International Journal of Statistics and Probability; Vol. 14, No. 1; March 2025
ISSN 1927-7032 E-ISSN 1927-7040
Published by Canadian Center of Science and Education

Algorithmic Construction of Bayesian Optimal Block Designs Using the Linear Mixed Effects Model

Dibaba B. Gemechu[1], Legesse K. Debusho[2] & Linda M. Haines[3]

[1] Department of Mathematics, Statistics and Actuarial Science, Namibia University of Science and Technology, Namibia

[2] Department of Statistics, University of South Africa, South Africa

[3]Department of Statistical Sciences, University of Cape Town , South Africa

Correspondence: Dibaba B. Gemechu, Department of Mathematics, Statistics and Actuarial Science, Namibia University of Science and Technology, Namibia

Received: December 12, 2024 Accepted: February 21, 2025 Online Published: March 31, 2025

doi:10.5539/ijsp.v14n1p50 URL: https://doi.org/10.5539/ijsp.v14n1p50

Abstract

In this paper a numerical method for construction of optimal Bayesian block designs of size two is considered. The main focus is on implementing prior information on the unknown error variance and variance of random block effects to calculate the *A*- and *D*-optimal designs. It is noted from the numerical results that the *A*- and *D*-optimal Bayesian block designs are insensitive to the shape of the prior distributions.

Keywords: Bayesian optimal design, *A*- and *D*-optimal designs, linear mixed model, block designs

1. Introduction

Statistics and Computing

Optimal or efficient block designs of size two for the estimation of pairwise contrasts of treatment effects have been extensively studied in the design literature, for example for two-colour microarray experiment (Kerr & Churchill, 2001a,b; Churchill, 2002; Yang & Speed, 2002; Kerr, 2003; Wit et al., 2005; Landgrebe et al., 2006; GroΣmann & Schwabe, 2007; Bailey et al., 2013). The design construction approaches adopted in these studies are based on fixed-effects models. Furthermore, locally optimal block designs under the linear mixed effects models setting, where the block effects are assumed to be random, were discussed by Sarkar et al. (2007) and Debusho et al. (2019) for a given value of $\theta \in [0, 1)$, where θ is a function of error variance and variance of random array effects. A locally optimal design depends on a given value of θ or variance components. It however does not account for uncertainty in variance components values and hence one may be tempted to identify a design as optimal or near-optimal when it is not (Bueno Filho & Gilmour, 2007). This is the known criticism against the locally optimal design (Dokoumetzidis & Aarons, 2007). One of the practical approach to obtain the variance components is to conduct a preliminary pilot experiment and use the results to get the initial estimates, then compute the optimal designs for the main experiment using them. This approach however is not feasible in reality due to resource constraints (Gondro & Kinghorn, 2008). A Bayesian optimal design extends the locally optimal approach by allowing the practitioner to specify a prior distribution for θ or the variance components. Note however that different prior distributions on the variance components or different design criteria may yield different Bayesian designs (Lohr, 1995).

The aim of this paper is therefore to obtain Bayesian-optimal block designs with a block size of two by introducing a prior distribution for θ. Attention is restricted to two criteria, Bayesian *A*-optimality for which the expected value of the trace of the variance-covariance matrix of all possible pairwise treatment means is minimized and Bayesian *D*-optimal for which the expected value of the determinant of the variance-covariance matrix of all possible pairwise treatment means is minimized. The method that we propose relies on a treatment exchange algorithm of Debusho et al. (2019). The model, information matrix, optimality criteria and prior distribution are introduced in Section 2. The algorithm used for numerical construction of Bayesian optimal designs is discussed in Section 3. In Section 4, the results and discussions are presented, and finally some concluding remarks are given in Section 5. The notation and terminology introduced in Debusho et al. (2019) will be used throughout.

2. Preliminaries

2.1 Model and Information Matrix

Consider an experiment comprising b blocks and v treatments replicated $r_1, \ldots, r_v$ times. Furthermore, suppose the treatment effects $\boldsymbol{\tau}$ are fixed and the block effects $\mathbf{b}$ are random. Then following Debusho et al. (2019), the appropriate linear mixed effects model can be expressed in a matrix form as

$$\mathbf{y} = \mathbf{1}\,\mu + \mathbf{X}_\tau\,\boldsymbol{\tau} + \mathbf{X}_b\,\mathbf{b} + \mathbf{e}, \tag{1}$$

