**Introduction layout**

**Set the scene**

This chapter discusses searching for optimal designs for two-phase proteomics experiments, where the Phase 1 experiment is a completely randomised design (CRD) with multiplexing technique such as iTRAQ in the Phase 2 experiment. The first phase (Phase 1) experiment involves the organisms that to be perturbed by the experimental conditions of interest. Since the protein abundances cannot be measured directly from the organisms, the second phase (Phase 2) experiment involves the iTRAQ experiment for measuring these protein abundances of the experimental materials from the Phase 1 experiment. This chapter considers completely randomised designs (CRD) with v treatments, r\_b biological replicates and r\_t technical replicates for first phase experiment. The second phase experiment is the randomised block design (RBD) where the numbers of MudPIT run, n\_R, and iTRAQ tag, n\_\gamma, correspond to the numbers of blocks and block size, respectively. These numbers of treatments, biological replicates, technical replicates, runs and tags are collectively known as the design parameters.

There are many ways to allocate the samples from the first phase experiment to second phase experiment. Let n denote the total number of samples, the number of ways that these samples can be assigned to each MudPIT run can be calculated as

.

For example, if n=10, then there are 210 ways to assign these samples to a run in a 4-plex experiment. An optimal design can be obtained from a specific allocation of these samples to each runs, based on various optimality criteria. A number of optimality criteria have been defined (John and Williams, 1987). This chapter focuses only on the MS- and A-optimality criteria which are further discussed in Section 4.

The method of finding the optimal designs for different class of design such as block, row-column and alpha designs has been previously discussed (citation here…). These methods are aiming to find the best allocation of the treatment factors to the block factors based on some optimality criteria defined. As for finding the optimal two-phase designs, the allocation of the block factors from the Phase 1 experiment to block factors of the Phase 2 experiment also need to be monitored, which complicates the search for the two-phase optimal designs. A suitable method of finding the two-phase optimal designs has not been described to date; the main reason is the optimality criterion is still yet to be defined for two-phase experiments.

This chapter describes the methods of generating two-phase optimal designs by optimising the objective function using the simulated annealing algorithm (SA). The optimality criterion is defined in the objective function, which is a mathematical expression describing the relationship between the each variable. SA is a well-known heuristic method for finding the values of variables that result in a maximum or minimum of an objective function (Kirkpatrick et al. 1983). In the case of finding optimal designs, the variables of the objective function correspond to the candidate designs. SA plays the role in continuously generating the new candidate design and comparing the values of the objective function with this new candidate design to the best design so far in search. The optimality criterion for finding the two-phase optimal design is defined in the objective function which is to be discussed in detail in Section 6. In addition, this chapter presents an improved version of SA which shown to speed up the search for the optimal designs in Section 7.

The aim of this chapter is to develop a method for generating the optimal two-phase designs focusing on the multiplexing proteomics experiments, namely MudPIT-iTRAQ experiments. For the layout of this chapter, Section 2 described the linear models of the Phase 1 and 2 experiments. Section 3 and 4 explain the information matrix and optimal criteria, respectively. The construction of the objective function is explained in Section 6. Section 7 discusses four components of the SA, the new initial design, temperature control, modified swapping and two-stage swapping method, and shows how these four components can improve the search for the optimal designs. Section 8 describes some examples to illustrate the objective function and SA. An overall summary of the results is presented in Section 9. Many different experiment examples are introduced throughout this chapter to aid in the explanation of how the new objective function and SA are improved for searching the optimal two-phase designs. The theoretical ANOVA table and average efficiency factors of treatments are generated for describing the properties of a design founded.

**Design parameters**

The chapter focuses only on two-phase experiments in which there are v = 2,…,8, r\_b = 2,…,10, r\_t = 2, n\_\gamma = 4,8 and n\_R = n/n\_\gamma.

Give linear model of the first phase experiments,

need to add the dummy index to the model to indicate the technical replicates.

Give the linear model of the second phase experiments.

Give the model in a matrix notation, which includes the design matrices of treatments, animals, runs and tags.

