Optimal block designs for experiments with responses drawn from a Poisson distribution

Stephen Bush (Stephen.Bush@uts.edu.au)
Department of Mathematical and Physical Sciences,
University of Technology Sydney, Australia

Katya Ruggiero (k.ruggiero@auckland.ac.nz)
Department of Statistics,
University of Auckland, New Zealand

January 5, 2016

Abstract

Optimal block designs for additive models achieve their efficiency by dividing experimental units among relatively homogenous blocks and allocating treatments equally to blocks. Responses in many modern experiments, however, are drawn from distributions such as the one- and two-parameter exponential families, e.g., RNA sequence counts from a negative binomial distribution. These violate additivity. Yet, designs generated by assuming additivity continue to be used, because better approaches are not available, and because the issues are not widely recognised. We solve this problem for single-factor experiments in which treatments, taking categorical values only, are arranged in blocks and responses drawn from a Poisson distribution. We derive expressions for two objective functions, based on D_A- and C-optimality, with efficient estimation of linear contrasts of the fixed effects parameters in a Poisson generalised linear mixed model (GLMM) being the objective. These objective functions are shown to be computational efficient, requiring no matrix inversion. Using simulated annealing to generate Poisson GLMM-based locally optimal designs, we show that the replication numbers of treatments in these designs are inversely proportional to the relative magnitudes of the treatments' expected counts. Importantly, for non-negligible treatment effect sizes, Poisson GLMM-based optimal designs may be substantially more efficient than their classically optimal counterparts.

1 Introduction

The introduction of gene expression microarrays (Schena et al., 1995) towards the end of the twentieth century initiated the start of a biotechnology revolution of rapidly evolving instruments capable of profiling a wide range of different molecular species at the cellular level. Today, high resolution instruments, such as next-generation sequencing (NGS) technologies (Craig et al., 2008), are capable of generating counts of individual copies of, for example, different gene transcripts. Experiments using these technologies are relatively expensive, resulting in studies with low numbers of biological replicates of treatments, making the efficient statistical design of such experiment critical. Yet, with few exceptions, classically optimal designs based on the assumption of unit-treatment additivity (i.e. functional independence of variances and covariances on their means) continue to be used.

Optimal statistical designs are central to conducting efficient comparative experiments, enabling contrasts of treatment parameters to be estimated without bias and minimum variance while requiring minimum effort, subjects, or other resources. A rich body of literature on optimal designs (John and Williams, 1995; Atkinson et al., 2007) has grown since the creation of this field of statistics (Smith, 1918). Until very recently however, the criteria used to define optimal designs has depended on the assumption of unit-treatment additivity.

Under additivity, optimal designs achieve their efficiency by dividing the experimental material into relatively homogeneous blocks and allocating treatments equally to blocks. Full efficiency, whereby a model's treatment parameters can be estimated independently of its block parameters, is attained when treatments can be arranged in a complete block design (Fisher, 1926), i.e. each treatment occurs equally frequently, usually once, in each block. However, as already noted, responses in many modern experiments are drawn from distributions such as the one and two parameter exponential families, e.g., RNA sequence counts from a negative binomial distribution (Auer and Doerge, 2010). These violate additivity. Our current work on

single-factor experiments shows that optimal block designs from the classical setting can be importantly non-optimal when additivity is violated.

Cox (1988) was arguably the first to consider the problem of randomised experiments in the context of responses drawn from an exponential family distribution, with the objective being the difference between treatment groups in the canonical parameter. The approach taken by Cox is to use conditioning to eliminate the blocking effects, and argued that local arguments can be made when in the presence of small effects and when asymptotic maximum likelihood theory is reasonable. The author also suggests that treating blocking as a random variable, which is the approach of this paper, is reasonable.

The last decade has witnessed increasing interest in research on methods for the optimal design of experiments based on the structure of general exponential distributions. In 2006, Khuri et al. presented a comprehensive review of design issues for generalised linear models (GLMs) in the absence of any random parameters, discussing the dependence of optimal designs on the values of the canonical parameters of the model; a problem which persists for generalised linear mixed models. Using compromise criteria, Woods et al. (2006) proposed a method for finding exact designs robust to misspecification of the model's functional form for experiments involving several explanatory variables, where the factors take values along a bounded continuum. Exact designs constrain the weights placed on each treatment combination so that, for a given sample size, the replication number of each treatment (combination) is an integer (following the terminology of Atkinson et al. (2007)). In contrast, continuous designs do not use this constraint, and replication numbers are obtained by making nearest integer approximations. Russell et al. (2009) present results for generating D-optimal continuous designs for Poisson regression where there are several continuous bounded factors. They also discuss the implementation of compromise designs to obtain designs that are robust to parameter misspecification. Niaparast and Schwabe (2013) investigated Poisson regression with a continuous predictor and random intercept. They argued that finding an optimal design using standard likelihood methods is cumbersome, even for simple models.

The optimal block designs of experiments with multiple bounded continuous factors and correlated non-normal responses was first considered by Woods and Van de Ven (2011). They used generalised estimating equations to incorporate block effects into the variance estimate for fixed effect parameters for GLMs. They considered both exchangeable and autoregressive correlation patterns, and presented two strategies for constructing block designs. The first strategy uses simulated annealing (Kirkpatrick (1984), Haines (1987) and Woods (2010)), while the second allocates the runs of the optimal unstructured design to blocks in an optimal way.

Yang and Mandal (2015) present results that give D-optimal continuous factorial designs for logistic regression, and consider the use of exchange algorithms to find D-optimal exact designs in the absence of blocking variables.

To date, work in developing methods for generating optimal block designs for experiments with responses drawn from an exponential family distribution has been carried out exclusively in the context of response surface models. We consider, in contrast, block designs for the much more widely applicable class of designs in which the values of the factor levels are fixed at the outset and play no role in design optimality. As far as we are aware, there are currently no methods available for generating optimal block designs in this setting. Yet, as demonstrated by next-generation sequencing experiments, there is a very real and pressing need for such methods.

In this paper, we develop methods for generating optimal block designs for category-valued single factor-experiments with responses drawn from a Poisson distribution, and with efficient estimation of contrasts of the model parameters being the objective. In Section 2 we develop the notation and definitions needed to specify the generalised linear mixed model (GLMM) for responses drawn from an exponential family of distributions and the corresponding pseudo-likelihood estimating equations. From these we derive the marginal Fisher information matrix for the estimation of the fixed effects in the model. In Section 3, we develop objective functions based on D_A- and $C-$optimality for the efficient estimation of the fixed effects parameters in a Poisson GLMM. While these optimality criteria generally result in objective functions which require matrix inversion, we show that for Poisson GLMMs with log link the objective functions can be simplified so that less inversion is necessary, leading to computational efficiency in the search for optimal designs. We use simulated annealing to search the set of competing designs for experiments of a given size, with the search space constrained to those designs which are locally optimal based on point prior estimates of the fixed effect parameters. Our key inputs into the simulated annealing algorithm are described in Section 4. In Section 5 we consider two examples, including a next-generation sequencing experiment, where we generate locally optimal block designs using the methods developed in Sections 3 and 4. These show that, for a fixed number of blocks with constant block size, the replication of treatments in Poisson GLMM-based optimal designs are inversely proportional to the relative magnitudes of the treatments' expected counts, which flies in the face of our traditional belief of optimality being achieved through the (near-) balanced

allocation of treatments across blocks. They further show that, for experiments with non-negligible effect sizes, the Poisson GLMM-based optimal designs may be substantially more efficient than optimal designs from the classical setting assuming additivity.

With these methods in hand, experimenters will be enabled in correctly answering their research questions with minimum effort, subjects, or other resources.

2 Models

In this section we introduce generalised linear mixed models as an extension of both generalised linear models and of linear mixed models. As we progress to derive expressions for design optimality criteria, we observe that features of efficient designs for both generalised linear models and of linear mixed models are present in efficient designs for generalised linear mixed models.

2.1 The generalised linear model

Consider an experiment in which t treatments are arranged in a completely randomised design comprising n experimental units, where n is a multiple of t. We define a linear model for an $n \times 1$ vector of observations as

$$\boldsymbol{y} = X\boldsymbol{\beta} + \boldsymbol{e}, \tag{1}$$

where $\boldsymbol{\beta} = (\alpha, \boldsymbol{\tau}^\mathsf{T})^\mathsf{T}$ is a $p \times 1$ vector of parameters containing the fixed effects α, denoting the overall mean of the observations, and $\boldsymbol{\tau} = (\tau_1, \ldots, \tau_t)^\mathsf{T}$, denoting the t treatment effect parameters. The $n \times p$ treatment design matrix, X, characterises the allocation of treatments to experimental units and, therefore, the fixed effect parameters associated with each observation in $\boldsymbol{y}$. The $n \times 1$ vector of residual errors, $\boldsymbol{e}$, is assumed to be independently and identically distributed normal with constant variance. If the response variable does not give rise to this error distribution, then an alternative model needs to be considered. One such alternative is the generalised linear model (GLM).

GLMs are used when the responses in $\boldsymbol{y}$ are assumed to arise from a distribution belonging to the exponential family of distributions. Such distributions include, for example, the binomial distribution for binary responses and the Poisson distribution for count responses. In a GLM the vector of mean responses, $\boldsymbol{\mu}$, and the linear predictors, $X\boldsymbol{\beta}$, are related by a canonical link function, $g(\cdot) = b'(\cdot)^{-1}$. This gives rise to the model form

$$\boldsymbol{\eta} = g(\boldsymbol{\mu}) = X\boldsymbol{\beta}. \tag{2}$$

It follows that $E(\boldsymbol{y}) = \boldsymbol{\mu} = b'(\boldsymbol{\theta})$ and $\mathrm{Var}(\boldsymbol{y}) = b''(\boldsymbol{\theta})a(\phi)$, where $b(\boldsymbol{\theta})$ and $a(\phi)$ denote functions of the natural parameter, $\boldsymbol{\theta}$, and the dispersion parameter, ϕ, respectively, for the independent observations in $\boldsymbol{y}$.

2.1.1 Fisher information matrix for GLMs

The maximum likelihood estimator, $\widehat{\boldsymbol{\beta}}$, of the fixed model effects, $\boldsymbol{\beta}$, in a GLM is asymptotically normally distributed. The covariance matrix of $\widehat{\boldsymbol{\beta}}$ is the inverse of the Fisher information matrix which is derived from the log-likelihood function of the GLM, i.e.

$$M(\xi, \boldsymbol{\beta}) = E\left\{ -\frac{\partial^2 \ell(\boldsymbol{\theta})}{\partial \boldsymbol{\beta}^2} \right\} = E\left\{ -\frac{\partial^2 \ell\left[\theta\big(g^{-1}(X\boldsymbol{\beta}; y, \phi)\big)\right]}{\partial \boldsymbol{\beta}^2} \right\} = X^\mathsf{T} W X,$$

where $W = (DVD)^{-1}$, $V = \mathrm{diag}[\mathrm{Var}(y_i)]$ and $D = \mathrm{diag}[\partial \eta_i / \partial \mu_i]$. Hence, the information matrix depends on the link function, since $\partial \eta_i / \partial \mu_i = g'(\mu_i)$, the design, ξ, through the design matrix X and the parameters in $\boldsymbol{\beta}$.