where $\mathbf{y}$ is the $n \times 1$ response vector with $n = b\,v$, μ is the overall mean, $\mathbf{1}$ denotes the $n \times 1$ vector of ones, $\boldsymbol{\tau}$ is a $v \times 1$ vector of fixed treatment effects with attendant $n \times v$ design matrix $\mathbf{X}_\tau$, $\mathbf{b}$ is a $b \times 1$ vector of random block effects with attendant $n \times b$ design matrix and $\mathbf{e}$ is an $n \times 1$ vector of error terms. The random effects vector $\mathbf{b}$ is taken to be distributed as $\mathcal{N}(\mathbf{0},\, \sigma_b^2\,\mathbf{I}_b)$ and the error term $\mathbf{e}$ as $\mathcal{N}(\mathbf{0},\, \sigma^2\,\mathbf{I}_n)$, where $\mathbf{I}_b$ and $\mathbf{I}_n$ denote the identity matrices of order b and n, respectively, and $\mathbf{b}$ and $\mathbf{e}$ are assumed to be independent. Under the above assumptions, the variance of the responses is given by $Var(\mathbf{y}) = \mathbf{V} = \sigma_b^2\,\mathbf{X}_b\,\mathbf{X}_b' + \sigma^2\,\mathbf{I}_n$ and its inverse is given by

$$\mathbf{V}^{-1} = \frac{1}{\sigma^2}\left(\mathbf{I}_n - \mathbf{X}_b\,\mathbf{W}\,\mathbf{X}_b'\right),$$

where $\mathbf{W} = diag\left(\sigma_b^2/(\sigma^2 + k_1\,\sigma_b^2), \ldots, \sigma_b^2/(\sigma^2 + k_b\,\sigma_b^2)\right)$ and $k_i, i = 1, \ldots, b$ are block sizes.

Observe that $\mathbf{X}_\tau'\,\mathbf{1} = \mathbf{r} = (r_1, \ldots, r_v)'$, $\mathbf{X}_\tau'\,\mathbf{X}_b = \mathbf{N}$ are the vector of treatment replications and treatments-blocks incidence matrix, respectively, and $\mathbf{X}_\tau'\,\mathbf{X}_\tau = \mathbf{R} = diag(r_1, \ldots, r_v)$. Ignoring the constant $1/\sigma^2$, the general treatment information matrix, also called the $\mathbf{C}$-matrix, is given by

$$\mathbf{C} = \mathbf{R} - \mathbf{NWN}' - (\mathbf{r} - \mathbf{WNk})\,(n - \mathbf{k}'\mathbf{Wk})^{-}\,(\mathbf{r} - \mathbf{WNk})'\,, \tag{2}$$

where $\mathbf{k} = (k_1, \ldots, k_b)$ is vector of block sizes. Details of the derivation of $\mathbf{C}$ is given in Appendix A. Since the block size is assumed to be 2, no treatment occurs more than once in a block and hence the design layout is a binary block design. Therefore, performing some straightforward matrix algebra, expression (2) will be reduced to

$$\mathbf{C} = \mathbf{R} - \frac{1}{k}\,\mathbf{N}\,\mathbf{N}' + \rho\left(\frac{1}{k}\,\mathbf{N}\,\mathbf{N}' - \frac{1}{b\,k}\,\mathbf{r}\,\mathbf{r}'\right), \tag{3}$$

where $k = 2$ is the block size and $\theta = \sigma^2/(\sigma^2 + k\,\sigma_b^2)$. Note also that $\theta = 0$ corresponds to a fixed effects model with $\mathbf{C} = \mathbf{R} - \frac{1}{k}\,\mathbf{N}\,\mathbf{N}'$.

The matrix $\mathbf{C}$ is symmetric and positive semi-definite and, since $\mathbf{C}\mathbf{1}_v = \mathbf{0}$, has $rank(\mathbf{C}) \leq v - 1$. In the present study attention is restricted to designs for which the associated information matrix has rank $v-1$. It thus follows that the designs are connected and that all treatment contrasts, and in particular pairwise differences, are estimable (Haines, 2015). In this paper, attention is restricted to connected block designs.

2.2 Optimal Design Criteria

One of the aims of an experiment is to conduct pairwise treatment comparisons. Thus it is natural to focus on design criteria which are based on the variance of the estimated pairwise treatment differences and to select designs from a set of competing designs comprising b blocks and v treatments for which the criterion of interest is in some sense optimal. Consider therefore all possible pairwise treatment differences expressed as $\mathbf{T}\boldsymbol{\alpha}$ where $\mathbf{T}$ is a $\binom{v}{2} \times v$ matrix with each row comprising an element 1, an element -1 and all other elements 0, and is such that $\mathbf{T}'\,\mathbf{T} = v\,\mathbf{I}_v - \mathbf{J}_v$ (Dey, 2010). Then the variance-covariance matrix of the best linear unbiased estimator of $\mathbf{T}\boldsymbol{\alpha}$ is given by $\mathbf{T}\,\mathbf{C}^-\,\mathbf{T}'$, where $\mathbf{C}^-$ is an arbitrary g-inverse of $\mathbf{C}$.

