D-Optimal Designs for the Mitscherlich Non-Linear Regression Function

ISSN 1066-5307, Mathematical Methods of Statistics, 2022, Vol. 31, No. 1, pp. 1–17. © Allerton Press, Inc., 2022.

D-Optimal Designs for the Mitscherlich Non-Linear Regression Function

Maliheh Heidari[1], **Md Abu Manju**[1], **Pieta C. IJzerman-Boon**[2], and **Edwin R. van den Heuvel**[1*]

[1]*Department of Mathematics and Computer Science, Eindhoven University of Technology, 5600 MB Eindhoven, The Netherlands*

[2]*Center for Mathematical Sciences, MSD, 5342 CC Oss, The Netherlands*

Received May 30, 2021; revised November 22, 2021; accepted March 9, 2022

Abstract—Mitscherlich's function is a well-known three-parameter non-linear regression function that quantifies the relation between a stimulus or a time variable and a response. It has many applications, in particular in the field of measurement reliability. Optimal designs for estimation of this function have been constructed only for normally distributed responses with homoscedastic variances. In this paper we generalize this literature to D-optimal designs for discrete and continuous responses having their distribution function in the exponential family. We also demonstrate that our D-optimal designs can be identical to and different from optimal designs for variance weighted linear regression.

DOI: 10.3103/S1066530722010033

Keywords: exponential family, generalized non-linear models, weighted least squares

1. INTRODUCTION

In different fields of science (e.g., chemistry, biology, medicine, and pharmacology) the relation between a stimulus or a time variable (x) and a response variable (y) is being studied. These relations help us quantify the beneficial or maximum levels of stimuli or time periods for the use of drugs, pollutants, foods, and other substances. These relations are typically non-linear in both the stimulus or time variable and the model parameters. For instance, the three-parameter Michaelis-Menten curve $\mathbb{E}(y|x) = \beta_1 + \beta_2 x/[\beta_3 + x]$ is frequently used for chemical and biological applications [5, 7, 20], the four-parameter logistic growth curve $\mathbb{E}(y|x) = \beta_1 + (\beta_4 - \beta_1)[1 + (x/\beta_2)^{\beta_3}]^{-1}$ is typically used for biological assays [8, 27], the four-parameter non-linear exponential growth or decay curve $\mathbb{E}(y|x) = \beta_1 + \beta_2 x + \beta_3 \exp\{\beta_4 x\}$ is used in biology and medical sciences [12, 17], and the three-parameter one-compartmental model $\mathbb{E}(y|x) = \beta_1[\exp\{-\beta_2 x\} - \exp\{-\beta_3 x\}]/[\beta_3 - \beta_2]$ is often used in pharmacokinetics [4, 9, 10].

Precise estimation of non-linear models may require a substantial amount of testing. Designing optimal experiments may therefore help reduce testing and possibly reduce also other resources (e.g., time, costs). A parameter estimation criterion for optimal designs is D-optimality [14], which maximizes the determinant of $\mathbf{X}^T\mathbf{X}$ for linear regression functions, with $\mathbf{X}$ the design matrix. For non-linear functions D-optimality is obtained by maximizing the determinant of the Fisher information matrix [14]. D-optimal designs have been studied for different types of non-linear functions for both continuous and count responses.

Under assumption of normality, $y_i = \mathbb{E}(y_i|x_i) + \varepsilon_i$, with $\varepsilon_i \sim \mathcal{N}(0, \sigma^2)$ i.i.d., [5] provided D-optimal designs for the two-parameter ($\beta_1 = 0$) Michaelis-Menten curve, while [7] discussed D-optimal designs for this two-parameter Michaelis-Menten curve under heteroscedastic residual errors, i.e., $\varepsilon_i \sim \mathcal{N}(0, \nu(\mathbb{E}(y_i|x_i)))$, with ν a known function. In [18], a D-optimal design for the three ($\beta_1 = 0$) and