2.1.2 Poisson GLM

Responses in many modern experiments are drawn from distributions such as the one- and two-parameter exponential families, e.g., RNA sequence counts from a negative binomial distribution. Here we focus exclusively on experiments in which responses from the ith treatment group are counts independently drawn from the one-parameter Poisson distribution, i.e. $\boldsymbol{y}_i \sim \mathrm{Poisson}(\boldsymbol{\lambda}_i)$, with canonical link function $g(\cdot) = \log(\cdot)$.

2.2 The linear mixed model

Consider now an experiment in which t treatments are arranged in a generalised block design with b blocking factors. We define the linear mixed model (LMM) for an $n \times 1$ vector of observations, $\boldsymbol{y}$, using the general matrix notation

$$\boldsymbol{y} = X\boldsymbol{\beta} + Z\mathbf{u} + \boldsymbol{e}, \tag{3}$$

where the linear component, $\boldsymbol{\mu} = X\boldsymbol{\beta}$, represents the expected responses of the marginal model, with fixed effects parameter vector, $\boldsymbol{\beta}$, and treatment design matrix, X, defined as in (1). The vector of block random effect parameters $\boldsymbol{u} = (\boldsymbol{u}_1^{\mathsf{T}}, \ldots, \boldsymbol{u}_b^{\mathsf{T}})^{\mathsf{T}}$ is multivariate normally (MVN) distributed, with sub-vector $\boldsymbol{u}_i = (u_{i1}, \ldots, u_{ib_i})^{\mathsf{T}} \sim MVN(\mathbf{0}, G_i)$, where $G_i = \sigma_i^2 I$, corresponding to the ith block factor, $i = 1, \ldots, b$. The $n \times b$ block design matrix, Z, characterises the association of experimental units and, therefore, random effect parameters with each observation in $\boldsymbol{y}$. Finally, the $n \times 1$ vector of residual error parameters $\boldsymbol{e} \sim \text{MVN}(0, R)$. In the following, we consider only the case where these errors are uncorrelated, i.e. $R = \sigma^2 I$.

2.2.1 Fisher information matrix for LMMs

The LMM estimating (or normal) equations are given by

$$\begin{bmatrix} X^{\mathsf{T}} R^{-1} X & X^{\mathsf{T}} R^{-1} Z \\ Z^{\mathsf{T}} R^{-1} X & Z^{\mathsf{T}} R^{-1} Z + G^{-1} \end{bmatrix} \begin{bmatrix} \boldsymbol{\beta} \\ \boldsymbol{u} \end{bmatrix} = \begin{bmatrix} X^{\mathsf{T}} R^{-1} \boldsymbol{y} \\ Z^{\mathsf{T}} R^{-1} \boldsymbol{y} \end{bmatrix}. \tag{4}$$

Solving the estimating equations in (4) for the fixed model effects, $\boldsymbol{\beta}$, and a design ξ yields the estimate $M(\xi, \boldsymbol{\beta}, \sigma, \sigma_u)\widehat{\boldsymbol{\beta}} = XV^{-1}\mathbf{y}$, where the information matrix $M(\xi, \boldsymbol{\beta}, \sigma, \sigma_u) = X^{\mathsf{T}} V^{-1} X$ and the weight matrix $V = ZGZ^{\mathsf{T}} + R$.

2.3 The generalised linear mixed model

Extending either the LMM in (3) to allow the observed responses to arise from a distribution in the exponential family, with linear predictor defined as in (2), or the GLM defined in (2) to also include random effects, yields the generalised linear mixed model (GLMM)

$$\boldsymbol{\eta} = g(\boldsymbol{\mu}) = g[\text{E}(\boldsymbol{y}|\boldsymbol{q})] = X\boldsymbol{\beta} + Z\mathbf{q}, \tag{5}$$

where $\boldsymbol{y}|\boldsymbol{q}$ denotes the vector of responses, conditional on the random effects $\boldsymbol{q} = (\boldsymbol{u}^{\mathsf{T}}, \boldsymbol{e}^{\mathsf{T}})^{\mathsf{T}}$, arising from an exponential family of distributions. The random effect parameter vector, $\boldsymbol{u}$, and vector of residual error parameters, $\boldsymbol{e}$, are defined as in (3). As expected, when the link function, $g(\cdot)$, is the identity, the GLMM reduces to the ordinary LMM.

2.3.1 Fisher information matrix for GLMMs

The pseudo-likelihood estimating equations for the GLMM defined in (5) are

$$\begin{bmatrix} X^{\mathsf{T}} W X & X^{\mathsf{T}} W Z \\ Z^{\mathsf{T}} W X & Z^{\mathsf{T}} W Z + G^{-1} \end{bmatrix} \begin{bmatrix} \boldsymbol{\tau} \\ \boldsymbol{u} \end{bmatrix} = \begin{bmatrix} X^{\mathsf{T}} W \boldsymbol{y}^{\star} \\ Z^{\mathsf{T}} W \boldsymbol{y}^{\star} \end{bmatrix}, \tag{6}$$

where $W = (DV_\mu^{1/2} A V_\mu^{1/2} D)^{-1}$ and $\boldsymbol{y}^{\star} = \boldsymbol{\eta} + (\boldsymbol{y} - \boldsymbol{\mu})g'(\boldsymbol{\mu})$ is a pseudo-variable. In general, $D = \partial\boldsymbol{\mu}/\partial\boldsymbol{\eta}$, $V_\mu = \text{diag}(\sqrt{\partial^2 b(\boldsymbol{\theta})/\partial\boldsymbol{\theta}^2})$ and $A = \text{diag}(1/a(\phi))$, where ϕ is the scale parameter of the response distribution. The Fisher information matrix for the estimation of both the fixed and random effects is

$$M(\xi, \boldsymbol{\beta}, \sigma, \sigma_u) = \begin{bmatrix} X^{\mathsf{T}} W X & X^{\mathsf{T}} W Z \\ Z^{\mathsf{T}} W X & Z^{\mathsf{T}} W Z + G^{-1} \end{bmatrix} \tag{7}$$

(Stroup, 2012). While it may be tempting to use the conditional form of the information matrix, $X^{\mathsf{T}} W X$, this does not ensure that the random effects are estimable. Instead, we follow Niaparast and Schwabe (2013) and Waite and Woods (2015) and use the marginal information matrix for the estimation of the fixed effects. For a generalised block design ξ, we partition the design matrix defined in (7) into four sub-matrices, i.e.

$$M(\xi, \boldsymbol{\beta}, \sigma, \sigma_u) = \begin{bmatrix} M_{11}(\xi, \boldsymbol{\beta}, \sigma, \sigma_u) & M_{12}(\xi, \boldsymbol{\beta}, \sigma, \sigma_u) \\ M_{21}(\xi, \boldsymbol{\beta}, \sigma, \sigma_u) & M_{22}(\xi, \boldsymbol{\beta}, \sigma, \sigma_u) \end{bmatrix} = \begin{bmatrix} M_{11} & M_{12} \\ M_{21} & M_{22} \end{bmatrix},$$

where sub-matrix M_{11} contains the information pertaining to the fixed effects of interest, the efficiencies of which we would like to optimise, and M_{22} contains the information for the remaining effects. For Poisson

regression in blocks, M_{11} contains the information for the fixed effects in $\boldsymbol{\beta}$ and M_{22} contains the information for the random effects. Then, from results on the inverse of a partitioned matrix (Harville, 1997, p. 98), the marginal information matrix for the estimation of $\boldsymbol{\beta}$ is given by

$$M_{\boldsymbol{\beta}}^{\mathrm{marg}}(\xi, \boldsymbol{\beta}, \sigma, \sigma_u) = M_{11} - M_{12}(M_{22})^{-1}M_{21}. \tag{8}$$

In the next section, we derive (8) for Poisson regression with unstructured treatments in blocks.

3 Optimal block designs for correlated count data

In this section we develop objective functions for the efficient estimation of the fixed effects in a Poisson GLMM for block designs with unstructured treatments.

Consider an experiment in which t treatments are arranged in b blocks of equal size k. Assuming observations y_{ij} from unit j in block i are conditionally Poisson-distributed with expected value given by the rate parameter $\lambda_{R(i,j)}$, where $R(i,j) \in \{1, \ldots, t\}$ denotes the label for the treatment randomised to the (i,j)th unit, $i = 1, 2, \ldots, b$ and $j = 1, 2, \ldots, k$. The GLMM for this situation can be written as

$$\eta_{R(i,j)} = \alpha + \tau_{R(i,j)} + u_i + e_{ij}, \tag{9}$$

where $\eta_{R(i,j)}$ denotes the response on the linear predictor scale, α is the overall mean and $\tau_{R(i,j)}$ is the fixed effect of treatment $R(i,j)$. The block effects, u_i, are assumed to be random $N(0, \sigma_u^2)$ with $\mathrm{cov}(u_i, u_{i'}) = \sigma_u^2$ for $i = i'$ and zero otherwise. The residual errors, e_{ij}, associated with each unit are assumed $N(0, \sigma^2)$ and mutually uncorrelated.

The model specified in (9) satisfies the GLMM definition in (5), where $\boldsymbol{\eta} = [\eta_{R(i,j)}]$, X is the treatment design matrix, $\boldsymbol{\tau} = (\tau_1, \cdots, \tau_t)^{\mathsf{T}}$ is a vector fixed effect treatment parameters and the vector of random effect parameters $\boldsymbol{q} = (\boldsymbol{u}^{\mathsf{T}}, \boldsymbol{e}^{\mathsf{T}})^{\mathsf{T}} = (u_1, \cdots, u_b, e_{11}, \cdots, e_{bk})^{\mathsf{T}}$. Since all blocks are of equal size k, then the block design matrix $Z = (Z_b | I_n)$, where $Z_b = I_b \otimes \boldsymbol{j}_k$, I_b is an identity matrix of order b, $\boldsymbol{j}_k$ is a $k \times 1$ vector of ones, $n = bk$, and $\otimes$ denotes the Kronecker (or outer) product.

