Proceedings of Korean Data And Information Science Society,
October 28-29, 2005, pp. 95-102

Factorial Design for cDNA and RNA of Microarray

Choi Kuey Chung, Cheong Min Ju
(kjchoi@chosun.ac.kr , chisolsong@hanmail.net)

1. Introduction

Although balanced factorial designs, characterized by Shah (1960) and Kshirsagar (1966), were developed mainly in the context of agricultural experiments, balanced designs have been found to be useful in other settings as well. For instance, balanced designs have also been studied extensively for bioassay experiments, see Gupta and Mukerjee (1996), Kshirsagar and Wang (1996), and the references cited therein. The purpose of this paper is to provide balanced factorial designs for cDNA microarray experiments. Since two experimental conditions are ybridized on each microarray slide, the arrays form blocks of size two. The blocks are incomplete if more than two conditions are under study. Using the ANOVA model, Kerr and Churchill (2001a,b), and Churchill (2002) provided a detailed discussion of several design issues involved in microarray experiments. They showed the inefficiency of the “common reference" design, the design that has been widely used for conducting microarray experiments. The authors introduced loop designs (or cyclic designs) for microarrays and showed that they offered considerable improvement over the "common reference" designs in terms of efficiency. However, loop designs become inefficient for larger number of treatment combinations. Kerr and Churchill (2001a) used the A-optimality criterion to construct designs. Glonek and Solomon (2003) stressed the importance of keeping the contrasts of interest in mind for constructing efficient designs. They argued that a design with high overall efficiency may

not be the most efficient design as far as estimation of the specific contrasts of interest is concerned. The authors considered admissible designs, and restricted their computer search of designs to the class of admissible designs.

The reader is referred to, for example, Lee (2004) for background details on microarrays. The anatomy of confounding and estimation of factorial effects in single replicate block designs with block size two is first briefly discussed in the next section. Applications of classical confounded designs and generalized cyclic designs to factorial microarrays are presented in section 3. The 2^2, 2^3 and 3×2 experiments are discussed in detail. Microarray designs for other experiments can be obtained similarly. Finally, some concluding remarks are made in section 4.

2. Confounded and unconfounded designs with block size two

Consider a factorial experiment involving m factors $F_1, F_2, \cdots, F_m$ at two levels each. Treatment combinations will be denoted by m -tuples $a_1 a_2 \cdots a_m$ where a_i=0 or 1 denote the two coded levels of the ith factor, $i = 1, 2, \cdots m$. Let D be a single replicate design having 2^{m-1}blocks of size two for the m-factor experiment obtained using the classical method of confounding, see e.g. Raghavarao (1971). Let the 2^{m-1}factorial effects (main effects and/or interactions) confounded between the blocks of D be denoted by B_i,

$i = 1, 2, \cdots, 2^{m-1}$, and let A_i, $i = 1, 2, \cdots, 2^{m-1}$ denote the factorial effects that are unconfounded, and hence estimable in D.

Let (t_{i1}, t_{i2}) denote the two treatment combinations in the ith block of D, and let (y_{i1}, y_{i2}) be the observations corresponding to them, $i = 1, 2, \cdots, 2^{m-1}$. Then, the contrasts of block totals $y_{i1} + y_{i2}$, $i = 1, 2, \cdots, 2^{m-1}$, estimate the block effects, the factorial effects B_i, $i = 1, 2, \cdots, 2^{m-1}$, are confounded with. Whereas, the unconfounded factorial effects A_i, $i = 1, 2, \cdots, 2^{m-1}$, are estimated using the within block comparisons $y_{i1} - y_{i2}$, $i = 1, 2, \cdots, 2^{m-1}$. Let

$$y_d = [(y_{11} - y_{12}), (y_{21} - y_{22}), \cdots, (y_{b1} - y_{b2})]'$$

where $b = 2^{m-1}$ denotes the number of blocks. Let $h_j = (h_{j1}, h_{j2}, \cdots, h_{jb})'$ be such that $E(h'_j y_d) = A_j$, $j = 1, 2, \cdots, 2^{m-1}$. Then, clearly h_j, $j = 1, 2, \cdots, 2^{m-1}$, form a complete set of mutually orthogonal column vectors of size 2^{m-1}. Also, $Var(h'_j y_d) = h'_j h_j \sigma^2$ where σ^2 denotes the constant variance of the difference $y_{i1} - y_{i2}$. In the case of a microarray experiment involving 2^{m-1} arrays, the within block comparison $y_{i1} - y_{i2}$ corresponds to the $\log_2$ expression ratio for the two treatment combinations hybridized on the th array.

A partially confounded design in which all the 2^{m-1} factorial effects are estimable may be obtained by adding further replicates that confound a different set of factorial effects. Although we have illustrated the method of confounding using 2^m experiments, the method is applicable in general to factors with number of levels a prime or power of a prime number. In fact, similar conclusions also

hold for balanced factorial designs with block size two obtained using methods other than the classical method of confounding.