*E-mail: e.r.v.d.heuvel@tue.nl

the four parameter logistic growth curve[1)] was provided, respectively. In [13], this work on logistic curves was extended to heteroscedastic residuals, i.e., $\varepsilon_i \sim \mathcal{N}(0, \sigma^2 \mathbb{E}(y_i|x_i)[1-\mathbb{E}(y_i|x_i)])$, when $\beta_1 = 0$ and $\beta_4 = 1$ holds. Under the same parameter restrictions, [13] also provided D-optimal designs for the asymmetric logistic growth curve, i.e., $\mathbb{E}(y_i|x_i) = [1 + (x/\beta_2)^{\beta_3}]^{-r}$, with $r > 0$. In [8], D-optimal designs for the full four and five parameter logistic growth curve ($\mathbb{E}(y_i|x_i) = \beta_1 + (\beta_4 - \beta_1)[1 + (x/\beta_2)^{\beta_3}]^{-r}$ were studied with residuals having a heteroscedastic variance of the form $[\mathbb{E}(y_i|x_i)]^\gamma$, with $\gamma > 0$. D-optimal designs for the two-parameter (where $\beta_2 = 0$ and $\beta_3 = 1$ and where $\beta_1 = \beta_2 = 0$) and three-parameter ($\beta_2 = 0$) exponential decay model were provided by [12] under the assumption of homoscedastic residuals, while [17] provided a D-optimal design for the full four-parameter exponential growth model (among others), also under homoscedastic residuals. Finally, [19] discussed D-optimal designs for the three-parameter one-compartmental model under homoscedastic residuals, while [8] studied D-optimality of this compartmental model under heteroscedastic residuals using $[\mathbb{E}(y_i|x_i)]^\gamma$ (again).

For count responses y_i, the Poisson, Binomial, and Negative Binomial distributions have been used frequently [11, 22, 25, 26, 28], but these papers discuss optimal designs for forms of $\mathbb{E}(y_i|x_i)$ that can be rewritten into a linear function in the parameters, i.e., satisfying the definition of generalized linear models [21]. Interestingly though, [11] provided D-optimal designs for the class of generalized linear models with distributions in the exponential family a few years earlier. Contrary to the work on generalized linear models, [20] discussed D-optimal designs for mixed effects Poisson regression with the full three-parameter Michaelis-Menten curve. The random part only affected the constant or intercept β_1 and they also discussed designs without this random component.

One specific or special non-linear regression function is the three-parameter Mitscherlich function [1], given by $\mathbb{E}(y|x) = \beta_1 + \beta_2 \exp\{\beta_3 x\}$, with $\beta_1 \in \mathbb{R}$, $\beta_2 \neq 0, \beta_3 \neq 0$, and with x the stimulus or the logarithmically transformed stimulus variable. Note that the original formulation of the Mitscherlich function in [23] assumed that the parameters β_2 and β_3 were both negative. In some areas [6, 12], the Mitscherlich function is referred to as the three parameter decay model when the variable x is time. In that case the parameter β_3 is typically considered negative. The reason that the Mitscherlich function is special, is that it can be naturally used to investigate violations of *linearity* in different directions, which is less obvious for the other non-linear functions just discussed. Indeed, linearity can be obtained in two ways:

$$\begin{aligned} \beta_3 = 1 &: \mathbb{E}(y|\log(x)) = \beta_1 + \beta_2 x, \\ \beta_1 = 0 &: \log(\mathbb{E}(y|x)) = \log(\beta_2) + \beta_3 x, \end{aligned} \tag{1}$$

with the log the natural logarithm. In case both constraints $\beta_1 = 0$ and $\beta_3 = 1$ are satisfied, the system may be referred to as *proportional* to stimulus x. This makes the Mitscherlich function a very relevant function for the validation of measurement systems where violation of linearity plays an important role.