Suppose that $\mathcal{D} = \mathcal{D}(v, b)$ represents the class of connected block designs with b blocks of size two and v treatments each member of $\mathcal{D}$ keeps $\mathbf{T}\boldsymbol{\alpha}$ estimable, and assume that the model in (1) can be used for the observations generated by $\mathcal{D}$. Then a design $d_{BA}^* \in \mathcal{D}$ is said to be Bayesian A-optimal if it minimizes

$$\Psi_{BA}(d) = E_\theta\{trace(\mathbf{T}\,\mathbf{C}^-\,\mathbf{T}')\},$$

where the expectation is over θ, that is if it minimizes the average of the trace of the variance-covariance matrix of the BLUE of $\mathbf{T}\boldsymbol{\alpha}$, equivalently if it minimizes

$$\Psi_{BA}(d) = \int trace(\mathbf{T}\,\mathbf{C}^-\mathbf{T}')\,f_\theta(\theta)\,d\,\theta, \tag{4}$$

where $f_\theta(\theta)$ is a prior distribution of θ. In addition a design $d_{BD}^* \in \mathcal{D}$ is said to be Bayesian D-optimal if it minimizes

$$\Psi_{BD}(d) = E_\theta\{|\mathbf{T}\,\mathbf{C}^-\,\mathbf{T}'|\},$$

where the expectation is over θ (Lohr, 1995), equivalently if it minimizes

$$\Psi_{BD}(d) = \int |\mathbf{T}\,\mathbf{C}^-\,\mathbf{T}'|\,f_\theta(\theta)\,d\,\theta. \tag{5}$$

2.3 Prior Ddistribution of θ

In this paper we used the inverse-Gamma (IG) distributions to describe the uncertainty for the variance components σ_b^2 and σ^2 (Verdinelli, 1996). Specifically,

$$\sigma^2 \sim Inverse-Gamma(a_1, b_1) \quad \text{and} \quad \sigma_b^2 \sim Inverse-Gamma(a_2, b_2)$$

where a_1, b_1 and a_2, b_2 are the parameters of the inverse-Gamma distributions. The assignment of univariate IG distributions for the variance components follows from the model assumption that **b** and **e** are independent. Hence, the joint probability density function of (σ^2, σ_b^2) is given by $f_{\sigma^2,\sigma_b^2} = f_{\sigma^2} \times f_{\sigma_b^2}$. Recall that $\theta = \sigma^2/(\sigma^2 + k\,\sigma_b^2)$. Now let $\alpha = \sigma^2 + k\,\sigma_b^2$, hence $\sigma^2 = \alpha\,\theta$ and $\sigma_b^2 = \alpha\,(1-\theta)/k$. The joint probability density function (pdf) of (θ, α), $f_{\theta,\alpha}(\theta, \alpha)$, can be derived using Jacobian transformation and then the marginal probability density function of θ using $f_\theta(\theta) = \int_0^\infty f_{\theta,\alpha}(\theta, \alpha) d\alpha$. Thus

$$f_\theta(\theta) = \frac{\theta^{a_2-1}\,(1-\theta)^{a_1-1}}{\overline{b}^{\,a_2}\,B(a_1, a_2)} \times \left(1 + \frac{(1-\overline{b})\,\theta}{\overline{b}}\right)^{-(a_1+a_2)}, \tag{6}$$

where $\overline{b} = b_1/k\,b_2 > 0$ and $B(a_1, a_2)$ is a Beta function with parameters a_1 and a_2. Details of the derivation of $f_\theta(\theta)$ are given in Appendix B.

The evaluations of the integrals in (4) and (5) using the expression of $f_\theta(\theta)$ in (6) are computer intensive. However, if $f_\theta(\theta)$ is some known distribution, a Monte Carlo integral approximation could be implemented by drawing a Monte Carlo sample from $f_\theta(\theta)$. For example, consider $b_1 = k\,b_2$, i.e. $\overline{b} = 1$. Then it follows from expression (6) that

$$f_\theta(\theta) = \frac{\theta^{a_2-1}\,(1-\theta)^{a_1-1}}{B(a_1, a_2)},$$

which is a Beta distribution with parameters a_1 and a_2, denoted $Beta(a_1, a_2)$, where a_2 and a_1 are the shape and scale parameters, respectively. We used the different Beta distributions to compute the Bayesian optimal designs numerically. The following general steps (Bueno Filho & Gilmour, 2007) are used for the numerical evaluation of the Bayesian criterion values:

1. For a given connected block design $d \in \mathcal{D}(v, b, k)$, compute the information matrix as a function of θ;
2. Express the criterion conditioned on θ;
3. Assign a prior distribution of θ; and
4. Compute the Bayesian criterion value by integrating out the θ using its corresponding prior distribution.