The Poisson GLMM with link function $g(\cdot) = \log(\cdot)$, as defined in (9), can be expressed as a Poisson–Log-normal mixture, since $y_{ij}|\lambda_{ij}v_{ij} \sim \mathrm{Poisson}(\lambda_{ij}v_{ij})$ and $\exp(v_{ij}) = e_{ij} \sim N(0, \sigma^2)$. It incorporates overdispersion through the residual parameters, e_{ij}, in a way that is consistent with how the random block effects are incorporated into the model. (See both Stroup (2012) and Nettleton (2014) for a detailed discussion of this approach). An alternative analogous model is the negative binomial model which can be expressed as a Poisson–Gamma mixture model which is often used, for example, in the analysis of next generation sequencing data. In the Poisson–Gamma mixture model, $y_{ij}|\lambda_{ij}v_{ij} \sim \mathrm{Poisson}(\lambda_{ij}v_{ij})$ where $v_{ij} \sim \Gamma(1/\phi, \phi)$ for scale parameter ϕ.

When searching for optimal designs, we need to specify a criterion which describes the relative amount of (usually) treatment information that is available from a design to achieve the objectives of the experiment. Many of the commonly used criteria are based on properties of the Fisher information matrix. Here we focus our attention on finding designs that estimate contrasts of the fixed treatment effects as efficiently as possible, while ensuring that the random effects remain estimable. We then derive these objective functions for the Poisson GLMM for experimental designs with an unstructured treatment factor and a single block factor.

3.1 Optimality Criteria

We consider two optimality criteria: D_A–optimality, or generalised D–optimality, and C–optimality for fixed effects. Our implementation of both of these criteria depends on properties of the partitioned Fisher information matrix in (7).

Atkinson et al. (2007) describe a D_A optimal design as the design that minimises the determinant $B^{\mathsf{T}}M(\xi, \boldsymbol{\beta}, \sigma, \sigma_b)^{-1}B$, where B is a set of linear contrasts of the model parameters. We define the D_A–optimal design, $\xi_{D_A}^*$, over a class of competing designs, $\mathfrak{X}$, as

$$\xi_{D_A}^* = \arg\min_{\xi \in \mathfrak{X}} \det\{B^{\mathsf{T}}M(\xi, \boldsymbol{\beta}, \sigma, \sigma_b)^{-1}B\},$$

where $\det(\cdot)$ denotes the determinant.

The C–optimality criterion is a modification of the A–optimality criterion. Atkinson et al. (2007) define an A–optimal design as the design that minimises the trace of the inverse of the Fisher information matrix over $\mathfrak{X}$. That is, the A–optimal design is the design, ξ_A^*, that is defined as

$$\xi_A^* = \arg\min_{\xi \in \mathfrak{X}} \mathrm{tr}\{M(\xi, \boldsymbol{\beta}, \sigma, \sigma_b)^{-1}\}.$$

Atkinson et al. (2007) then define the C–optimal design, ξ_C^*, as the design

$$\xi_C^* = \arg\min_{\xi \in \mathfrak{X}} \operatorname{tr}\{B^\mathsf{T} M(\xi, \boldsymbol{\beta}, \sigma, \sigma_b)^{-1} B\}.$$

We now derive the expression for the D_A– and C–optimality objective functions for the estimation of linear combinations of the treatment effects in the model in (9).

3.2 Objective Functions

Since contrasts of the fixed treatment effects are of interest, let the contrasts in B be linear combinations of the entries in $\boldsymbol{\beta}$. In particular, we would like to estimate a set of orthogonal contrasts that form a basis for the degrees of freedom for treatment effects, so that the marginal information matrix is given by

$$B^\mathsf{T} M_{\boldsymbol{\beta}}^{\mathrm{marg}}(\xi, \boldsymbol{\beta}, \sigma, \sigma_b)^{-1} B = B^\mathsf{T} \left\{ M_{11} - M_{12}(M_{22})^{-1} M_{21} \right\}^{-1} B \tag{10}$$

where $M_{11} = X^\mathsf{T} W X$, $M_{12} = M_{21}^\mathsf{T} = X^\mathsf{T} W Z$, $M_{22} = Z^\mathsf{T} W Z + G^{-1}$.

In the case of a Poisson GLMM with a single blocking factor, it follows from the pseudo-likelihood estimating equations in (6) that $D = \operatorname{diag}(\lambda_1^{-1}, \ldots, \lambda_t^{-1})$, $V_\lambda^{1/2} = \operatorname{diag}(\lambda_1^{1/2}, \ldots, \lambda_t^{1/2})$, $A = \operatorname{diag}[1/a(\phi)] = I_n$ and, hence, the weight matrix $W = \operatorname{diag}(\lambda_{R(i,j)})$. For a block design with b blocks of size k, the block design matrix $Z = (I_b \otimes \boldsymbol{j}_k | I_{bk})$ and the diagonal covariance matrix corresponding to the random effects assuming $\operatorname{cov}(u_i, e_{ij}) = 0$ for all i and j, is

$$\operatorname{var}(\boldsymbol{q}) = \operatorname{var}\begin{bmatrix} \mathbf{u} \\ \boldsymbol{e} \end{bmatrix} = G = \begin{bmatrix} \sigma_u^2 I_b & \mathbf{0} \\ \mathbf{0} & \sigma^2 I_{bk} \end{bmatrix}.$$

Substituting these matrix results into M_{22} gives

$$M_{22} = (I_b \otimes \boldsymbol{j}_k | I_{bk})^\mathsf{T} \operatorname{diag}(\lambda_{R(i,j)})(I_b \otimes \boldsymbol{j}_k | I_{bk}) + \begin{bmatrix} (1/\sigma_u^2) I_b & \mathbf{0} \\ \mathbf{0} & (1/\sigma^2) I_{bk} \end{bmatrix}$$

Applying the results on the inverse of a sum (Henderson and Searle, 1981) to M_{22}, i.e.

$$M_{22} = (G^{-1} + Z^\mathsf{T} W Z)^{-1} = G - G Z^\mathsf{T} (W^{-1} + Z G Z^\mathsf{T})^{-1} Z G,$$

and the fact that $W^{-1} + Z G Z^\mathsf{T}$ is block diagonal with the sub-matrix corresponding to the ith block given by

$$(W^{-1} + Z G Z^\mathsf{T})_i = \operatorname{diag}\left(\sigma^2 + \lambda_{R(i,j)}^{-1}\right) + \sigma_b^2 \boldsymbol{j}_k \boldsymbol{j}_k^\mathsf{T},$$

we obtain

$$(W^{-1} + Z G Z^\mathsf{T})_i^{-1} = \operatorname{diag}\left(\frac{1}{\sigma^2 + \lambda_{R(i,j)}^{-1}}\right) + \frac{\boldsymbol{\ell}_i \boldsymbol{\ell}_i^\mathsf{T}}{\sigma_b^2 \left\{1 + (\boldsymbol{\ell}_i^{1/2})^\mathsf{T} \boldsymbol{\ell}_i^{1/2}\right\}},$$

where

$$\boldsymbol{\ell}_i = \sigma_b^2 \left[\frac{1}{\sigma^2 + \lambda_{R(i,1)}^{-1}}, \cdots, \frac{1}{\sigma^2 + \lambda_{R(i,k)}^{-1}}\right].$$

It follows that $M_{\boldsymbol{\beta}}^{\mathrm{marg}}(\xi, \boldsymbol{\beta}, \sigma, \sigma_b)$ is block diagonal with the ith sub-matrix, $M_{\boldsymbol{\beta}}^{\mathrm{marg}}(\xi, \boldsymbol{\beta}, \sigma, \sigma_b)_i = X_i^\mathsf{T} \Omega_i X_i$, where X_i contains the rows if the design matrix corresponding to block i and

$$\Omega_i = \operatorname{diag}\left(\frac{1}{\sigma^2 + \lambda_{R(i,j)}^{-1}}\right) - \frac{\boldsymbol{\ell}_i \boldsymbol{\ell}_i^\mathsf{T}}{\sigma_b^2 \left\{1 + (\boldsymbol{\ell}_i^{1/2})^\mathsf{T} \boldsymbol{\ell}_i^{1/2}\right\}}.$$

Since B contains only contrasts of the fixed effects, the expression for $B^\mathsf{T} M(\xi, \boldsymbol{\beta}, \sigma, \sigma_b)^{-1} B$ can be expressed in terms of the marginal information matrix for the fixed effects. It follows that the objective function for the D_A–optimal design is given by

$$\xi_{D_A}^* = \arg\min_{\xi \in \mathfrak{X}} \det\left\{B^\mathsf{T} \left(\sum_{i=1}^{b} X_i^\mathsf{T} \Omega_i X_i\right)^{-1} B\right\},$$

6

while the objective function for the C–optimal design is

$$\xi_C^* = \arg\min_{\xi \in \mathfrak{X}} \operatorname{tr} \left\{ B^\mathsf{T} \left(\sum_{i=1}^{b} X_i^\mathsf{T} \Omega_i X_i \right)^{-1} B \right\}.$$

A full derivation is provided in the Supplementary Material. The following example illustrates the structure of these objective functions.

Example 1 Suppose that we wish to find the optimal arrangement of $t = 3$ treatments in $b = 2$ blocks of size $k = 3$, and will observe a count response that we wish to model by the Poisson GLMM $\eta_{ij} = \alpha + \tau_{R(i,j)} + u_i + e_{ij}$, where $y_{ij}|u_i, e_{ij} \sim \text{Poisson}(\exp(\eta_{ij}))$ for $i = 1, 2, 3$, and $j = 1, 2$.

The components of the Fisher information matrix, defined in (10), are $Z = (I_2 \otimes j_3, I_6)$, $W = \text{diag}(\lambda_{R(i,j)})$, and $G = \text{diag}(\sigma_u^2 j_2^\mathsf{T}, \sigma^2 j_6^\mathsf{T})$. It follows that

$$M_{22} = [I_2 \otimes j_3 | I_6]^\mathsf{T} \text{diag}(\lambda_{R(i)}) [I_2 \otimes j_3 | I_6] + \begin{bmatrix} (1/\sigma_u^2) I_2 & \mathbf{0} \\ \mathbf{0} & (1/\sigma^2) I_6 \end{bmatrix}$$

$$= \left[\begin{array}{cc|ccc|ccc} \lambda_{R(1,\cdot)} & 0 & \lambda_{R(1,1)} & \lambda_{R(1,2)} & \lambda_{R(1,3)} & 0 & 0 & 0 \\ 0 & \lambda_{R(2,\cdot)} & 0 & 0 & 0 & \lambda_{R(2,1)} & \lambda_{R(2,2)} & \lambda_{R(2,3)} \\ \hline \lambda_{R(1,1)} & 0 & \lambda_{R(1,1)} & 0 & 0 & 0 & 0 & 0 \\ \lambda_{R(1,2)} & 0 & 0 & \lambda_{R(1,2)} & 0 & 0 & 0 & 0 \\ \lambda_{R(1,3)} & 0 & 0 & 0 & \lambda_{R(1,3)} & 0 & 0 & 0 \\ \hline 0 & \lambda_{R(2,1)} & 0 & 0 & 0 & \lambda_{R(2,1)} & 0 & 0 \\ 0 & \lambda_{R(2,2)} & 0 & 0 & 0 & 0 & \lambda_{R(2,2)} & 0 \\ 0 & \lambda_{R(2,3)} & 0 & 0 & 0 & 0 & 0 & \lambda_{R(2,3)} \end{array} \right]$$

$$+ \begin{bmatrix} (1/\sigma_u^2) I_2 & \mathbf{0} \\ \mathbf{0} & (1/\sigma^2) I_6 \end{bmatrix}$$

where $\lambda_{R(i,\cdot)} = \sum_{j=1}^{k} \lambda_{R(i,j)}$ and

$$\ell_i = \left[\frac{\sigma_b^2}{\sigma^2 + \lambda_{R(i,1)}^{-1}}, \frac{\sigma_b^2}{\sigma^2 + \lambda_{R(i,2)}^{-1}}, \frac{\sigma_b^2}{\sigma^2 + \lambda_{R(i,3)}^{-1}} \right].$$

