Mathematica Aeterna, Vol. 5, 2015, no. 3, 441 - 447

Implementation of chemical balance weighing design

Małgorzata Graczyk

Department of Mathematical and Statistical Methods
Poznań University of Life Sciences
Wojska Polskiego 28, 60-637 Poznań
magra@up.poznan.pl

Abstract

In this paper, an implementation of the design matrix of the chemical balance weighing design is presented. We give the relation between the design matrix of two-colour microarrays experiment and the design matrix of the chemical balance weighing design. Moreover, the equality of the estimators of all contrasts in the model of two-colour microarrays experiment and in introduced model of the chemical balance weighing design is dished up. An application of presented theory is indicated.

Mathematics Subject Classification: 62K05

Keywords: chemical balance weighing design, two-colour microarray

1 Introduction

In order to present the relation between the chemical balance weighing design and two-colour microarrays experiment, first off, let us consider the microarrays experiment. At the beginning let us recall a few properties of such models. Some technique of design and optimality issues in microarrays experiments has been described in [8] [9], [6], [14] and [2]. A general overview of statistical designs in microarray experiments is presented in [1]. Many papers have been published in selection optimal microarrays. To select efficient designs for microarray experiments, in [10] a minimax approach is considered. The importance of robustness in the context of microarray experiments has been studied in several papers including [11] one. Here, two criteria of the robustness of microarray designs against missing observations are considered. In [15], simulated annealing are used to find near-optimal (A- and D-optimal) microarray designs in one-factor experiments.

Let us consider s treatments and a arrays. The statistical analysis is based on the gene-specific model

$$\log_2(y_{ijl}) = t_i + \alpha_j + \delta_l + \epsilon_{ijl}, \tag{1}$$

where y_{ijl} describes the intensity of treatment i coloured in dye l on array j, $l =$ green or red, t_i denotes the effect of i^{th} treatment, α_j the effect of j^{th} array, δ_l denotes the effect of l^{th} colour and ϵ_{ijl} are the error terms for $i = 1, 2, ..., s$, $j = 1, 2, ..., a$. Suppose that array j has treatments i and k coloured green and red respectively.

For analysis using intra-array information only, model (1) can be replaced by

$$\log_2\left(\frac{y_{ij\ green}}{y_{kj\ red}}\right) = t_i - t_k + \delta_{green} - \delta_{red} + \epsilon_{ij\ green} + \epsilon_{jk\ red} \tag{2}$$

We ignore the dye effect, $\delta_{green} = \delta_{red}$ and we take $z_j = \log_2(y_{ij\ green}) - \log_2(y_{jk\ red})$, $j = 1, 2, ..., a$, $i, k = 1, 2, ..., s$, $i \neq k$. Thus in the matrix notation, model (2) we write in the simpler form

$$\mathbf{z} = \mathbf{X}\mathbf{t} + \mathbf{q} \tag{3}$$

where $\mathbf{z} = (z_1, z_2, ..., z_a)^{'}$ is the vector of log ratios of the dye intensities, $\mathbf{t} = (t_1, t_2, ..., t_s)^{'}$ is vector of unknown treatment effects, $\mathbf{q}$ is the random vector of errors with $\mathrm{E}(\mathbf{q}) = \mathbf{0}_a$ and $\mathrm{Var}(\mathbf{q}) = \sigma^2\mathbf{I}_a$. Furthermore, $\mathbf{X}$ is an $a \times s$ design matrix, with each row containing exactly one 1 and one -1, all other elements being equal to zero. Furthermore, we said that any design $\mathbf{X}$ is singular or nonsingular depending on whether the matrix $\mathbf{X}^{'}\mathbf{X}$ is singular or nonsingular, respectively. In the considered case, the matrix $\mathbf{X}^{'}\mathbf{X}$ is singular and the observed log intensity ratios are given by $\mathbf{z}$, then the least squares estimates of the parameters are given in the vector form by $\hat{\mathbf{t}} = \left(\mathbf{X}^{'}\mathbf{X}\right)^{-}\mathbf{X}^{'}\mathbf{z}$ and $\mathrm{Var}\left(\hat{\mathbf{t}}\right) = \sigma^2\left(\mathbf{X}^{'}\mathbf{X}\right)^{-}$.