Note that the optimal designs depend on the prior information of the variance components through the prior distribution of θ.

3. Algorithm for Constructing Bayesian Optimal Designs

In this paper we have used the following algorithm, which has been adopted from the treatment exchange algorithm of Debusho et al. (2019) for a search of Bayesian *A*- and *D*-optimal block designs using a Beta distribution as a prior for θ. For a given parametric combination (v, b) and prior distribution of θ, the steps used in the algorithm are presented below:

Step 1: Selection of a connected initial block design:

(a) Select an initial design with b blocks at random.

(b) If the initial design does not comprises all v treatments, that is if $rank(\mathbf{R}) < v$, where $\mathbf{R}$ is a diagonal matrix of treatment replication, go to Step 1(a) else go to the next step. At this stage, we can use $\theta = 0$ to simplify the search for initial connected block design.

(c) If the initial design is not connected, that if $rank(\mathbf{C}) < v - 1$, go to Step 1(a).

(d) For a generated initial connected block design, derive the optimality criterion as a function of θ and compute its corresponding Bayesian optimality criterion score values using the Monte Carlo Method.

Step 2: Treatment Exchange Procedure: Given initial connected design comprises of all v treatments

(a) Set $j = 1$ for the first block ($j = 1, \ldots, b$) and $i = 1$ for one of the treatment ($i = 1, 2$) in the jth block. Delete the treatment in the (i, j)th cell of the initial design and replace the deleted treatment by the rest of $v - 2$ treatments not contained in the jth array taking one at a time. List all new possible designs that will be generated as a result of treatment exchange.

(b) Check the connectedness of all the candidate designs in Step 2(a). For simplicity $\theta = 0$ can be used to check connectedness. For the connected designs, compute their corresponding Bayesian optimality score values using the Monte Carlo Method and select the best design.

(c) Set $i = 2$ and repeat the process from Step 2(a) using the improved design.

(d) If $j < b$, set $j = j + 1$ and go to Step 2(a) using the improved design in Step 2(c). Otherwise, select the design for which the score value is optimal.

Step 3: Run Steps 1 and 2 for a fixed number of iterations (say $n = 100$), this is fixed by the user. Stop.

In Steps 1(d) and 2(b) the following procedures are followed for computation of optimality criterion scores:

(i) Take a random sample of size M from the prior distribution of θ, i.e. a Monte Carlo sample. Thus, we have $\boldsymbol{\theta} = (\theta_1, \theta_2, \ldots, \theta_M)$. We used Monte Carlo sample of size $M = 10,000$ for all the examples described in this paper.

(ii) For each selected values of $\theta_j \in \theta$ compute the A-optimality criterion score

$$\Psi_{BA}(d/\theta_l) = trace(\mathbf{T}\,\mathbf{C}^{-}(\theta_l)\,\mathbf{T}'), \quad l = 1, 2, \ldots, M.$$

(iii) Compute the average of the optimality criterion scores as

$$E(\Psi_{BA}(d)) \approx \frac{1}{M} \sum_{l=1}^{M} \Psi_{BA}(d/\theta_l).$$

Note that this is a Monte Carlo integral approximation of the optimality criterion given in expression (4).

Note also that the same algorithm can be invoked in order to construct Bayesian D-optimal block designs. The algorithm for both A- and D-optimality has been coded and implemented in the R-programming language.

4. Results and Discussion

The Bayesian optimal designs were calculated for the parametric combinations $3 \leq v \leq 20$ and $b = v$ the same number of arrays as treatments, $b = v + 1$ one extra array and $b = v + 2$ two extra arrays using nine different Beta priors for θ. These values were selected for illustration purpose, however our R code can also generate the optimal designs for other parametric combinations. The prior distributions vary in shapes and includes one uniform on $[0, 1)$ $Beta(1, 1)$, two positively skewed ($Beta(0.5, 1.5)$ and $Beta(1.25, 5)$), two negatively skewed ($Beta(1.5, 0.5)$ and $Beta(5, 1.5)$), two bimodal ($Beta(0.3, 0.3)$ and $Beta(0.5, 0.75)$) and two symmetric ($Beta(5, 5)$ and $Beta(5, 10)$) which tend to normality. The Bayesian A- and D-optimal block designs results are summarized as follows:

(i) When $3 \leq v = b \leq 20$, the A- and D-optimal designs are the loop designs $C_v(v)$ for all the priors used (see Figure 1(a)), where $C_s(v)$ represents a design for v treatments in v arrays whose graphs contain a circuit of length s (Bailey, 2007).