The ith sub-matrix of the marginal information matrix, corresponding to block i in the design, is given by

$$M_\beta^{\text{marg}}(\xi, \boldsymbol{\beta}, \sigma, \sigma_b)_i$$

$$= \text{diag} \left(\frac{1}{\sigma^2 + \lambda_{R(1,j)}^{-1}} \right) - \frac{\ell_1 \ell_1^\mathsf{T}}{\sigma_b^2 \left(1 + (\ell_1^{1/2})^\mathsf{T} \ell_1^{1/2} \right)}$$

$$= \begin{bmatrix} \frac{1}{\sigma^2 + \lambda_{R(1,1)}^{-1}} & 0 & 0 \\ 0 & \frac{1}{\sigma^2 + \lambda_{R(1,2)}^{-1}} & 0 \\ 0 & 0 & \frac{1}{\sigma^2 + \lambda_{R(1,3)}^{-1}} \end{bmatrix}$$

$$- \frac{1}{1 + \frac{\sigma_b^2}{\sigma^2 + \lambda_{R(1,1)}^{-1}} + \frac{\sigma_b^2}{\sigma^2 + \lambda_{R(1,2)}^{-1}} + \frac{\sigma_b^2}{\sigma^2 + \lambda_{R(1,3)}^{-1}}} \times$$

$$\begin{bmatrix} \frac{\sigma_b^2}{(\sigma^2 + \lambda_{R(1,1)}^{-1})^2} & \frac{\sigma_b^2}{(\sigma^2 + \lambda_{R(1,1)}^{-1})(\sigma^2 + \lambda_{R(1,2)}^{-1})} & \frac{\sigma_b^2}{(\sigma^2 + \lambda_{R(1,1)}^{-1})(\sigma^2 + \lambda_{R(1,3)}^{-1})} \\ \frac{\sigma_b^2}{(\sigma^2 + \lambda_{R(1,2)}^{-1})(\sigma^2 + \lambda_{R(1,1)}^{-1})} & \frac{\sigma_b^2}{(\sigma^2 + \lambda_{R(1,2)}^{-1})^2} & \frac{\sigma_b^2}{(\sigma^2 + \lambda_{R(1,2)}^{-1})(\sigma^2 + \lambda_{R(1,3)}^{-1})} \\ \frac{\sigma_b^2}{(\sigma^2 + \lambda_{R(1,3)}^{-1})(\sigma^2 + \lambda_{R(1,1)}^{-1})} & \frac{\sigma_b^2}{(\sigma^2 + \lambda_{R(1,3)}^{-1})(\sigma^2 + \lambda_{R(1,2)}^{-1})} & \frac{\sigma_b^2}{(\sigma^2 + \lambda_{R(1,3)}^{-1})^2} \end{bmatrix},$$

with the corresponding structure of the second block taking a similar form.

To investigate the optimal designs that are produced, Table 1 gives the D_A–optimal and C–optimal designs for a variety of treatment means and and values of σ_b^2, with $\sigma^2 = 0.25$. We

observe that as the size of the block variance increases relative to the treatment means, the optimal design becomes more balanced. For small block variances the effect of the different treatment variances becomes more dominant in determining the optimal design.

[Table 1 about here.]

4 Locally optimal block designs using simulated annealing

Generating D_A–optimal or C–optimal block designs requires an iterative search algorithm for which we have elected to use simulated annealing (SA) (Kirkpatrick et al., 1983). Since the design criterion to be minimised for both D_A– and C–optimality is the generalised variance of $B\boldsymbol{\beta}$, which for count data is functionally dependent on the treatment group means, we constrain the SA algorithm to search for locally optimal designs, i.e. designs that are optimal for a set of point priors for the expected treatment counts, λ_h, $h = 1, \ldots, t$, and the variance components between blocks, σ_u^2, and residuals, σ^2.

Three key inputs are required by the SA algorithm: a starting design, an objective function and candidate generator procedure, which we now discuss.

The SA algorithm is initialised with a starting design, D_0, generated by randomly assigning treatments to blocks with objective function value $O(D_0)$. At each iteration a new design, D_i, is generated by random exchanges of treatments in randomly selected experimental units in design D_{i-1} at the previous iteration, where $D_{i-1} = D_0$ at the first iteration. Since the SA algorithm searches for candidate designs which minimise the objective function, D_i always replaces D_{i-1} if $O(D_i) < O(D_{i-1})$, and has a small probability of replacing D_{i-1} even it is a slightly worse design. The acceptance probability of a worse design depends on the so-called temperature of the algorithm, which is initially set high to enable the algorithm to escape local optima in early iterations. As the iterations continue the temperature gradually cools, and with it the probability of accepting worse designs. In this way, the algorithm converges to the global optimum within the constrained set of competing designs.

As discussed in section 3.2, we consider two objective functions based on the D_A–optimality and C–optimality criteria, both defined in Atkinson et al. (2007). The goal is search the space of candidate designs that minimise these functions.

In contrast to block designs from the classical setting, where optimal efficiency is achieved by allocating treatments as equally as possible among blocks, optimal designs based on responses drawn from Poisson distributions have treatment replication inversely proportional to their treatment means. Our selection of a starting design, therefore, makes no assumption of equal replication or balance. Consequently, a candidate design generating procedure which makes random exchanges of treatments between blocks is unsatisfactory. Instead, we propose starting with a random design and then substitute the treatment assigned to a randomly chosen experimental unit in the design with a randomly chosen treatment from the treatment set.

Our preliminary testing of this strategy showed that, for some sets of design parameters (i.e. number of treatments, blocks and block size) and point priors, the SA algorithm converged very slowly and sometimes would get caught in local minima, even for reasonably high initial temperatures. To overcome these limitations, our candidate design generating procedure includes an option for $m \geqslant 1$ substitutions to be made at each iteration. A vector of probabilities $P = \{(p_1, \ldots, p_m) : p_1 \geq \cdots \geq p_m \text{ and } p_1 + \cdots + p_m = 1\}$, where p_m denotes the probability that m experimental units will have treatment substitutions at a given iteration. The SA algorithm for performing the optimisation described above is implemented in the `designGLMM R` package which is available from the Comprehensive R Archive Network.

We now consider two examples where we find optimal block designs for experiments in which responses are drawn from a Poisson distribution using our SA algorithm.

5 Examples

5.1 Differential striatal gene expression between two strains of mouse

We consider a comparative experiment to assess the level of differential striatal gene expression between two mice strains using the Illumina GAIIx next-generation sequencing (NGS) platform (Bottomly et al., 2011). cDNA, copied from amplified RNA isolated from cells in the striatum of twenty-one mice – ten from the C57BL/6J strain (strain 1) and eleven from the DBA/2J strain (strain 2) – was loaded into individual lanes (plots) of three flow cells (blocks) for sequencing. The design used by Bottomly et al. (2011) comprised two flow cells with three replicates of strain 1 and four replicates of strain 2, and a third flow cell with four replicates of strain 1 and three replicates of strain 2. The question that we wish to answer is whether this design, or a different design with the same number of samples, is optimal for the estimation of strain effects.

8

Table 2 presents counts from four of the 36536 identified genes (labelled A = ENSMUSG00000046994, B = ENSMUSG00000039967, C = ENSMUSG00000050141 and D = ENSMUSG00000033826) in this experiment, selected to represent the different effect sizes observed across the entire data set (see ReCount resource (Frazee et al., 2011)). A per-gene GLMM of the form presented in (9) was fitted to the count data yielding the parameter estimates shown in Table 3. We now use the effect sizes obtained from these estimates as point priors in searching for optimal designs for the estimation of the strain effect.

Table 3 shows that the size of the strain effect in genes A and B are quite small, with the relative abundances being approximately equal to 1. However, these genes do differ when we consider the ratio of the between flow cell variation to the within flow cell variation (i.e. σ_u^2/σ^2) of each. For gene A the between flow cell variation is 0.4 times that within cells, while for gene B this variance ratio is an order of magnitude larger. The C–optimal design based on the point priors estimated from the gene A data consists of three flow cells, each comprising four replicates of strain 1 and three replicates of strain 2, while the C–optimal design based on the point priors estimated from the gene B data consists of three flow cells, each comprising three replicates of strain 1 and four replicates of strain 2.

In contrast, the strain effect is very large for genes C and D with 2000 more copies of gene C in strain 1 than strain 2 and, conversely, almost 28 times the number of copies of gene D in strain 2 than strain 1. For both of these genes the magnitude of the variation between flow cells is comparable with that for genes A and B, however the within flow cell variation for genes C and D appears negligible. The C–optimal design based on the point priors estimated from the gene C data consists of three flow cells, each comprising one replicate of strain 1 and replicates of strain 2, while the C–optimal design based on the point priors estimated from the gene D data consists of three flow cells, each comprising five replicates of strain 1 and two replicates of strain 2. Neither of these designs, nor those identified as optimal for genes A and B, is the same as the design used by Bottomly et al. (2011).

For each of genes A – D we now consider the relative performances of eight alternative designs, $D_1 - D_8$ shown in Table 4, in which the twenty-one striatum cDNA samples are arranged in three blocks (flow cells), with seven samples per block. The treatments (strains) assigned to each flow cell are denoted by $1^{r_1}2^{r_2}$, where r_h denotes the number of replicates of strain h, $h = 1, 2$. Designs D_3, D_6, D_8 and D_2 are the C–optimal designs given above for genes A to D, respectively. Designs D_4, with two blocks containing 1^42^3 and one block containing 1^32^4, and D_5, with two blocks each containing 1^32^4 and one block containing 1^32^4, would be considered optimal and isomorphic under unit-treatment additivity.