Different optimality criteria such as A-, D- and E-optimality have been proposed in the context of selecting efficient design for microarray experiments. Such criteria are functions of the matrix $\mathbf{X}^{'}\mathbf{X}$. For some general theory of experimental designs and some details of optimality criteria we refer the reader to [12]. Here, we consider D-optimal designs. Usually, the design $\mathbf{X}$ is D-optimal if it minimizes the determinant of $(\mathbf{X}^{'}\mathbf{X})^{-1}$. It is equivalent to the determining maximum of the determinant of $\mathbf{X}^{'}\mathbf{X}$. In [3], as D-optimal design consider such design which maximizes the value of $\prod_{i=1}^{s-1}\lambda_i$, where λ_i are nonzero eigenvalues of $\mathbf{X}^{'}\mathbf{X}$. On the other hand, it is worth noting that the model (3) we can treat as usually used model of the chemical balance weighing design, in which $\mathbf{z}$ is random vector of the observations, $\mathbf{X}$ is the design matrix having elements -1, 0, 1, $\mathbf{t}$ is vector of unknown parameters and $\mathbf{q}$ is random vector of errors for

which $\mathrm{E}(\mathbf{q}) = \mathbf{0}_a$ and $\mathrm{Var}(\mathbf{q}) = \sigma^2\mathbf{I}_a$. Here, in the notation of weighing designs, the number of arrays a corresponds with the number of measurements and the number of treatments s corresponds with the number of objects. The purpose of the paper is to bring together two approaches, the first one connected with two-colour microarrays experiments and the second one that deals with weighing designs. For that reason, we present the relations between weighing designs and microarray experiments. This aim is implemented on the base of the construction the design matrix of microarray experiment.

2 The main result

The important point to note here is the form of the matrix $\mathbf{X}$. For the design matrix $\mathbf{X}$ having the form as in the model (3), the matrix $\mathbf{X}'\mathbf{X}$ is singular. For that reason the following model is implemented

$$\mathbf{z}_1 = \mathbf{X}_1\mathbf{t} + \mathbf{q}_1, \tag{4}$$

where the matrix $\mathbf{X}_1$ is constructed from the matrix $\mathbf{X}$ by adding one row with elements equal to 1, which is interpreted as additionally measurement, i.e.

$$\mathbf{X}_1 = \begin{bmatrix} \mathbf{X} \\ \mathbf{1}_s' \end{bmatrix}. \tag{5}$$

Here $\mathbf{X}_1$ is an $(a+1) \times s$ design matrix, $\mathbf{X}$ is an $a \times s$ matrix and has the form as described in the model (3) and, what's more, $a = 0.5s(s-1)$. Besides, $\mathbf{z}_1 = [\mathbf{z}' \ z_0]$, $\mathbf{q}_1 = [\mathbf{q}' \ q_0]$, $\mathrm{E}(\mathbf{q}_1) = \mathbf{0}_{a+1}$, $\mathrm{Var}(\mathbf{q}_1) = \sigma^2\mathbf{I}_{a+1}$. In the formula (5), the design matrix $\mathbf{X}$ of the form given in (3), can be expressed as $\mathbf{X} = 2\mathbf{N}' - \mathbf{1}_a\mathbf{1}_s'$, where $\mathbf{N}$ is the incidence matrix of the balanced incomplete block design with the parameters v, $b = 0.5v(v-1)$, $r = v-1$, $k = 2$, $\lambda = 1$, where, according to the above notation, $v = s$ and $b = a$.

Lemma 2.1 *For $\mathbf{X}_1$ given in the form (5), the matrix $\mathbf{X}_1'\mathbf{X}_1$ is nonsingular.*

Proof. An easy computation shows that for the matrix $\mathbf{X}$ of the form given in the model (3), we have $\mathbf{X}'\mathbf{X} = s\mathbf{I}_s - \mathbf{1}_s\mathbf{1}_s'$. Besides, for the matrix $\mathbf{X}_1$ of the form (5) it is obvious that $\mathbf{X}_1'\mathbf{X}_1 = s\mathbf{I}_s$, which completes the proof.
There are many problems of practical interest of microarrays experiments. In many application there is a need to estimate the elementary contrasts of the vector of treatment effects $\mathbf{t}$. So, let us consider the function

$$\mathbf{c}'\mathbf{t} = t_i - t_k \ \text{ for } \ i,k = 1,2,...,s, \ i > k. \tag{6}$$