(ii) When $b = v + 1; 3 \leq v \leq 20$, the A- and D-optimal designs are the parallel path designs $PP(x, y, z)$ of lengths x, y and z between two vertices with $x + y + z = b$ for all the priors used (see Figure 1(b)).

(iii) When $b = v + 2; 3 \leq v \leq 20$, the A- and D-optimal designs are the complete graph K_4 on four vertices with each of its six edges replace by paths (Bailey 2007, p.379) for all the priors used (see Figure 1(c)).

The A- and D-optimal designs are robust to specification of the prior distribution, so they are insensitive to the exact shape of the prior distribution.

The numerical results in Debusho et al. (2019) show that for some set of θ in the interval $[0, 1)$, v and b values, the A-optimal designs are the loop designs, parallel path designs $PP(x, y, z)$ of lengths x, y and z between two vertices with $x + y + z = b$ and the complete graph K_4 on four vertices with each of its six edges replace by paths for $v = b$, $b = v + 1$ and $b = v + 2$, respectively. However, for the parametric combinations considered in Debusho et al. (2019) the D-optimal designs are invariant of the θ values, for example for $v = b$ the D-optimal designs are the loop designs for all θ in the interval $[0, 1)$. Hence, for such v and b values the results might suggest that there is no advantage to using the Bayesian method for construction of D-optimal block designs of size two.

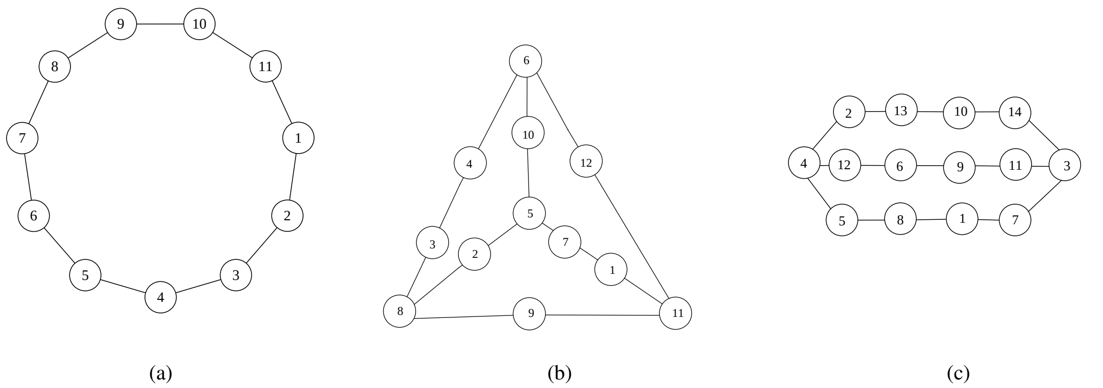

Figure 1. (a) Loop design for eleven treatments $C_{11}(11)$; (b) Parallel path design $PP(5, 5, 5)$ for 14 treatments in 15 blocks; and (c) Complete graph K_4 for 12 treatments in 14 blocks

5. Concluding Remarks

In this paper we introduced a numerical method for construction of optimal Bayesian block designs of size two. The main focus is on implementing prior information on the error variance σ^2 and the variance of random array effects σ_b^2, through $\theta = \sigma^2/(\sigma^2 + k\sigma_b^2)$, where k is the block size, by using Monte Carlo samples from appropriate prior distribution. The results presented, show that for Beta priors with different shapes, the A- and D-optimal designs are the loop designs, $PP(x, y, z)$ of lengths x, y and z between two vertices with $x + y + z = b$ and the complete graph K_4 for $v = b$, $b = v + 1$ and $b = v + 2$, respectively.

In this paper it is assumed that there is no row effect. However, if we take into account the effect of row then the two size block design can be considered as a $2 \times b$ row-column design (Bailey, 2007). The method to the construction of optimal Bayesian block designs reported here has been extended to the row-column designs and results relating to this extension will be reported elsewhere.

Acknowledgements

The authors acknowledge journal's editor for the feedback provided.

Authors contributions

DBG Dr. drafted the manuscript, LKD and LMH revised and edited the manuscript. All authors read and approved of the final manuscript. All authors contributed equally to the study design.

Funding

The publication of this article was supported financially by the Department of Mathematics, Statistics and Actuarial Science at Namibia University of Science and Technology (NUST).

Competing interests

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper. Informed consent Obtained.

Ethics approval

The Publication Ethics Committee of the Canadian Center of Science and Education. The journals policies adhere to the Core Practices established by the Committee on Publication Ethics (COPE).

Provenance and peer review

Not commissioned; externally double-blind peer reviewed.

Data sharing statement

No additional data are available.