Figure 1 shows the per-gene relative efficiencies of designs $D_1 - D_8$ using the point priors in Table 4, where here we define relative efficiency of design D_i as $O_g(D_i)/\max\{O_g(D_1),\ldots,O_g(D_8)\}$, i.e. the ratio of the value of the objective function, based on C–optimality, for design D_i relative to the largest value of the objective function across all eight designs, based on the point priors of gene g, $g = $ A, B, C, D. Figure 1 shows that designs D_4 and D_5 are optimal for genes A and B which each have a negligible strain effect. Note that because these design are near-balanced they are also D_A–optimal. These designs are not optimal, however, for genes C and D, where the strain effects are quite large. Indeed, the larger the strain effect, the more substantial the loss in efficiency.

[Table 2 about here.]

[Table 3 about here.]

[Table 4 about here.]

[Figure 1 about here.]

5.2 Begging behaviour of nestling barn owls

Roulin and Bersier (2007) investigated the begging behaviour of nestling barn owls. They recorded the number of begging vocalisations, or calls, made by an offspring to its parent in the 15 minutes prior to the parent owl's arrival at the nest. Of interest were the treatment factors gender (of the parent) and satiety (food-deprived and food-satiated juvenile). See Zuur et al. (2007) for a detailed discussion of these data. Suppose that in a future similar study the researchers want to investigate 15 barn owl broods (blocks) each comprising 10 nestlings, what would be the optimal design? Treating the four combinations of gender and satiety as four levels of a single treatment factor, we fitted a Poisson GLMM to the data given in (Roulin and Bersier, 2007) to obtain estimates of the requisite point priors. From this analysis we found that the mean number of calls for Deprived Females was $\lambda_1 = 1.33$, for Deprived Males was $\lambda_2 = 1.36$, for Satiated Females was $\lambda_3 = 0.44$ and for Satiated Males was $\lambda_4 = 0.54$. The between nest standard deviation was $\sigma_u = 1.11$ and the within nest excess variation was $\sigma = 0.47$. The C–optimal design for this experiment is the

one in which all 15 broods comprise the treatment allocation $1^3 2^3 3^2 4^2$ to nestlings. On the other hand, the classically optimal design would consist of two broods each with the treatment allocations $1^3 2^3 3^2 4^2$, $1^3 2^2 3^3 4^2$, $1^3 2^2 3^2 4^3$, $1^2 2^3 3^3 4^2$, $1^2 2^3 3^2 4^3$, and $1^2 2^2 3^3 4^3$, with the remaining three broods being selected from these six combinations such that the treatments are as balanced as possible ($1^3 2^3 3^2 4^2$, $1^2 2^2 3^3 4^3$, and $1^2 2^3 3^2 4^3$, for instance).

6 Discussion

In this paper, we find optimal designs for Poisson regression with a single unstructured treatment factor and a single variable that creates blocks of equal size. The methods discussed here are implemented in the `R` statistical software package `designGLMM`, which is available on the Comprehensive R Archive Network (cran.r-project.org) under a GPL3 license. We observe that for experiments where the treatment means are sufficiently different, and the block effect is not dominant, the optimal designs differ from those for linear models. This is because in a non–linear setting the Fisher information matrix, and hence the optimal designs, depend on the values of the model parameters.

The optimal designs for generalised linear mixed models depend on the functional form of the model and the values of the model parameters. In this paper, we have used a Poisson–Lognormal model, as discussed in Stroup (2012) and Nettleton (2014). Many other model configurations are available for modelling count data, most notably the negative binomial model (Lawless (1987)). Hilbe (2011) presents yet other possibilities, including alternate mean–variance relationships and hurdle models. Use of these alternate modelling approaches may yield optimal designs for the estimation of treatment effects which differ from those based on the Poisson GLMM.

The objective functions considered in this paper assume that the treatment effects are of primary interest. Specifically, a set of linear combinations of treatment effects are to be estimated with as small variance as possible, while being distinguishable from block effects. In some experiments, researchers are also interested in estimating the block effects efficiently, which may give rise to different optimal designs.

In Example 1 of Section 5, we considered the optimal design of a NGS experiment in which seven samples were placed onto individual lanes of three different flow cells. In generating optimal designs for this experiment, we considered only the variability between chips, and not between lanes. Auer and Doerge (2010) suggest that variation between lanes should also be a consideration. Furthermore, some NGS experiments use a process called *barcoding* to place multiple samples onto a single lane. This would suggest that more complex design structures, such as row–column designs, may be appropriate.

Additional complications arise from the design of NGS experiments. For instance, Auer and Doerge (2010, Eq.3) use an offset term, such as $\log(c_{ij})$, to normalise the number of reads per lane, and is common practise in the modelling of NGS data (see, for example Mortazavi et al. 2008).

We are currently looking at how we can address some of these issues in the optimal design of NGS experiments. Other areas which require further investigation include incorporating prior distributions for each of the model parameters to develop Bayesian optimal designs, and the investigation of alternative search algorithms that may be more efficient than simulated annealing in finding optimal designs.

Supplementary Materials

Web Appendix A, referenced in Section 3, is available with this paper at the Biometrics website on Wiley Online Library.

References

Atkinson, A., Donev, A., and Tobias, R. (2007). *Optimum experimental designs, with SAS.* Oxford, UK: Oxford Univiversity Press.

Auer, P. L. and Doerge, R. W. (2010). Statistical design and analysis of RNA sequencing data. *Genetics* **185,** 405–416.

Bottomly, D., Walter, N., Hunter, J., Darakjian, P., Kawane, S., Buck, K., Searles, R., Mooney, M., McWeeney, S., and Hitzemann, R. (2011). Evaluating gene expression in C57BL/6J and DBA/2J mouse striatum using RNA-seq and microarrays. *PloS One* **6,** e17820.

Cox, D. (1988). A note on design when response has an exponential family distribution. *Biometrika* **75,** 161–164.

Craig, D. W., Pearson, J. V., Szelinger, S., Sekar, A., Redman, M., Corneveaux, J. J., Pawlowski, T. L., Laub, T., Nunn, G., Stephan, D. A., et al. (2008). Identification of genetic variants using bar-coded multiplexed sequencing. *Nature Methods* **5,** 887–893.

Fisher, R. A. (1926). The arrangement of field experiments. *Journal of the Ministry of Agriculture of Great Britain* **33,** 503–513.

Frazee, A., Langmead, B., and Leek, J. (2011). Recount: a multi-experiment resource of analysis-ready RNA-seq gene count datasets. *BMC Bioinformatics* **12,** 449.

Haines, L. M. (1987). The application of the annealing algorithm to the construction of exact optimal designs for linear–regression models. *Technometrics* **29,** 439–447.

Harville, D. A. (1997). *Matrix algebra from a statistician's perspective.* New York: Springer-Verlag.

Henderson, H. V. and Searle, S. R. (1981). On deriving the inverse of a sum of matrices. *Siam Review* **23,** 53–60.

Hilbe, J. M. (2011). *Negative binomial regression.* New York: Cambridge University Press.

John, J. A. and Williams, E. R. (1995). *Cyclic and computer generated designs.* London: Chapman & Hall.

Khuri, A. I., Mukherjee, B., Sinha, B. K., and Ghosh, M. (2006). Design issues for generalized linear models: A review. *Statistical Science* **21,** 376–399.

Kirkpatrick, S. (1984). Optimization by simulated annealing: Quantitative studies. *Journal of statistical physics* **34,** 975–986.

Kirkpatrick, S., Gelatt, C., and Vecchi, M. (1983). Optimization by simulated annealing. *Science* **220,** 671–680.

Lawless, J. F. (1987). Negative binomial and mixed poisson regression. *Canadian Journal of Statistics* **15,** 209–225.

Nettleton, D. (2014). Design of RNA sequencing experiments. In Datta, S. and Nettleton, D., editors, *Statistical analysis of next generation sequencing data*, volume 4 of *Frontiers in Probability and the Statistical Sciences*, chapter 5, pages 93–119. Switzerland: Springer International Publishing.

Niaparast, M. and Schwabe, R. (2013). Optimal design for quasi-likelihood estimation in poisson regression with random coefficients. *Journal of Statistical Planning and Inference* **143,** 296–306.

Roulin, A. and Bersier, L. (2007). Nestling barn owls beg more intensely in the presence of their mother than in the presence of their father. *Animal Behaviour* **74,** 1099–1106.

Russell, K., Woods, D., Lewis, S., and Eccleston, J. (2009). D-optimal designs for poisson regression models. *Statistica Sinica* **19,** 721–730.

Schena, M., Shalon, D., Davis, R. W., and Brown, P. O. (1995). Quantitative monitoring of gene-expression patterns with a complementary-DNA microarray. *Science* **270,** 467–470.

Smith, K. (1918). On the standard deviations of adjusted and interpolated values of an observed polynomial function and its constants and the guidance they give towards a proper choice of the distribution of observations. *Biometrika* **12,** 185.

Stroup, W. W. (2012). *Generalized linear mixed models: modern concepts, methods and applications.* Boca Raton FL: CRC press.

Waite, T. W. and Woods, D. C. (2015). Designs for generalized linear models with random block effects via information matrix approximations. *Biometrika*, in press.

Woods, D. (2010). Robust designs for binary data: applications of simulated annealing. *Journal of Statistical Computation and Simulation* **80,** 29–41.

Woods, D., Lewis, S., Eccleston, J., and Russell, K. (2006). Designs for generalized linear models with several variables and model uncertainty. *Technometrics* **48,** 284–292.

Woods, D. C. and Van de Ven, P. (2011). Blocked designs for experiments with correlated non-normal response. *Technometrics* **53,** 173–182.

Yang, J. and Mandal, A. (2015). D-optimal factorial designs under generalized linear models. *Communications in Statistics - Simulation and Computation* **44,** 2264–2277.

Zuur, A., Ieno, E., Walker, N., Saveliev, A., and Smith, G. (2007). *Mixed effects models and extensions in ecology with R.* New York: Springer Science and Business Media.