Lemma 2.2 *In the model (3) and (4) each function of the form (6) is estimable.*

Proof. We first consider the model (3). It is worth emphasizing, the design $\mathbf{X}$ is connected if all linear contrasts in $\mathbf{t}$ are estimable. Authors in [13] showed that the design is connected if and only if the matrix $\mathbf{X}^{'}\mathbf{X}$ has $s-1$ nonzero eigenvalues. For the design matrix $\mathbf{X}$ in the model (3), the matrix $\mathbf{X}^{'}\mathbf{X}$ has $s-1$ nonzero eigenvalues. Therefore, any contrast is estimable. Next, let us examine the model (4). For the matrix $\mathbf{X}_1$ given by the form (5) we have $\mathbf{X}_1^{'}\mathbf{X}_1 = s\mathbf{I}_s$. Thus the matrix $\mathbf{X}_1^{'}\mathbf{X}_1$ is nonsingular. It implies all treatment effects are estimable, particularly all linear contrasts are estimable, i.e. the function $\mathbf{c}^{'}\mathbf{t}$ is estimable. So, the proof is completed.

Theorem 2.3 *The form of the estimator of any function* $\mathbf{c}^{'}\mathbf{t}$ *given in (6) and it's variance is the same in the models (3) and (4).*

Proof. Our proof starts with the observation that one of possible generalized inverses of $\mathbf{X}^{'}\mathbf{X}$ is given as $\left(\mathbf{X}^{'}\mathbf{X}\right)^{-} = \frac{1}{s}\begin{bmatrix} \mathbf{I}_{s-1} + \mathbf{1}_{s-1}\mathbf{1}_{s-1}^{'} & \mathbf{0}_{s-1} \\ \mathbf{0}_{s-1}^{'} & 0 \end{bmatrix}$. In the model (3), the estimator of $\mathbf{c}^{'}\mathbf{t}$ is equal $\mathbf{c}^{'}(\mathbf{X}^{'}\mathbf{X})^{-}\mathbf{X}^{'}\mathbf{z}$. Because $\mathbf{c}^{'}(\mathbf{X}^{'}\mathbf{X})^{-} = s^{-1}\mathbf{c}^{'}$ then we obtain that the estimator of $\mathbf{c}^{'}\mathbf{t}$ is equal $s^{-1}\mathbf{c}^{'}\mathbf{X}^{'}\mathbf{z}$. The variance of the difference of the estimators of treatment effects is given as $\mathrm{Var}(\hat{t}_i - \hat{t}_k) = 2s^{-1}\sigma^2$, $i,k = 1,2,...,s$, $i > k$. Next, the procedure is to show that in the model (4) we have $\hat{\mathbf{c}^{'}\mathbf{t}} = \mathbf{c}^{'}(\mathbf{X}_1^{'}\mathbf{X}_1)^{-1}\mathbf{X}_1^{'}\mathbf{z}_1 = s^{-1}\mathbf{c}^{'}\left(\mathbf{X}^{'}\mathbf{z} + \mathbf{1}_s z_0\right)$. In addition from $\mathbf{c}^{'}\mathbf{1}_s = 0$ we obtain the same form of the estimator of $\mathbf{c}^{'}\mathbf{t}$ as in the model (3). Furthermore, $\mathrm{Var}(\hat{t}_i) = s^{-1}\sigma^2$ and simultaneously $\mathrm{Var}(\hat{t}_i - \hat{t}_k) = 2s^{-1}\sigma^2$, $i,k = 1,2,...,s$, $i > k$. In this way, we obtain the thesis of Theorem.

3 The illustration of above theory

Our paper is concerned with some issues regard to the planning of microarrays experiments. In order to place some simple results in a conceptual frame work for the designs of two-coloured experiments, we demonstrate the utility of our approach by determining optimal designs for the following example. As the application of the theory presented so far, under assumption that we ignore the dye effect, let us consider the experiment in which we have at our disposal three objects, i.e. $s = 3$ and, according to the above theory, three microarrays, i.e. $a = 3$. Afterwards, observed log intensity ratios are given by $\mathbf{z}^{'} = [z_1\ z_2\ z_3]$ and for $j = 1,2,3$, let $z_j = \log_2\left(\frac{y_{ij\ green}}{y_{jk\ red}}\right) = t_i - t_k + \epsilon_{ij\ green} + \epsilon_{jk\ red} = t_i - t_k + q_j$, where $i,j,k = 1,2,3$, $i > k$. The design matrix would have the form given as $\mathbf{X} = \begin{bmatrix} 1 & 0 & -1 \\ -1 & 1 & 0 \\ 0 & -1 & 1 \end{bmatrix}$. So $z_j = \mathbf{x}_j\mathbf{t} + q_j$, where $\mathbf{x}_j$ denotes j^{th} row of the matrix $\mathbf{X}$. The matrix $\mathbf{X}^{'}\mathbf{X}$ of this design is singular. The matrix $\mathbf{X}$ of such structure