As far as we know, D-optimal designs for the Mitscherlich non-linear function have only been discussed under the assumption of a normally distributed response y with homoscedastic residual variances [1, 6, 12]. For measurement systems these restrictive distributional assumptions may only be true for special cases. In (micro)biology the distribution of the biological response may typically deviate from normality. For instance, the biological response can be discrete, when a number of events or microorganisms is being observed. Furthermore, in chemical and (micro)biological analyses, the residual variance may typically depend on the level of the response, i.e., a mean-variance relation may exist. Thus, there are many applications where an extension of the current D-optimal designs for the Mitscherlich function is needed. Here we will generalize the D-optimal designs for estimation of the Mitscherlich function, when the discrete or continuous distribution function for the response y is from the exponential family in its natural form [2, 21]. We also consider the situation where the dispersion

[1)] Although the Michaelis–Menten curve is in principle a special form of the logistic growth curve (using reparametrization $\beta_1(L) = \beta_1(M)$, $\beta_2(L) = \beta_3^{-1}(M)$, $\beta_3(L) = 1$, $\beta_4(L) = \beta_1(M) + \beta_2(M)$, and stimulus $x(L) = x^{-1}(M)$), one may be inclined to think that literature on optimal designs for Michaelis–Menten is implied by optimal designs for logistic curves. However, optimal designs are typically minimally optimal, which means that the number of stimulus values in the optimal design is set equal to the number of parameters in the non-linear function and both functions have different numbers of parameters. Additionally, when they do study for instance two-parameter versions, the two-parameter logistic curve ($\beta_1(L) = 0$ and $\beta_4(L) = 1$) is typically different from the two-parameter Michaelis–Menten curve ($\beta_1(M) = 0$), since $\beta_3(L)$ is not considered equal to one, making research on these non-linear curves disjoint.

parameter is not known. Due to the general formulation of our D-optimal design, the complicated proofs that have been used by others for their special settings, become rather simple.

The next section will introduce our generalized non-linear model, the log-likelihood function, Fisher's information matrix, and the D-optimality criterion. In Section 3 we present our main result. We will construct the D-optimal design for the Mitscherlich non-linear function using three stimuli levels (minimally D-optimal [19]). We also provide examples for the well-known distributions in the exponential family of distributions. In Section 4 we discuss transformations of the Mitscherlich non-linear function to further generalize our results. We also show that our work generalizes the D-optimal designs in literature [1, 6, 12]. Furthermore, we show that our D-optimal design can be constructed from a D-optimal design for weighted linear regression, extending [1] directly to the distributions in the exponential family. However, when heteroscedastic residual variances are introduced, D-optimality can not be obtained through weighted linear regression anymore, demonstrating that D-optimality and weighted least squares are different approaches and only identical under specific conditions. We finalize with Section 5 summarizing and discussing our work.

2. STATISTICAL MODEL

Let y_{ij} be response $j \in \{1,2,...,n_i\}$ at stimulus x_i, $i \in \{1,2,...,m\}$, and all y_{ij} being mutually independently distributed. The distribution of y_{ij} is an element of the exponential family[2)] having density $f(y|\theta_i,\phi) = \exp\{[y\theta_i - b(\theta_i)]/a(\phi) + c(y,\phi)\}$, with $y \in \mathbb{R}$, θ_i an unknown parameter that will depend on stimulus x_i, ϕ an (un)known dispersion parameter, and $a(\cdot)$, $b(\cdot)$, and $c(\cdot,\cdot)$ known functions [21]. It is assumed that the range of y does not depend on θ_i and ϕ. Furthermore, function $b(\cdot)$ is at least twice differentiable, with $b'(\cdot)$ and $b''(\cdot)$ the first and second derivative. As a consequence, we have $\mathbb{E}(y_{ij}|x_i) \equiv \mu_i = b'(\theta_i)$ and $\mathsf{VAR}(y_{ij}|x_i) = b''(\theta_i)a(\phi)$. Using the canonical link function g, the relation between θ_i and μ_i is given by $\theta_i = g(\mu_i)$. Our model includes the well-known distributions Poisson, Binomial, Negative Binomial, Gaussian, Gamma, and Inverse Gaussian with their canonical link functions.