A   Full derivation of $M_{\boldsymbol{\beta}}^{\mathbf{marg}}(\xi, \boldsymbol{\beta}, \sigma, \sigma_b)$

In this section, we present a full derivation for the marginal information matrix for the estimation of a set of contrasts of the fixed parameters $B\boldsymbol{\beta}$ for a Poisson GLMM. The marginal information matrix will is given by

$$B^T M_{\boldsymbol{\beta}}^{\mathrm{marg}}(\xi, \boldsymbol{\beta}, \sigma, \sigma_b)^{-1} B = B^T \left[M_{11} - M_{12} \{M_{22}\}^{-1} M_{21} \right]^{-1} B$$

where $M_{11} = X^\mathsf{T} W X$, $M_{12} = M_{21}^\mathsf{T} = X^\mathsf{T} W Z$, and $M_{22} = Z^\mathsf{T} W Z + G^{-1}$. In this formulation, $W = (D V_\lambda^{1/2} A V_\lambda^{1/2} D)^{-1}$, and G is the covariance matrix of random effects. For Poisson regression, we have

$$D = diag \left[\frac{\partial g(\boldsymbol{\lambda}|\boldsymbol{bu})}{\partial \boldsymbol{\lambda}} \right] = diag \left[\frac{\partial \log(\boldsymbol{\lambda})}{\partial \boldsymbol{\lambda}} \right] = diag[\lambda_1^{-1}, \lambda_2^{-1}, \dots, \lambda_t^{-1}],$$

$$V_\lambda^{1/2} = diag \left[\left(\frac{\partial^2 b(\theta)}{\partial \theta^2} \right)^{1/2} \right] = diag \left[\left(\frac{\partial^2 \exp(\boldsymbol{\eta})}{\partial \boldsymbol{\eta}^2} \right)^{1/2} \right] = diag[(\lambda_1, \lambda_2, \dots, \lambda_t)^{1/2}],$$

$$A = diag[1/a(\phi)] = I_N,$$

and hence $W = diag[\lambda_{R(i,j)}]$. If we assume that $\mathrm{cov}(u_i, e_{ij}) = 0$ for all i, j then

$$\mathrm{Var}(\boldsymbol{q}) = \mathrm{var} \left[\begin{array}{c} \boldsymbol{u} \\ \boldsymbol{e} \end{array} \right] = G = \left[\begin{array}{cc} \sigma_b^2 I_b & \mathbf{0} \\ \mathbf{0} & \sigma^2 I_{bk} \end{array} \right].$$

We can use this information to simplify the expression for $(M_{22})^{-1}$ so that it does not require matrix inversion. This will improve computation times for the simulated annealing algorithm.

If we consider a block design with b blocks of size k. We then have that the block design matrix, $Z = [I_b \otimes \boldsymbol{j}_k | I_{bk}]$, weight matrix $W = diag(\lambda_{R(i,j)})$ and diagonal covariance matrix corresponding to the random effects G. Then

$$M_{22} = Z^T W Z + G^{-1}$$

$$= [\boldsymbol{I}_b \otimes \boldsymbol{j}_k | \boldsymbol{I}_{bk}]^T diag(\lambda_{R(i,j)}) [\boldsymbol{I}_b \otimes \boldsymbol{j}_k | \boldsymbol{I}_{bk}] + \left[\begin{array}{cc} \frac{1}{\sigma_b^2} \boldsymbol{I}_b & \mathbf{0} \\ \mathbf{0} & \frac{1}{\sigma^2} \boldsymbol{I}_{bk} \end{array} \right]$$

Using the inverse sum of matrices result of Henderson and Searle (1981) that

$$(H + JKL)^{-1} = H^{-1} - H^{-1} J (K^{-1} + L H^{-1} J)^{-1} L H^{-1},$$

we obtain

$$(G^{-1} + Z^T W Z)^{-1} = G - G Z^T (W^{-1} + Z G Z^T)^{-1} Z G.$$

Now

$$W^{-1} + Z G Z^T = diag \left(\lambda_{R(i,j)}^{-1} \right) + [I_b \otimes \boldsymbol{j}_k | I_{bk}] \left[\begin{array}{cc} \sigma_b^2 I_b & \mathbf{0} \\ \mathbf{0} & \sigma^2 I_{bk} \end{array} \right] \left[\begin{array}{c} I_b \otimes \boldsymbol{j}_k^T \\ I_{bk} \end{array} \right]$$

$$= diag \left(\lambda_{R(i,j)}^{-1} \right) + \left[\sigma_b^2 I_b \otimes \boldsymbol{j}_k \boldsymbol{j}_k^T + \sigma^2 I_{bk} \right]$$

$$= diag \left(\sigma^2 + \lambda_{R(i,j)}^{-1} \right) + \sigma_b^2 \left[\begin{array}{cccc} \boldsymbol{j}_k \boldsymbol{j}_k^T & \mathbf{0} & \cdots & \mathbf{0} \\ \mathbf{0} & \boldsymbol{j}_k \boldsymbol{j}_k^T & \cdots & \mathbf{0} \\ \vdots & \vdots & \ddots & \vdots \\ \mathbf{0} & \mathbf{0} & \cdots & \boldsymbol{j}_k \boldsymbol{j}_k^T \end{array} \right]$$

Notice that this matrix is block diagonal, with b blocks of size $k \times k$ with similar structure. We can then express the $(i,i)^{\text{th}}$ block as

$$(W^{-1} + ZGZ^T)_i = diag\left(\sigma^2 + \lambda_{R(i,j)}^{-1}\right) + \sigma_b^2 \boldsymbol{j}_k \boldsymbol{j}_k^T$$

This is of the form $(H + \boldsymbol{a}\boldsymbol{b}^T)$, where H is invertable and square and $\boldsymbol{a}$ and $\boldsymbol{b}$ are column vectors, so we can invert this block using the Sherman-Morrison formula. Then

$$(W^{-1} + ZGZ^T)_i^{-1} = diag\left(\frac{1}{\sigma^2 + \lambda_{R(i,j)}^{-1}}\right) - \frac{diag\left(\frac{1}{\sigma^2 + \lambda_{R(i,j)}^{-1}}\right) \times \sigma_b^2 \boldsymbol{j}_k \boldsymbol{j}_k^T \times diag\left(\frac{1}{\sigma^2 + \lambda_{R(i,j)}^{-1}}\right)}{1 + \sigma_b^2 \boldsymbol{j}_k^T diag\left(\frac{1}{\sigma^2 + \lambda_{R(i,j)}^{-1}}\right) \boldsymbol{j}_k}$$

If we let

$$\boldsymbol{\ell}_i = \left[\frac{\sigma_b^2}{\sigma^2 + \lambda_{R(i,1)}^{-1}}, \frac{\sigma_b^2}{\sigma^2 + \lambda_{R(i,2)}^{-1}}, \cdots \frac{\sigma_b^2}{\sigma^2 + \lambda_{R(i,k)}^{-1}}\right]$$

then

$$(W^{-1} + ZGZ^T)_i^{-1} = diag\left(\frac{1}{\sigma^2 + \lambda_{R(i,j)}^{-1}}\right) + \frac{\boldsymbol{\ell}_i \boldsymbol{\ell}_i^T}{\sigma_b^2(1 + (\boldsymbol{\ell}_i^{1/2})^T \boldsymbol{\ell}_i^{1/2})}$$

Next, we can add the additional components that are not dependent on X. So

$$W - WZ(ZWZ^T + G^{-1})^{-1}Z^TW = W - WZ(G - GZ^T(W^{-1} + ZGZ^T)^{-1}ZG)Z^TW$$
$$= W - WZGZ^TW + WZGZ^T(W^{-1} + ZGZ^T)^{-1}(WZGZ^T)^T$$

Since $ZGZ^T = [\sigma_b^2 I_b \otimes \boldsymbol{j}_i \boldsymbol{j}_i^T + \sigma^2 I_{bk}]$, we have

$$WZGZ^TW = \begin{bmatrix} \sigma_b^2 \boldsymbol{\lambda}_1 \boldsymbol{\lambda}_1^T & \boldsymbol{0} & \cdots & \boldsymbol{0} \\ \boldsymbol{0} & \sigma_b^2 \boldsymbol{\lambda}_2 \boldsymbol{\lambda}_2^T & \cdots & \boldsymbol{0} \\ \vdots & \vdots & \ddots & \vdots \\ \boldsymbol{0} & \boldsymbol{0} & \cdots & \sigma_b^2 \boldsymbol{\lambda}_b \boldsymbol{\lambda}_b^T \end{bmatrix} \times \sigma^2 W^2$$

$$WZGZ^T = \begin{bmatrix} \sigma_b^2 \boldsymbol{\lambda}_1 \boldsymbol{j}_k^T & \boldsymbol{0} & \cdots & \boldsymbol{0} \\ \boldsymbol{0} & \sigma_b^2 \boldsymbol{\lambda}_2 \boldsymbol{j}_k^T & \cdots & \boldsymbol{0} \\ \vdots & \vdots & \ddots & \vdots \\ \boldsymbol{0} & \boldsymbol{0} & \cdots & \sigma_b^2 \boldsymbol{\lambda}_b \boldsymbol{j}_k^T \end{bmatrix} \times \sigma^2 W$$

where $\boldsymbol{\lambda}_i = (\lambda_{R(i,1)}, \lambda_{R(i,2)}, \cdots, \lambda_{R(i,k)})^T$. Since each of these matrices are block diagonal, the $(i,i)^{\text{th}}$ block