we can treat as commonly used the matrix of the chemical balance weighing design. We refer readers to [5]. Here, authors provide detailed accounts of the relevant background concerned on singular chemical balance weighing designs. In mentioned paper, the design matrix of such structure is considered. They propose to take into account the matrix $\mathbf{X}_1$ of the form $\begin{bmatrix} 1 & 0 & -1 \\ -1 & 1 & 0 \\ 0 & -1 & 1 \\ 1 & 1 & 1 \end{bmatrix}$. This is the design matrix of the D-optimal chemical balance weighing design. Thereafter, for $\mathbf{X}$ we obtain as in (3) $\hat{\mathbf{t}} = \begin{bmatrix} \hat{t}_1 \\ \hat{t}_2 \\ \hat{t}_3 \end{bmatrix} = (\mathbf{X}'\mathbf{X})^{-}\mathbf{X}'\mathbf{z} = \frac{1}{3}\begin{bmatrix} 2 & 1 & 0 \\ 1 & 2 & 0 \\ 0 & 0 & 0 \end{bmatrix}\begin{bmatrix} 1 & 0 & -1 \\ -1 & 1 & 0 \\ 0 & -1 & 1 \end{bmatrix}\begin{bmatrix} z_1 \\ z_2 \\ z_3 \end{bmatrix}$. Thus $\hat{\mathbf{t}} = \frac{1}{3}\begin{bmatrix} z_1 + 2z_2 + z_3 \\ -z_1 + z_2 + 2z_3 \\ 0 \end{bmatrix}$. Therefore $\hat{t}_1 - \hat{t}_2 = \frac{1}{3}(2z_1 + z_2 - z_3)$, $\hat{t}_1 - \hat{t}_3 = \frac{1}{3}(z_1 + 2z_2 + z_3)$, $\hat{t}_2 - \hat{t}_3 = \frac{1}{3}(-z_1 + z_2 + 2z_3)$. For $\mathbf{X}_1$, we obtain $\hat{\mathbf{t}} = \begin{bmatrix} \hat{t}_1 \\ \hat{t}_2 \\ \hat{t}_3 \end{bmatrix} = \frac{1}{s}\mathbf{X}_1'\mathbf{z}_1 = \frac{1}{3}\begin{bmatrix} z_1 + z_2 + z_0 \\ -z_1 + z_3 + z_0 \\ -z_2 - z_3 + z_0 \end{bmatrix}$. Moreover, $\hat{t}_1 - \hat{t}_2 = \frac{1}{3}(2z_1 + z_2 - z_3)$, $\hat{t}_1 - \hat{t}_3 = \frac{1}{3}(z_1 + 2z_2 + z_3)$, $\hat{t}_2 - \hat{t}_3 = \frac{1}{3}(-z_1 + z_2 + 2z_3)$. As you can see the contrasts are the same, besides the measurement z_0 is not important. In both models, $\mathrm{Var}(\hat{t}_i - \hat{t}_k) = \frac{2}{3}\sigma^2,\ i,k = 1,2,3,\ i > k$. It means, the information on the contrasts in both models is the same.

4 Discussion

The results in this paper we obtained under the assumption that the design $\mathbf{X}_1$ in (4) is optimal in that sense that its information matrix is proportional to the identity matrix. In the literature, such design is called optimal in the sense of Hotelling, see [7], [4]. The design constructed in this form is very suitable for checking the properties of designs and allows to examine the properties of nonsingular designs, that is very easy and desirable. In addition, the design $\mathbf{X}_1$ given by the formula (4) satisfies the criterion of D-optimality in usual sense: the determinant of the information matrix of the design is maximal, so the general variance of the estimators is minimal. Hence, especially when the matrix $\mathbf{X}$ is of the form described in the model (3), the information matrix of this design is singular. In this situation, the criterion of D-optimality as the maximum of the product of the nonzero eigenvalues of the information matrix