The Mitscherlich function we will study is $\mu_i = \beta_1 + \beta_2 x_i^{\beta_3}$ with constraints $\beta_2 > 0$, $\beta_3 > 0$, and $x_i \geq 0$ the stimulus of interest. Note that we allow a stimulus that can be equal to zero, which was not implemented in earlier formulations. Restrictions on parameter β_1 are determined by the type of distribution for y_{ij}. For instance, $\beta_1 \in \mathbb{R}$ is allowed for the normal distribution, $\beta_1 \geq 0$ is needed for the Poisson distribution, and $\beta_1 > 0$ is required for the Gamma distribution. Our choice for the Mitscherlich function fits very well with measurement system analysis where we expect typically non-negative values when we choose certain levels for the stimulus. Thus we will assume that $\beta_1 \geq 0$.

2.1. Maximum Likelihood Estimation

If we define $\mathbf{y}_i = (y_{i1}, y_{i2}, ..., y_{in_i})^T$, $\mathbf{y} = (\mathbf{y}_1, \mathbf{y}_2, ..., \mathbf{y}_m)^T$, and $\boldsymbol{\beta} = (\beta_1, \beta_2, \beta_3)^T$, the log-likelihood function can be written as

$$\begin{aligned}\ell(\boldsymbol{\beta},\phi|\mathbf{y}) &= \sum_{i=1}^{m}\sum_{j=1}^{n_i}\left[(y_{ij}\theta_i - b(\theta_i))/a(\phi) + c(y_{ij},\phi)\right] \\ &= \tfrac{1}{a(\phi)}\sum_{i=1}^{m}\left[y_{i.}g(\mu_i) - n_i b(g(\mu_i))\right] + \sum_{i=1}^{m}\sum_{j=1}^{n_i} c(y_{ij},\phi), \qquad (2)\end{aligned}$$

where $y_{i.} = \sum_{j=1}^{n_i} y_{ij}$ is the sum of the observations at stimulus x_i. The maximum likelihood estimates (MLEs) for the parameters $\boldsymbol{\beta}$ and ϕ can be obtained by solving the following likelihood equations:

$$\ell'_{\beta_k} = \frac{\partial \ell(\boldsymbol{\beta},\phi|\mathbf{y})}{\partial \beta_k} = \frac{1}{a(\phi)}\sum_{i=1}^{m}(y_{i.} - n_i\mu_i)\, g'(\mu_i)\,\frac{\partial \mu_i}{\partial \beta_k} = 0 \quad \forall k = 1,2,3,$$

[2)]Note that there exists a more general formulation of the exponential family of distributions of the form $f(y|\theta_i) = \exp\{T(y)\eta(\theta_i) - A(\theta_i) + B(y)\}$, but that we have selected its more restrictive natural form with θ_i the canonical parameter when ϕ is known [2]. Furthermore, if ϕ is unknown our formulation may not be a two-parameter exponential family anymore [21]. Irrespective of its formal definition, we will focus on densities $f(y|\theta_i,\phi) = \exp\{[y\theta_i - b(\theta_i)]/a(\phi) + c(y,\phi)\}$ where ϕ is allowed to be unknown.