References

Bailey, R. A. (2007). Designs for two-colour microarray experiments. *Journal of the Royal Statistical Society Series C: Applied Statistics*, *56*(4), 365–394.

Bailey, R. A., Schiffl, K., & Hilgers, R.-D. (2013). A note on robustness of d-optimal block designs for two-colour microarray experiments. *Journal of Statistical Planning and Inference*, *143*(7), 1195–1202.

Bueno Filho, J. S. d. S., & Gilmour, S. G. (2007). Block designs for random treatment effects. *Journal of statistical planning and inference*, *137*(4), 1446–1451.

Churchill, G. A. (2002). Fundamentals of experimental design for cdna microarrays. *Nature genetics*, *32*(4), 490–495.

Debusho, L. K., Gemechu, D. B., & Haines, L. M. (2019). Algorithmic construction of optimal block designs for two-colour cdna microarray experiments using the linear mixed effects model. *Communications in Statistics-Simulation and Computation*, *48*(7), 1948–1963.

Dey, A. (2010). *Incomplete block designs*. World Scientific.

Dokoumetzidis, A., & Aarons, L. (2007). Bayesian optimal designs for pharmacokinetic models: sensitivity to uncertainty. *Journal of biopharmaceutical statistics*, *17*(5), 851–867.

Gondro, C., & Kinghorn, B. P. (2008). Optimization of cdna microarray experimental designs using an evolutionary algorithm. *IEEE/ACM Transactions on Computational Biology and Bioinformatics*, *5*(4), 630–638.

Gradshteyn, I., & Ryzhik, I. (1994). Table of integrals, series and products. *New York: Academic Press*.

GroΣmann, H., & Schwabe, R. (2007). *The relationship between optimal designs for microarray and paired comparison experiments*. Univ., Fak. für Mathematik.

Haines, L. M. (2015). Introduction to linear models. *Handbook of Design and Analysis of Experiments (Edited by A. Dean, M. Morris, J. Stufken and D. Bingham)*, 63–95.

Kerr, M. K. (2003). Design considerations for efficient and effective microarray studies. *Biometrics*, *59*(4), 822–828.

Kerr, M. K., & Churchill, G. A. (2001a). Experimental design for gene expression microarrays. *Biostatistics*, *2*(2), 183–201.

Kerr, M. K., & Churchill, G. A. (2001b). Statistical design and the analysis of gene expression microarray data. *Genetics Research*, *77*(2), 123–128.

Landgrebe, J., Bretz, F., & Brunner, E. (2006). Efficient design and analysis of two colour factorial microarray experiments. *Computational statistics & data analysis*, *50*(2), 499–517.

Lohr, S. L. (1995). Optimal bayesian design of experiments for the one-way random effects model. *Biometrika*, *82*(1), 175–186.

Sarkar, A., Parsad, R., Rathore, A., & Gupta, V. (2007). Efficient block designs for 2-colour microarray experiments. *Journal of Indian Society of Agricultural Statistics*, *61*(2), 227–248.

Verdinelli, I. (1996). *Bayesian design for the normal linear model with unknown error variance* (Tech. Rep.). Technical Report 647, Carnegie Mellon Univ.

Wit, E., Nobile, A., & Khanin, R. (2005). Near-optimal designs for dual channel microarray studies. *Journal of the Royal Statistical Society Series C: Applied Statistics*, *54*(5), 817–830.

Yang, Y. H., & Speed, T. (2002). Design issues for cdna microarray experiments. *Nature Reviews Genetics*, *3*(8), 579–588.

Appendix A: Derivation of C-matrix

The model in (1) can also be expressed in a general linear mixed model form as

$$\mathbf{y} = \mathbf{X}\boldsymbol{\beta} + \mathbf{Z}\,\mathbf{b} + \mathbf{e} \tag{7}$$

where $\mathbf{X} = [\mathbf{1}\ \ \mathbf{X}_\tau]$, $\boldsymbol{\beta}' = (\mu\ \boldsymbol{\theta}')$ and $\mathbf{Z} = \mathbf{X}_b$. Under the assumptions $\mathbf{b} \sim \mathcal{N}(\mathbf{0},\ \sigma_b^2\,\mathbf{I}_b)$, $\mathbf{e} \sim \mathcal{N}(\mathbf{0},\ \sigma^2\,\mathbf{I}_n)$ and $\mathbf{b}$ and $\mathbf{e}$ are independent, the mean and variance of $\mathbf{y}$ are given by $E(\mathbf{y}) = \mathbf{X}\boldsymbol{\beta}$ and $\mathbf{V} = \sigma_b^2\,\mathbf{X}_b\,\mathbf{X}_b' + \sigma^2\,\mathbf{I}_n$, respectively. The normal equations related to vector of fixed effects, $\boldsymbol{\beta}$, in model (7) are readily derived and can be expressed succinctly as

