**

Abbreviations:

This write-up describes in generating the two-phase optimal designs for the following design parameters:

Phase 1 experiment – completely randomised designs with 2-8 treatments, 2-10 biological replicates, 2 technical replicates,

Phase 2 experiment – randomised block designs with 4/8 tags.

Applying the simulated annealing algorithm, some optimal designs have found. This write-up describes the patterns of these designs. From studying these patterns, it can be a good starting point to construct a generic method for designing the two-phase MudPIT-iTRAQ experiments.

The most trivial designs for the MudPIT-iTRAQ two-phase experiments are the designs with 2 treatment groups and two technical replicates with four-plex system. The patterns can then be divided into two main groups based on number of biological replicates from the first phase experiments: a) designs with even number of biological replicates and b) designs odd number of biological replicates.

For the designs with 2 treatment groups and two technical replicates with four-plex system, the number of biological replicates is identical to the number of MudPIT run for the second phase experiments.

For the rest of this write-up, the allocations are illustrated as a matrix where the rows correspond to the runs and columns correspond to the tags. In addition, the upper case letter denotes the animal ID and the lower case letter denotes the treatments.

The current objective function for the simulated annealing method is to find the design with the highest average efficiency factor when allocating the animals of Phase 1 experiment to the runs and tags of the Phase 2 experiments.

The orthogonal projector for the within runs and tags stratum can be written as

The information matrix can be written as

where denotes the animal design matrix.

Since the design criterion is to maximise the average efficiency factor, the aim is to find that can compute the highest the average efficiency factor from the information matrix.

I have found out that the design with the higher average efficiency of treatment in the between animals within runs stratum was hard to find using the current simulated annealing algorithm.

Therefore, I believe I need to modify the swapping method of the simulated annealing algorithm to enable me to find the optimal design quickly and efficiently. The current swapping method is to swap any random pair of observations throughout the design. Note that the way how I have designed the experiments with 2 technical replicates is to group a pair of animals and treatments and assigned them to a quadrant comprising 2 runs and 2 tags. Hence, the new swapping method is to swap any two random pairs of animals and treatments.

In addition, I have also modified the objective function of the simulated annealing algorithm. The current objection function gives a weighted sum of the average efficiency factors of animals and treatments. However, for an experiment consisting of more than 2 treatments, there will be more than 1 degree of freedom (DF) associated with the treatment effects in some stratum of the ANOVA table. Furthermore, for iTRAQ experiments, if I only try to find the design with the highest average efficiency factor for treatments in the between animals within runs stratum, then I can obtain a design that has 100% of average efficiency factor for the treatments, but some DF of treatments are lost to the other strata. This situation is also known as *disconnected design*. Therefore, to avoid generating such the disconnected designs when searching for the optimal design using the simulated annealing algorithm, the modified objection function gives a weighted sum of the average efficiency factors of animals and treatments and the proportion of the DF of treatment of current design to the total DF of treatment, i.e.

Using this new simulated annealing algorithm, I have be able find the designs with higher average efficiency factors of treatment as well as for the experiments with even biological replicates. The results can be shown in a table as follows,

3. Results table and then summarise