of $W - WZ(ZWZ^T + G^{-1})^{-1}Z^T W$ becomes

$$(W - WZ(ZWZ^T + G^{-1})^{-1}Z^T W)_i$$
$$= W - WZGZ^T W + WZGZ^T \times (W^{-1} + ZGZ^T)^{-1} \times (WZGZ^T)^T$$
$$= diag(\lambda_{R(i,j)}) - (\sigma_b^2 \boldsymbol{\lambda}_i \boldsymbol{\lambda}_i^T + \sigma^2 diag(\lambda_{R(i,j)})^2)$$
$$+ (\sigma_b^2 \boldsymbol{\lambda}_i \boldsymbol{j}_k^T + \sigma^2 diag(\lambda_{R(i,j)})) \times \left(diag\left(\frac{1}{\sigma^2 + \lambda_{R(i,j)}^{-1}} \right) - \frac{\boldsymbol{\ell}_i \boldsymbol{\ell}_i^T}{\sigma_b^2(1 + (\boldsymbol{\ell}_i^{1/2})^T \boldsymbol{\ell}_i^{1/2})} \right) \times (\sigma_b^2 \boldsymbol{\lambda}_i \boldsymbol{j}_k^T + \sigma^2 diag(\lambda_{R(i,j)}))^T$$
$$= diag(\lambda_{R(i,j)} - \sigma^2 \lambda_{R(i,j)}^2) - \sigma_b^2 \boldsymbol{\lambda}_i \boldsymbol{\lambda}_i^T$$
$$+ \left(\sigma_b^2 \boldsymbol{\lambda}_i \boldsymbol{j}_k^T \times diag\left(\frac{1}{\sigma^2 + \lambda_{R(i,j)}^{-1}} \right) - \sigma_b^2 \boldsymbol{\lambda}_i \boldsymbol{j}_k^T \times \frac{\boldsymbol{\ell}_i \boldsymbol{\ell}_i^T}{\sigma_b^2(1 + (\boldsymbol{\ell}_i^{1/2})^T \boldsymbol{\ell}_i^{1/2})} + \sigma^2 diag(\lambda_{R(i,j)}) \times diag\left(\frac{1}{\sigma^2 + \lambda_{R(i,j)}^{-1}} \right) \right.$$
$$\left. - \sigma^2 diag(\lambda_{R(i,j)}) \times \frac{\boldsymbol{\ell}_i \boldsymbol{\ell}_i^T}{\sigma_b^2(1 + (\boldsymbol{\ell}_i^{1/2})^T \boldsymbol{\ell}_i^{1/2})} \right) \times (\sigma_b^2 \boldsymbol{\lambda}_i \boldsymbol{j}_k^T + \sigma^2 diag(\lambda_{R(i,j)}))^T$$
$$= diag(\lambda_{R(i,j)} - \sigma^2 \lambda_{R(i,j)}^2) - \sigma_b^2 \boldsymbol{\lambda}_i \boldsymbol{\lambda}_i^T$$
$$+ \left(\boldsymbol{\lambda}_i \boldsymbol{\ell}_i^T - \frac{\boldsymbol{\lambda}_i \boldsymbol{j}_k^T \boldsymbol{\ell}_i \boldsymbol{\ell}_i^T}{(1 + (\boldsymbol{\ell}_i^{1/2})^T \boldsymbol{\ell}_i^{1/2})} + diag\left(\frac{\sigma^2 \lambda_{R(i,j)}}{\sigma^2 + \lambda_{R(i,j)}^{-1}} \right) - \frac{\sigma^2 \boldsymbol{m}_i \boldsymbol{\ell}_i^T}{\sigma_b^2(1 + (\boldsymbol{\ell}_i^{1/2})^T \boldsymbol{\ell}_i^{1/2})} \right) \times (\sigma_b^2 \boldsymbol{\lambda}_i \boldsymbol{j}_k^T + \sigma^2 diag(\lambda_{R(i,j)}))^T$$
$$= diag(\lambda_{R(i,j)} - \sigma^2 \lambda_{R(i,j)}^2) - \sigma_b^2 \boldsymbol{\lambda}_i \boldsymbol{\lambda}_i^T + \sigma_b^2 \boldsymbol{\lambda}_i \boldsymbol{\ell}_i^T \boldsymbol{j}_k \boldsymbol{\lambda}_i^T - \sigma_b^2 \frac{\boldsymbol{\lambda}_i \boldsymbol{j}_k^T \boldsymbol{\ell}_i \boldsymbol{\ell}_i^T \boldsymbol{j}_k \boldsymbol{\lambda}_i^T}{(1 + (\boldsymbol{\ell}_i^{1/2})^T \boldsymbol{\ell}_i^{1/2})} + \sigma^2 \boldsymbol{m}_i \boldsymbol{\lambda}_i^T$$
$$- \frac{\sigma^2 \boldsymbol{m}_i \boldsymbol{\ell}_i^T \boldsymbol{j}_k \boldsymbol{\lambda}_i^T}{(1 + (\boldsymbol{\ell}_i^{1/2})^T \boldsymbol{\ell}_i^{1/2})} + \sigma^2 \boldsymbol{\lambda}_i \boldsymbol{m}_i^T - \frac{\sigma^2 \boldsymbol{\lambda}_i \boldsymbol{j}_k^T \boldsymbol{\ell}_i \boldsymbol{m}_i^T}{(1 + (\boldsymbol{\ell}_i^{1/2})^T \boldsymbol{\ell}_i^{1/2})} + diag\left(\frac{\sigma^4 \lambda_{R(i,j)}^2}{\sigma^2 + \lambda_{R(i,j)}^{-1}} \right) - \frac{\sigma^4 \boldsymbol{m}_i \boldsymbol{m}_i^T}{\sigma_b^2(1 + (\boldsymbol{\ell}_i^{1/2})^T \boldsymbol{\ell}_i^{1/2})}$$

where

$$\boldsymbol{m}_i = \left(\frac{\sigma_b^2 \lambda_{R(i,1)}}{\sigma^2 + \lambda_{R(i,1)}^{-1}}, \frac{\sigma_b^2 \lambda_{R(i,2)}}{\sigma^2 + \lambda_{R(i,2)}^{-1}}, \cdots \frac{\sigma_b^2 \lambda_{R(i,k)}}{\sigma^2 + \lambda_{R(i,k)}^{-1}} \right)^T$$

Since $\boldsymbol{\ell}_i^T \boldsymbol{j}_k = (\boldsymbol{\ell}_i^{1/2})^T \boldsymbol{\ell}_i^{1/2}$ and $\boldsymbol{j}_k^T \boldsymbol{\ell}_i = (\boldsymbol{\ell}_i^{1/2})^T \boldsymbol{\ell}_i^{1/2}$, which are constants, we obtain

$$(W - WZ(ZWZ^T + G^{-1})^{-1}Z^T W)_i$$
$$= diag(\lambda_{R(i,j)} - \sigma^2 \lambda_{R(i,j)}^2) - \sigma_b^2 \boldsymbol{\lambda}_i \boldsymbol{\lambda}_i^T + \sigma_b^2 \boldsymbol{\lambda}_i (\boldsymbol{\ell}_i^{1/2})^T \boldsymbol{\ell}_i^{1/2} \boldsymbol{\lambda}_i^T - \sigma_b^2 \frac{\boldsymbol{\lambda}_i (\boldsymbol{\ell}_i^{1/2})^T \boldsymbol{\ell}_i^{1/2} (\boldsymbol{\ell}_i^{1/2})^T \boldsymbol{\ell}_i^{1/2} \boldsymbol{\lambda}_i^T}{(1 + (\boldsymbol{\ell}_i^{1/2})^T \boldsymbol{\ell}_i^{1/2})} + \sigma^2 \boldsymbol{m}_i \boldsymbol{\lambda}_i^T$$
$$- \frac{\sigma^2 \boldsymbol{m}_i (\boldsymbol{\ell}_i^{1/2})^T \boldsymbol{\ell}_i^{1/2} \boldsymbol{\lambda}_i^T}{(1 + (\boldsymbol{\ell}_i^{1/2})^T \boldsymbol{\ell}_i^{1/2})} + \sigma^2 \boldsymbol{\lambda}_i \boldsymbol{m}_i^T - \frac{\sigma^2 \boldsymbol{\lambda}_i (\boldsymbol{\ell}_i^{1/2})^T \boldsymbol{\ell}_i^{1/2} \boldsymbol{m}_i^T}{(1 + (\boldsymbol{\ell}_i^{1/2})^T \boldsymbol{\ell}_i^{1/2})} + diag\left(\frac{\sigma^4 \lambda_{R(i,j)}^2}{\sigma^2 + \lambda_{R(i,j)}^{-1}} \right) - \frac{\sigma^4 \boldsymbol{m}_i \boldsymbol{m}_i^T}{\sigma_b^2(1 + (\boldsymbol{\ell}_i^{1/2})^T \boldsymbol{\ell}_i^{1/2})}$$
$$= diag\left(\lambda_{R(i,j)} - \sigma^2 \lambda_{R(i,j)}^2 + \frac{\sigma^4 \lambda_{R(i,j)}^2}{\sigma^2 + \lambda_{R(i,j)}^{-1}} \right) + \left(\sigma_b^2 (\boldsymbol{\ell}_i^{1/2})^T \boldsymbol{\ell}_i^{1/2} - \sigma_b^2 - \frac{\sigma_b^2 \boldsymbol{\ell}_i^T \boldsymbol{\ell}_i}{(1 + (\boldsymbol{\ell}_i^{1/2})^T \boldsymbol{\ell}_i^{1/2})} \right) \boldsymbol{\lambda}_i \boldsymbol{\lambda}_i^T$$
$$+ \left(\sigma^2 - \frac{\sigma^2 (\boldsymbol{\ell}_i^{1/2})^T \boldsymbol{\ell}_i^{1/2}}{(1 + (\boldsymbol{\ell}_i^{1/2})^T \boldsymbol{\ell}_i^{1/2})} \right) \boldsymbol{m}_i \boldsymbol{\lambda}_i^T + \left(\sigma^2 - \frac{\sigma^2 (\boldsymbol{\ell}_i^{1/2})^T \boldsymbol{\ell}_i^{1/2}}{(1 + (\boldsymbol{\ell}_i^{1/2})^T \boldsymbol{\ell}_i^{1/2})} \right) \boldsymbol{\lambda}_i \boldsymbol{m}_i^T - \frac{\sigma^4 \boldsymbol{m}_i \boldsymbol{m}_i^T}{\sigma_b^2(1 + (\boldsymbol{\ell}_i^{1/2})^T \boldsymbol{\ell}_i^{1/2})}$$
$$= diag\left(\frac{1}{\sigma^2 + \lambda_{R(i,j)}^{-1}} \right) - \left(\frac{\sigma_b^2}{(1 + (\boldsymbol{\ell}_i^{1/2})^T \boldsymbol{\ell}_i^{1/2})} \right) \boldsymbol{\lambda}_i \boldsymbol{\lambda}_i^T + \left(\frac{\sigma^2}{(1 + (\boldsymbol{\ell}_i^{1/2})^T \boldsymbol{\ell}_i^{1/2})} \right) \boldsymbol{m}_i \boldsymbol{\lambda}_i^T + \left(\frac{\sigma^2}{(1 + (\boldsymbol{\ell}_i^{1/2})^T \boldsymbol{\ell}_i^{1/2})} \right) \boldsymbol{\lambda}_i \boldsymbol{m}_i^T$$
$$- \frac{\sigma^4 \boldsymbol{m}_i \boldsymbol{m}_i^T}{\sigma_b^2(1 + (\boldsymbol{\ell}_i^{1/2})^T \boldsymbol{\ell}_i^{1/2})}$$
$$= diag\left(\frac{1}{\sigma^2 + \lambda_{R(i,j)}^{-1}} \right) + \frac{1}{\sigma_b^2(1 + (\boldsymbol{\ell}_i^{1/2})^T \boldsymbol{\ell}_i^{1/2})} \left(-\sigma_b^4 \boldsymbol{\lambda}_i \boldsymbol{\lambda}_i^T + \sigma^2 \sigma_b^2 \boldsymbol{m}_i \boldsymbol{\lambda}_i^T + \sigma^2 \sigma_b^2 \boldsymbol{\lambda}_i \boldsymbol{m}_i^T - \sigma^4 \boldsymbol{m}_i \boldsymbol{m}_i^T \right)$$
$$= diag\left(\frac{1}{\sigma^2 + \lambda_{R(i,j)}^{-1}} \right) - \frac{1}{\sigma_b^2(1 + (\boldsymbol{\ell}_i^{1/2})^T \boldsymbol{\ell}_i^{1/2})} \left(\sigma^2 \boldsymbol{m}_i - \sigma_b^2 \boldsymbol{\lambda} \right) \left(\sigma^2 \boldsymbol{m}_i - \sigma_b^2 \boldsymbol{\lambda} \right)^T$$