$$\left(\mathbf{X}'\,\mathbf{V}^{-1}\,\mathbf{X}\right)\hat{\boldsymbol{\beta}} = \mathbf{X}'\,\mathbf{V}^{-1}\,\mathbf{y}$$

where $\hat{\boldsymbol{\beta}}$ is the maximum likelihood estimator of $\boldsymbol{\beta}$ and this can be re-expressed as

$$\begin{pmatrix} \mathbf{1}_n'\,\mathbf{V}^{-1}\,\mathbf{1}_n & \mathbf{1}_n'\,\mathbf{V}^{-1}\,\mathbf{X}_\tau \\ \mathbf{X}_\tau'\,\mathbf{V}^{-1}\,\mathbf{1}_n & \mathbf{X}_\tau'\,\mathbf{V}^{-1}\,\mathbf{X}_\tau \end{pmatrix}\begin{pmatrix} \hat{\mu} \\ \hat{\boldsymbol{\tau}} \end{pmatrix} = \begin{pmatrix} \mathbf{1}_n'\,\mathbf{V}^{-1}\,\mathbf{y} \\ \mathbf{X}_\tau'\,\mathbf{V}^{-1}\,\mathbf{y} \end{pmatrix} \tag{8}$$

Furthermore, performing some straightforward matrix algebra, the reduced normal equations for treatment effects, $\boldsymbol{\tau}$, eliminating the mean can be given by

$$\left[\mathbf{X}_\tau'\,\mathbf{V}^{-1}\,\mathbf{X}_\tau - \left(\mathbf{X}_\tau'\,\mathbf{V}^{-1}\,\mathbf{1}_n\right)\left(\mathbf{1}_n'\,\mathbf{V}^{-1}\,\mathbf{1}_n\right)^{-1}\mathbf{1}_n'\,\mathbf{V}^{-1}\,\mathbf{X}_\tau\right]\hat{\boldsymbol{\tau}} =$$

$$\left[\mathbf{X}_\tau'\,\mathbf{V}^{-1} - \left(\mathbf{X}_\tau'\,\mathbf{V}^{-1}\,\mathbf{1}_n\right)\left(\mathbf{1}_n'\,\mathbf{V}^{-1}\,\mathbf{1}_n\right)^{-1}\mathbf{1}_n'\,\mathbf{V}^{-1}\right]\mathbf{y} \tag{9}$$

which we write in abbreviated form as

$$\mathbf{C}\,\hat{\boldsymbol{\tau}} = \mathbf{q} \tag{10}$$

with $\mathbf{C} = \mathbf{X}_\tau'\,\mathbf{V}^{-1}\,\mathbf{X}_\tau - \left(\mathbf{X}_\tau'\,\mathbf{V}^{-1}\,\mathbf{1}_n\right)\left(\mathbf{1}_n'\,\mathbf{V}^{-1}\,\mathbf{1}_n\right)^{-1}\mathbf{1}_n'\,\mathbf{V}^{-1}\,\mathbf{X}_\tau$ and $\mathbf{q}$ is the term in the right hand side of expression (9).

Using the inverse of the variance matrix, $\mathbf{V} = \sigma_b^2\,\mathbf{X}_b\,\mathbf{X}_b' + \sigma^2\,\mathbf{I}_n$, which is given by

$$\mathbf{V}^{-1} = \frac{1}{\sigma^2}\left(\mathbf{I}_n - \mathbf{X}_b\,\mathbf{W}\,\mathbf{X}_b'\right)$$

where $\mathbf{W} = diag\left(\sigma_b^2/(\sigma^2 + k_1\,\sigma_b^2), \ldots, \sigma_b^2/(\sigma^2 + k_b\,\sigma_b^2)\right)$, the terms in the $\mathbf{C}$ matrix can be expressed as

1. $\mathbf{X}_\tau'\,\mathbf{V}^{-1}\,\mathbf{X}_\tau = \mathbf{R} - \mathbf{N}\,\mathbf{W}\,\mathbf{N}'$,
2. $\mathbf{X}_\tau'\,\mathbf{V}^{-1}\,\mathbf{1}_n = \mathbf{r} - \mathbf{N}\,\mathbf{W}\,\mathbf{k}$, and
3. $\mathbf{1}_n'\,\mathbf{V}^{-1}\,\mathbf{1}_n = n - \mathbf{k}'\,\mathbf{W}\,\mathbf{k}$.