Define the incidence matrix

N

Concurrence matrix

NN’

**Information matrix**

Information matrix can be defined using the concurrence matrix

I = rI – (1/k)NN’

**############################################################**

**Optimal design of experiments**

Fisher information is widely used in [optimal experimental design](http://en.wikipedia.org/wiki/Optimal_design). Because of the reciprocity of estimator-variance and Fisher information, ***minimizing* the *variance*** corresponds to ***maximizing* the *information***.

When the [linear](http://en.wikipedia.org/wiki/Linear_model) (or [linearized](http://en.wikipedia.org/wiki/Nonlinear_regression)) [statistical model](http://en.wikipedia.org/wiki/Statistical_model) has several [parameters](http://en.wikipedia.org/wiki/Parameter), the [mean](http://en.wikipedia.org/wiki/Expected_value) of the parameter-estimator is a [vector](http://en.wikipedia.org/wiki/Column_vector) and its variance is a [matrix](http://en.wikipedia.org/wiki/Matrix_(mathematics)). The [inverse matrix](http://en.wikipedia.org/wiki/Inverse_matrix) of the variance-matrix is called the "information matrix". Because the variance of the estimator of a parameter vector is a matrix, the problem of "minimizing the variance" is complicated. Using [statistical theory](http://en.wikipedia.org/wiki/Statistical_theory), statisticians compress the information-matrix using real-valued [summary statistics](http://en.wikipedia.org/wiki/Summary_statistics); being real-valued functions, these "information criteria" can be maximized.

Traditionally, statisticians have evaluated estimators and designs by considering some [summary statistic](http://en.wikipedia.org/wiki/Summary_statistics) of the covariance matrix (of a [mean](http://en.wikipedia.org/wiki/Expected_value)-[unbiased](http://en.wikipedia.org/wiki/Unbiased) [estimator](http://en.wikipedia.org/wiki/Estimator)), usually with positive real values (like the [determinant](http://en.wikipedia.org/wiki/Determinant) or [matrix trace](http://en.wikipedia.org/wiki/Matrix_trace)). Working with positive real-numbers brings several advantages: If the estimator of a single parameter has a positive variance, then the variance and the Fisher information are both positive real numbers; hence they are members of the convex cone of nonnegative real numbers (whose nonzero members have reciprocals in this same cone). For several parameters, the covariance-matrices and information-matrices are elements of the convex cone of nonnegative-definite symmetric matrices in a [partially](http://en.wikipedia.org/wiki/Partial_order) [ordered vector space](http://en.wikipedia.org/wiki/Ordered_vector_space), under the [Loewner](http://en.wikipedia.org/wiki/Charles_Loewner" \o "Charles Loewner)(Löwner) order. This cone is closed under matrix-matrix addition, under matrix-inversion, and under the multiplication of positive real-numbers and matrices. An exposition of matrix theory and the Loewner-order appears in Pukelsheim[[15]](http://en.wikipedia.org/wiki/Fisher_information" \l "cite_note-14).

The traditional optimality-criteria are the [information](http://en.wikipedia.org/wiki/Information)-matrix's [invariants](http://en.wikipedia.org/wiki/Invariant_theory); algebraically, the traditional optimality-criteria are functional of the [eigenvalues](http://en.wikipedia.org/wiki/Eigenvalue) of the (Fisher) information matrix: see [optimal design](http://en.wikipedia.org/wiki/Optimal_design).

##########################################################################

Instead of considering the design for the second phase experiment is a randomised block design, the second phase experiment can be considered as the two-column designs where the rows and columns correspond the runs and tags, respectively.

Define the orthogonal projector of the Between Runs and Tags stratum.

Give the animal information matrix with respect to Within Runs and Tags stratum.

Give the treatment information matrix with respect to Within Runs and Tags stratum.

**MS- and A- optimality criteria**

Define both optimality criteria.

MS-optimality criteria

A-optimality criteria

**Objective function**

What SA does is to establish a search path in search space, each step of the search path corresponds to the new candidate design generated from iteration of the search as illustrate in Figure .