Since

$$\sigma^2 \boldsymbol{m}_i - \sigma_b^2 \boldsymbol{\lambda} = \left(\frac{\sigma^2 \sigma_b^2 \lambda_{R(i,1)}}{\sigma^2 + \lambda_{R(i,1)}^{-1}}, \frac{\sigma^2 \sigma_b^2 \lambda_{R(i,2)}}{\sigma^2 + \lambda_{R(i,2)}^{-1}}, \cdots \frac{\sigma^2 \sigma_b^2 \lambda_{R(i,k)}}{\sigma^2 + \lambda_{R(i,k)}^{-1}} \right)^T - \left(\sigma_b^2 \lambda_{R(i,1)}, \sigma_b^2 \lambda_{R(i,2)}, \cdots, \sigma_b^2 \lambda_{R(i,k)} \right)^T$$

$$= \left(\frac{\sigma^2 \sigma_b^2 \lambda_{R(i,1)} - \sigma^2 \sigma_b^2 \lambda_{R(i,1)} - \sigma_b^2}{\sigma^2 + \lambda_{R(i,1)}^{-1}}, \frac{\sigma^2 \sigma_b^2 \lambda_{R(i,2)} - \sigma^2 \sigma_b^2 \lambda_{R(i,2)} - \sigma_b^2}{\sigma^2 + \lambda_{R(i,2)}^{-1}}, \cdots \frac{\sigma^2 \sigma_b^2 \lambda_{R(i,k)} - \sigma^2 \sigma_b^2 \lambda_{R(i,k)} - \sigma_b^2}{\sigma^2 + \lambda_{R(i,k)}^{-1}} \right)^T$$

$$= - \left(\frac{\sigma_b^2}{\sigma^2 + \lambda_{R(i,1)}^{-1}}, \frac{\sigma_b^2}{\sigma^2 + \lambda_{R(i,2)}^{-1}}, \cdots \frac{\sigma_b^2}{\sigma^2 + \lambda_{R(i,k)}^{-1}} \right)^T$$

$$= -\boldsymbol{\ell}_i,$$

we can simplify to obtain

$$(W - WZ(ZWZ^T + G^{-1})^{-1} Z^T W)_i = diag \left(\frac{1}{\sigma^2 + \lambda_{R(i,j)}^{-1}} \right) - \frac{\boldsymbol{\ell}_i \boldsymbol{\ell}_i^T}{\sigma_b^2 (1 + (\boldsymbol{\ell}_i^{1/2})^T \boldsymbol{\ell}_i^{1/2})}$$

It follows that $M_{\boldsymbol{\beta}}^{\text{marg}}(\xi, \boldsymbol{\beta}, \sigma, \sigma_b)$ is block diagonal with the (i,i) block, $M_{\boldsymbol{\beta}}^{\text{marg}}(\xi, \boldsymbol{\beta}, \sigma, \sigma_b)_i$, given by

$$M_{\boldsymbol{\beta}}^{\text{marg}}(\xi, \boldsymbol{\beta}, \sigma, \sigma_b)_i = X_i^T \left(diag \left(\frac{1}{\sigma^2 + \lambda_{R(i,j)}^{-1}} \right) - \frac{\boldsymbol{\ell}_i \boldsymbol{\ell}_i^T}{\sigma_b^2 (1 + (\boldsymbol{\ell}_i^{1/2})^T \boldsymbol{\ell}_i^{1/2})} \right) X_i,$$

where X_i contains the rows if the design matrix corresponding to block i.
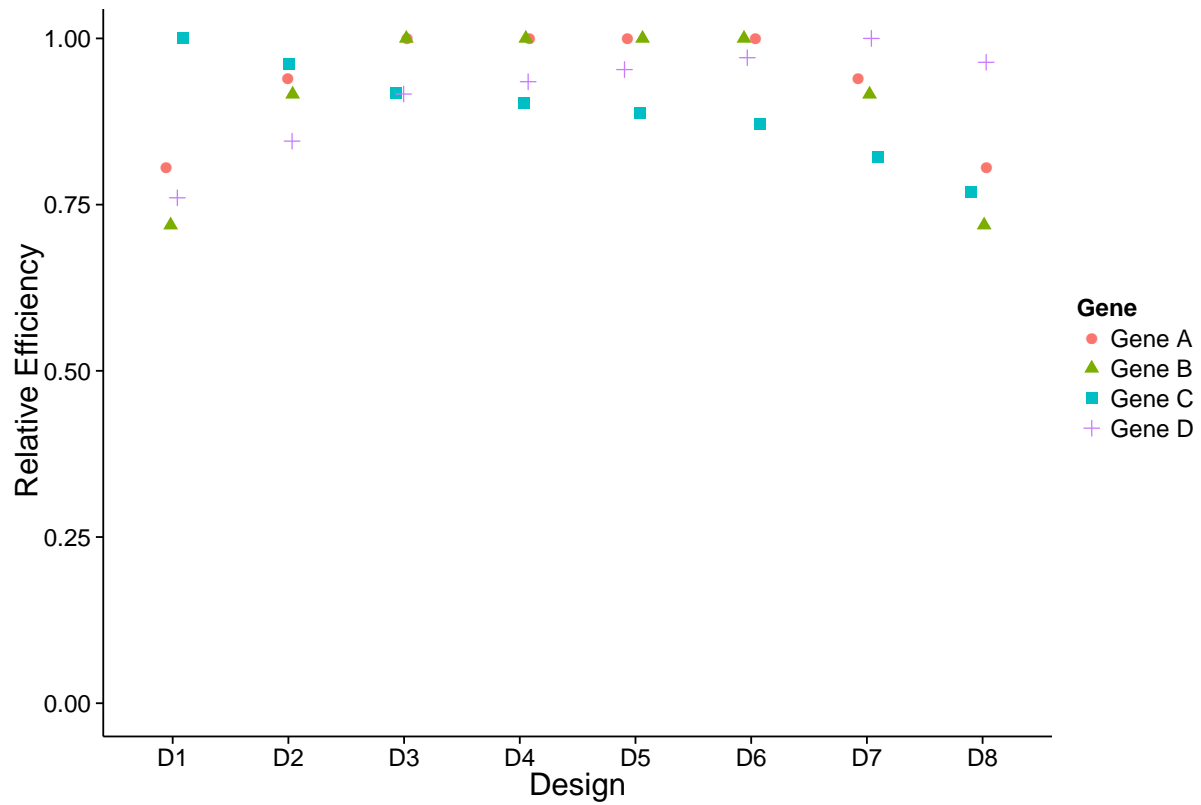
Figure 1: Per-gene relative efficiencies of eight designs, $D_1 - D_8$, defined in Table 4 based on point priors of four genes, A – D, given in Table 3.

Table 1: D_A- and C-optimal designs for three treatments with expected counts $(\lambda_1, \lambda_2, \lambda_3) = (1,1,1)$, $(1,1,2)$, $(1,2,4)$ or $(1,4,16)$, block variance $\sigma_b^2 = 0.016$, 0.25, or 4 and $\sigma^2 = 0.25$. The relative efficiencies of the randomised complete block design are given in the final column

| λ_1 | λ_2 | λ_3 | σ_b^2 | σ^2 | D_A–optimality | | C–optimality | |
					Optimal design	BIBD Efficiency	Optimal design	BIBD Efficiency
1	1	1	0.016	0.25	(1,2,3), (1,2,3)	1	(1,2,3), (1,2,3)	1
1	1	2	0.016	0.25	(1,2,3), (1,2,3)	1	(1,1,2), (1,2,3)	0.988
1	2	4	0.016	0.25	(1,2,3), (1,2,3)	1	(1,1,2), (1,2,3)	0.919
1	4	16	0.016	0.25	(1,2,3), (1,2,3)	1	(1,1,2), (1,2,3)	0.851
1	1	1	0.25	0.25	(1,2,3), (1,2,3)	1	(1,2,3), (1,2,3)	1
1	1	2	0.25	0.25	(1,2,3), (1,2,3)	1	(1,2,3), (1,2,3)	1
1	2	4	0.25	0.25	(1,2,3), (1,2,3)	1	(1,1,2), (1,2,3)	0.990
1	4	16	0.25	0.25	(1,2,3), (1,2,3)	1	(1,1,2), (1,2,3)	0.923
1	1	1	4	0.25	(1,2,3), (1,2,3)	1	(1,2,3), (1,2,3)	1
1	1	2	4	0.25	(1,2,3), (1,2,3)	1	(1,2,3), (1,2,3)	1
1	2	4	4	0.25	(1,2,3), (1,2,3)	1	(1,2,3), (1,2,3)	1
1	4	16	4	0.25	(1,2,3), (1,2,3)	1	(1,2,3), (1,2,3)	1

Table 2: Gene counts for four selected genes from two strains of mice: C57BL/6J (1) and DBA/2J (2). The columns within each subtable correspond to the seven lanes into which individual cDNA samples were loaded within a flow cell.

Flow cell	Gene†	Strain						
		1	1	1	2	2	2	2
1	A	132	134	140	112	134	100	115
1	B	794	922	606	507	688	510	659
1	C	34	59	52	1	0	0	1
1	D	10	12	7	9	38	29	19

Flow cell	Gene†	Strain						
		1	1	1	1	2	2	2
2	A	101	68	64	102	132	139	110
2	B	758	722	731	803	1080	614	961
2	C	43	29	30	31	0	0	1
2	D	41	1	12	3	2	33	61

Flow cell	Gene†	Strain						
		1	1	1	2	2	2	2
3	A	174	194	194	146	155	157	128
3	B	1169	1353	1343	1359	1437	1426	1512
3	C	64	41	56	6	1	1	1
3	D	18	5	5	35	50	31	45

†Genes: A = ENSMUSG00000046994; B = ENSMUSG00000039967
C = ENSMUSG00000050141; D = ENSMUSG00000033826

Table 3: Parameter estimates on the link scale $(\alpha, \tau_1, \sigma, \sigma_u)$ and the response scale (λ_1, λ_2) from fitting a per-gene Poisson GLMM to gene counts.

| Gene | Link scale | | | | Response scale | |
	α	τ_1	σ	σ_u	λ_1	λ_2
A	4.85767	0.00050	0.20104	0.12874	128.66	128.79
B	6.81209	-0.00001	0.13382	0.27905	908.77	908.76
C	3.78631	-3.73949	0.00000	0.19885	1855.30	1.05
D	1.87168	1.66639	0.00002	0.26546	1.23	34.40

Table 4: Eight block designs for $t = 2$ strains arranged in $b = 3$ blocks of size $k = 7$. Treatments within a block are denoted by $1^{r_1}2^{r_2}$, where r_h denotes the number of replicates of strain h, $h = 1, 2$. The blocks in designs D_4 and D_5 each have two different combinations of replicates of strains 1 and 2, indicated by the multiplier c in $c \times 1^{r_1}2^{r_2}$, $c = 1, 2$.

				Design				
D_1	D_2	D_3	D_4	D_5	D_6	D_7	D_8
$1^6 2^1$	$1^5 2^2$	$1^4 2^3$	$2 \times 1^4 2^3$	$2 \times 1^3 2^4$	$1^3 2^4$	$1^2 2^5$	$1^1 2^6$
			$1 \times 1^3 2^4$	$1 \times 1^4 2^3$			