The above results follows the general properties $\mathbf{X}_\tau'\mathbf{1} = \mathbf{r}$, $\mathbf{X}_\tau'\mathbf{X}_\tau = \mathbf{R}$, $\mathbf{X}_\tau'\mathbf{X}_b = \mathbf{N}$ and $\mathbf{X}_b'\mathbf{1} = \mathbf{k}$. Thus, substituting these results in expression (9) it is readily follows, ignoring the constant $\frac{1}{\sigma^2}$, that

$$\mathbf{C} = \mathbf{R} - \mathbf{N}\,\mathbf{W}\,\mathbf{N}' - (\mathbf{r} - \mathbf{N}\,\mathbf{W}\,\mathbf{k})\,(n - \mathbf{k}'\,\mathbf{W}\,\mathbf{k})^{-}\,(\mathbf{r} - \mathbf{N}\,\mathbf{W}\,\mathbf{k})'\,.$$

Appendix B

Assume that σ^2 and σ_b^2 have independent inverse Gamma (IG) distributions, say $\sigma^2 \sim IG(a_1, b_1)$ and $\sigma_b^2 \sim IG(a_2, b_2)$, where $a_1 > 0, b_1 > 0, a_2 > 0$ and $b_2 > 0$, a_1 and a_2 are the shape parameters and b_1 and b_2 are scale parameters. The independence of σ^2 and σ_b^2 implies that their joint probability density function of (σ^2, σ_b^2), f_{σ^2,σ_b^2} is given by

$$f_{\sigma^2,\sigma_b^2}(\sigma^2, \sigma_b^2) = \frac{b_1^{a_1} b_2^{a_2}}{\Gamma(a_1)\Gamma(a_2)}(\sigma^2)^{-a_1-1}(\sigma_b^2)^{-a_2-1} exp\left(-\left(\frac{b_1}{\sigma^2} + \frac{b_2}{\sigma_b^2}\right)\right).$$

Recall that $\theta = \sigma^2/(\sigma^2 + k\,\sigma_b^2)$. Now let $\alpha = \sigma^2 + k\,\sigma_b^2$, hence $\sigma^2 = \alpha\,\theta$ and $\sigma_b^2 = \alpha\,(1-\theta)/k$. Then applying the bivariate transformation using the Jacobian, the joint probability density function (pdf) of θ and α, $f_{\theta,\alpha}(\theta, \alpha)$ is given by

$$\begin{aligned} f_{\theta,\alpha}(\theta, \alpha) &= |J| \times f^{-1}_{\sigma^2,\sigma_b^2}(\sigma^2, \sigma_b^2) \qquad (11) \\ &= \frac{b_1^{a_1} b_2^{a_2}}{\Gamma(a_1)\Gamma(a_2)}\theta^{-a_1-1}(1-\theta)^{-a_2-1}k^{a_2}\alpha^{-(a_1+a_2+1)} exp\left(-\left(\frac{a_1(1-\theta) + k\,b_2\theta}{\alpha\,\theta(1-\theta)}\right)\right) \qquad (12) \end{aligned}$$

where $\sigma^2 = \alpha\,\theta$ and $\sigma_b^2 = \alpha\,(1-\theta)/k$ and hence $J = -\alpha/k$. Some straightforward algebra yields the pdf of θ as

$$f_\theta(\theta) = \frac{\theta^{a_2-1}(1-\theta)^{a_1-1}}{\overline{b}^{\,a_2} B(a_1, a_2)}\left(1 + \frac{(1-\overline{b})\theta}{\overline{b}}\right)^{-(a_2+a_1)}$$

where $\overline{b} = \frac{b_1}{k\,b_2} > 0$, and $B(a_1, a_2) = \frac{\Gamma(a_1)\Gamma(a_2)}{\Gamma(a_1+a_2)}$. Observe that $f_\theta(\theta) \geq 0$ for all $\theta \in (0, 1)$ and

$$\begin{aligned} \int_0^1 f_\theta(\theta)d\theta &= \frac{1}{\overline{b}^{a_2} B(a_1, a_2)} \int_0^1 \theta^{a_2-1}(1-\theta)^{a_1-1}\left(1 + \frac{(1-\overline{b})\theta}{\overline{b}}\right)^{-a_2-a_1} d\theta \qquad (13) \\ &= \frac{1}{\overline{b}^{\,a_2} B(a_1, a_2)}\left(1 + \frac{(1-\overline{b})}{\overline{b}}\right)^{-a_2} B(a_2, a_1) \\ &= 1 \end{aligned}$$

since $\int_0^1 x^{a_5-1}(1-x)^{a_6-1}(1+a_7x)^{-a_5-a_6}dx = (1+a_7)^{-a_5}B(a_5, a_6)$, for $a_5 > 0, a_6 > 0,\ a_7 > -1$ (Gradshteyn & Ryzhik, 1994).