of the design is considered. Moreover, determined variance of the estimator of comparisons of these effects is also equal. However, in the model (4) we consider the design matrix which is of full column rank, so we are working under usually established optimality criteria. This simplification may, however, have practical benefits. Taking as the design matrix the matrix of the chemical balance weighing design we receive the matrix of optimal design. Therefore, in that design we obtain the estimators having the smallest general variance. Here, some issues of the planning of experiments are discussed under assumption $a = 0.5s(s-1)$. A discussion of this assumption is needed. For $a < 0.5s(s-1)$, the product of the nonzero eigenvalues of the information matrix of the design is always smaller as the product of the nonzero eigenvalues of the information matrix of the design for $a = 0.5s(s-1)$. Thus the variance of the estimator of any contrast of comparisons of the treatment effects in the design with $a = 0.5s(s-1)$ is not bigger than for the design with $a < 0.5s(s-1)$. This is our motivation to consider $a = 0.5s(s-1)$. A major step in the statistical analysis of microarray experiments is to estimate all linear contrasts of the vector of treatment effects $\mathbf{t}$. It is worth emphasizing that the above considerations imply that in both models (3) and (4) the form of the estimator of all possible comparisons of treatment effects is the same. It demands noting that the matrix $\mathbf{X}_1$ is introduced in order to point out the properties of the design of microarray experiment. We do not expect that it will be treated as the plan of that kind of experiment. Therefore, in that case, we do not perform experiment with one more measuring as the equality of models is derived. In particular, the statement of this measurement is not contributed to assessing of treatments effects and should be treated as theoretical construction (technical operation) introduced in the purpose of analyzing the properties of experimental designs. Obviously, the choice of the best (optimal) design matrix for the microarray experiment depends on the researcher questions under investigation and the number of arrays at our disposal. For a given experimental effort and any practical constrains on the problem, we seek to optimize the choice of the design matrix on the key of biological effects and statistical aspects. This paper should be treated as proposition in the theme of studying the properties of design matrices used in the planning of microarrays experiments. In addition, it is hoped that the present endeavor will generate further interest in the above considerations.

References

[1] Atkinson A.C., Donev A.N., Tobias R.D., Optimum Experimental Designs, with SAS, Oxford: Oxford University Press, 2007.

[2] Bailey R.A., Designs for two-colour microarray experiments, Appl. Statist., 56, , 2007, 365-394.

[3] Bailey R.A., Schiffl K., Hielgers R.D., A note on robustness of D-optimal designs for two-colour microarray experiments, Journal of Statistical Planning and Inference, 143,, 2013, 1195-1202.

[4] Banerjee K.S., Weighing Designs for Chemistry, Medicine. Economics, Operations Research, Statistics. Marcel Dekker Inc., New York, , 1975.

[5] Ceranka B., Graczyk M., Notes about singular chemical balance weighing design, Acta Universitate Lodziensis, Folia Oeconomica, 235, 2010, 241-245.

[6] Churchill G.A., Fundamentals of experimental design for cDNA microarrays, Nature Genetics (Suppl.), 3, 2002, 490-495.

[7] Hotelling H., Some improvements in weighing and other experimental techniques, Ann. Math. Statist., 15, 1944, 297-306.

[8] Kerr M.K., Churchill G.A., Experimental design for gene expression microarrays. Biostatistics, 2, 2001a, 183-201.

[9] Kerr M.K., Churchill G.A., Statistical design and the analysis of gene expression microarray data. Genet. Res., 77, 2001b, 123-128.

[10] Landgrebe J., Bretz F., Brunner E., Efficient design and analysis of two-color factorial microarray design. Comput. Stat. Data Anal., 50, 2006, 499-517.

[11] Latif H., M., Bretz F., Brunner E., Robustness considerations in selecting efficient two-color microarray designs Bioinformatics, 25 (18), 2009, 2355-2361.

[12] Pukelsheim F., Optimal Designs of Experiments, JohnWiley Sons, New York, 1993.

[13] Shah K.R., Sinha B.K., Theory of Optimal Designs, Springer-Verlag, Berlin, 1989.

[14] Yang Y.H., Speed T., Design issues for cDNA microarray experiments. Nature Genetics (Suppl.), 3, 2002, 579-588.

[15] Wit E., Nobile A., Khanin R., Near-optimal designs for dual channel microarray studies, Appl. Stat., 54, 2005, 817-830.

Received: May, 2015