3. Factorial Microarrays

3.1 2^2 experiments

In a 2^2 experiment we have two factors F_1 and F_2 at two levels each. As before, the two levels of each factor are coded as 0 and 1. There are two more factors in addition to F_1 and F_2. These are the two nuisance factors: dye at two levels and array or slide at b levels, where b denotes the total number of slides used in the experiment. Throughout, we will denote dye by factor F_d with its two levels coded as 0 (cy5:red) and 1 (cy3:green). The model for the data can be written as

$$(y_{i1} - y_{i2}) = (\delta_1 - \delta_2) + (\gamma_{t_{i1}} - \gamma_{t_{i2}}) + \epsilon_{i1i2}$$

where the within block comparison from the i th block, $y_{i1} - y_{i2}$, corresponds to the $\log_2$ expression ratio for the two treatment combinations t_{i1} and t_{i2} hybridized on the th array. The $\delta_1 - \delta_2$ is the dye effect, γ_j is the effect of the jth treatment combination, and ϵ_{i1i2} are random errors independently distributed with mean zero and constant variance σ^2. The $\log_2$ expression ratios are assumed to have been subjected to normalization and background correction methods. Note that the block effects get eliminated from the within block comparisons $y_{i1} - y_{i2}$. The dye effect $\delta_1 - \delta_2$ will also get

automatically eliminated from estimates of the factor main effect and the interaction if the treatments are orthogonal to dyes, i.e. if the two levels of dye appear equally often with each treatment combination. Note also that the above model is applicable for each gene separately; thus the treatment effects, and dye effect are gene specific.

The single replicate designs obtained using the method of confounding along with the effect confounded in each replicate are given below in Table 1.

Table 1

Designs for 2^2 factorial

Design	Blocks	Effect confounded
D_{11}	[00, 01], [10, 11]	F_1
D_{12}	[00, 10], [01, 11]	F_2
D_{13}	[00, 11], [01, 10]	F_1F_2

A microarray design is a row-column design with F_g (dye) as the row factor and blocks (slides) as the column factor; the levels 0 (red) and 1(green) of F_g correspond to the first and second rows respectively. Using D_{11} and D_{12} we get the following microarray design, where columns correspond to the four slides:

00	11	10	01
01	10	00	11

All the treatment combinations in the first row are labeled with cy5 and those in the second row are labeled with cy3. For instance, the first column [00, 01] represents a slide hybridized with treatment

combinations 00 and 01, which are labeled with red and green dyes respectively. The treatment combinations in the columns are arranged such that rows are orthogonal to treatment combinations. The orthogonality of rows and treatments can be achieved only when the number of slides used is a multiple of the number of treatment combinations. The above microarray design estimates the main effects F_1 and F_2 with 50% efficiency, whereas, since the interaction F_1F_2 is unconfounded both in D_{11} and D_{12}, it is estimated with full efficiency. This design is in fact the common loop (CL) design given by Kerr and Churchill (2001a). Taking two replications of D_{13} yields the following design, referred to as cross-swap (CS) design by Landgrebe et al. (2005):

00	01	11	10
11	10	00	01

Landgrebe et al. (2005) considered two designs in 16 slides; one design using four replications of CL and the other using two replications each of CL and CS. In fact, D_{11}, D_{12}, D_{13} can be appropriately combined to obtain additional designs with desired efficiencies of estimation of $F1$, F_2 and F_1F_2. For example, taking 3 replications of D_{11}, 2 replications of D_{12} and 3 replication of D_{13} yields the following microarray design in 16 slides:

00	10	01	11	00	10	10	01	00	11	01	11	00	01	11	10
01	11	00	10	01	11	00	11	10	01	10	00	11	10	00	01

The design estimates F_1, F_2 and F_1F_2 with efficiencies 62.5%, 75%

and 62.5% respectively. The choice of a design will depend upon the desired efficiency or precision of estimation of the contrasts of interest.

References

Glonek, G., Solomon, P. (2004). Factorial and time course designs for cdna microarray experiments. Biostatistics 5:89-111.

Gupta, S. (1987). Generating generalized cyclic designs with factorial balance. Commun. Statist. - Theor. Meth. 16:1885-1900.

Gupta, S., Mukerjee, R. (1996). Developments in incomplete block designs for parallel line assays. In: Ghosh, S., Rao, C.R. ed., Handbook of Statistics 13:875-902. Amsterdam: North-Holland.

Kerr, M., Churchill, G. (2001a). Experimental designs for gene expression microarrays. Bio statistics 2:183-201.

Kerr, M., Churchill, G. (2001b). Statistical design and analysis of gene expression microarray data. Genet. Res. 77:123-128.

Kshirsagar, A.M. (1966). Balanced factorial designs. J. R. Statist. Soc. B 28:59-67.

Kshirsagar, A.M., Wang, W. (1996). Use of Lattice square designs in

bioassays. J. Biophar. Statist. 6:185-199.

Landgrebe, J., Bretz, F., Brunner, E. (2005). Efficient design and analysis of two colour factorial microarray experiments. Comp. Statist. and Data Analysis (In press)

Lee, M-L. T. (2004). Analysis of Microarray Gene Expression Data. New York: Springer.

Lewis, S.M. and Tuck, M.G. (1985). Paired comparison designs for factorial experiments. Appl. Statist. 34:227-234.

Raghavarao, D. (1971). Constructions and Combinatorial Problems in Design of Experiments. New York: Wiley.

Shah, B.V. (1960). Balanced factorial experiments. Ann. Math. Statist. 31:791-794.

Wit, E., Nobile, E., Khanin, R. (2004). Simulated Annealing for Near Optimal Dual-Channel Microarray Designs. Technical Report No. 04-7, Department of Statistics, University of Glasglow, U.K.

Yang, Y., Dudoit, S., Luu, P., Lin, D., Peng, V., Ngai, J., Speed, T. (2002). Normalization for cdna microarray data: a robust composite method addressing single and multiple slide systematic variation. Nucleic Acids Res. 30:e15.

Yang, Y., Speed, T. (2002). Design issues for cdna microarray experiments. Nature Rev.: Genetics 3:579-588.