$$\ell'_{\phi} = \frac{\partial \ell\left(\boldsymbol{\beta}, \phi | \mathbf{y}\right)}{\partial \phi} = -\frac{a'\left(\phi\right)}{a^2\left(\phi\right)} \sum_{i=1}^{m} \left[y_{i.} g(\mu_i) - n_i b(g(\mu_i))\right] + \sum_{i=1}^{m} \sum_{j=1}^{n_i} c'(y_{ij}, \phi) = 0, \tag{3}$$

where $g'(\mu_i) = \partial g(\mu_i)/\partial \mu_i$, $a'(\phi) = \partial a(\phi)/\partial \phi$, and $c'(y_{ij}, \phi) = \partial c(y_{ij}, \phi)/\partial \phi$. The 4×4 Fisher information matrix $I_{4\times 4}(\boldsymbol{\beta}, \phi)$ is obtained by the (negative) expected values of the derivatives of the score functions in (3), but the elements of $I_{4\times 4}(\boldsymbol{\beta}, \phi)$ are also equal to the variances and covariances of the score functions [16, Theorem 1.1, p. 406]. Using the derivatives of the score functions and taking expectations (see for details Appendix A), the variances and covariances of the score functions become

$$\begin{gathered}
\mathsf{VAR}(\ell'_{\beta_k}) = \frac{1}{a(\phi)} \sum_{i=1}^{m} n_i g'(\mu_i) \left(\frac{\partial \mu_i}{\partial \beta_k}\right)^2, \\
\mathsf{VAR}(\ell'_{\phi}) = \sum_{i=1}^{m} \sum_{j=1}^{n_i} \left[\left(\frac{a''\left(\phi\right)}{a'\left(\phi\right)} - \frac{2a'\left(\phi\right)}{a\left(\phi\right)}\right) \mathbb{E}\left(\frac{\partial c(y_{ij}, \phi)}{\partial \phi}\right) - \mathbb{E}\left(\frac{\partial^2 c(y_{ij}, \phi)}{(\partial \phi)^2}\right)\right], \\
\mathsf{COV}(\ell'_{\beta_r}, \ell'_{\beta_s}) = \frac{1}{a(\phi)} \sum_{i=1}^{m} n_i g'(\mu_i) \left(\frac{\partial \mu_i}{\partial \beta_r}\right) \left(\frac{\partial \mu_i}{\partial \beta_s}\right), \quad r \neq s, \\
\mathsf{COV}(\ell'_{\beta_k}, \ell'_{\phi}) = 0.
\end{gathered} \tag{4}$$

Orthogonality of the score functions for the location parameters $\boldsymbol{\beta}$ and the score function for the dispersion parameter ϕ has been obtained earlier [3]. Furthermore, when ϕ would be known (e.g., $\phi = 1$), the Fisher information matrix $I_{4\times 4}(\boldsymbol{\beta}, \phi)$ reduces to a 3×3 matrix $I_{3\times 3}(\boldsymbol{\beta})/a(\phi)$, fully determined by the score functions ℓ'_{β_k} in (4), and with $I_{3\times 3}(\boldsymbol{\beta})$ independent of ϕ. Note that $\mathsf{COV}(\ell'_{\beta_k}, \ell'_{\phi}) = 0$ for all $k \in \{1, 2, 3\}$, implies that the covariance of the MLEs for β_k and ϕ is zero too, but this does not necessarily imply that the variance $\mathsf{VAR}(\hat{\boldsymbol{\beta}})$ of MLE $\hat{\boldsymbol{\beta}}$ is independent of ϕ or the variance $\mathsf{VAR}(\hat{\phi})$ of MLE $\hat{\phi}$ is independent of $\boldsymbol{\beta}$, since the corresponding elements of the inverse Fisher information may still depend on ϕ or $\boldsymbol{\beta}$ through its density, respectively.

2.2. D-Optimality Criterion

D-Optimality is defined by maximizing the determinant of the Fisher information matrix $I_{4\times 4}(\boldsymbol{\beta}, \phi)$, see [14, 15]. Due to the (asymptotic) independence of the ML estimators $\hat{\boldsymbol{\beta}}$ and $\hat{\phi}$, the determinant of the Fisher information matrix can be rewritten as $|I_{4\times 4}(\boldsymbol{\beta}, \phi)| = \mathsf{VAR}(\ell'_{\phi}) |I_{3\times 3}(\boldsymbol{\beta})| / a(\phi)$. In case the variance $\mathsf{VAR}(\hat{\phi})$ of MLE $\hat{\phi}$ is independent of $\boldsymbol{\beta}$, i.e.,

$$\frac{\partial \mathsf{VAR}(\ell'_{\phi})}{\partial \beta_k} = 0, \quad \forall k \in \{1, 2, 3\}, \tag{5}$$

we can focus on determinant $|I_{3\times 3}(\boldsymbol{\beta})|$, as if the dispersion parameter ϕ would be known.[3)] Condition (5) is satisfied for exponential families of distributions of the form $f(y|\theta, \eta) = \exp\{\eta[y\theta_i - b(\theta_i)] + d_1(y) + d_2(\eta) + \eta c(y)\}$, where $\eta = 1/a(\phi)$, since the derivative $\partial \ell'_{\eta}/\partial \eta$ of the score function ℓ'_{η} is independent of y and $\boldsymbol{\beta}$ (see [29, formula (1.22), p. 8]). Condition (5) holds for all well-known distribution functions Poisson, Binomial, Negative Binomial, Gaussian, Gamma, and inverse Gaussian (see [29, Table 1.1]), but we can not rule out that there may exist an exotic distribution function in our formulation of the exponential family of distributions for which condition (5) would not hold.

We are interested in the smallest number of stimuli that would maximize determinant $|I_{4\times 4}(\boldsymbol{\beta}, \phi)|$, i.e., the locally minimal D-optimality criterion [19]. Assuming that condition (5) holds true, we can focus

[3)] Note that we do not need a fourth stimulus to be able to estimate parameter ϕ. The reason is that the MLE of $\boldsymbol{\beta}$ can be obtained independently of the estimation of ϕ because the likelihood equations for $\boldsymbol{\beta}$ do not involve the parameter ϕ, see (3). Additionally, ϕ can be estimated from the variability in the observations y_{ij} if $n > 1$, since $\mathsf{VAR}(y_{ij}|x_i) = b''(\theta_i)a(\phi)$ and θ_i can be estimated with MLE $\hat{\boldsymbol{\beta}}$ and x_i.

on only three stimuli x_1, x_2, and x_3, since determinant $|I_{3\times3}(\boldsymbol{\beta})|$ contains only three parameters. Thus we are looking for stimuli x_1, x_2, and x_3, with $x_1 < x_2 < x_3$, such that

$$\underset{L\leq x_1<x_2<x_3\leq U}{\arg\max} |I_{3\times3}(\boldsymbol{\beta})|, \tag{6}$$

with $L \geq 0$ and $U < \infty$ a known lower and upper bound on the range of stimuli, respectively, typically determined by practical limitations. With the help of Matlab we were able to express the determinant $|I_{3\times3}(\boldsymbol{\beta})|$ in an explicit form equal to

$$\beta_2^2 \left[(x_1x_2)^{\beta_3}\log\left(\frac{x_2}{x_1}\right) - (x_1x_3)^{\beta_3}\log\left(\frac{x_3}{x_1}\right) + (x_2x_3)^{\beta_3}\log\left(\frac{x_3}{x_2}\right)\right]^2 \prod_{i=1}^{3}[n_i g'(\mu_i)]. \tag{7}$$

It is important to realize that the sample sizes n_1, n_2, and n_3 do not influence the choice of stimuli x_1, x_2, and x_3 for maximization of (7), since only the product $n_1n_2n_3$ is involved in (7). Thus if the optimal design is known and the total sample size $n = n_1 + n_2 + n_3$ is determined, it would be best to choose the same sample size in each stimulus to maximize precision, or equivalently to minimize the variance of the parameter estimates.

3. MAIN RESULTS: D-OPTIMAL DESIGNS

Here we will focus on finding the optimal values for x_1, x_2, and x_3 that would maximize determinant $|I_{3\times3}(\boldsymbol{\beta})|$ in (7) under constraint $L \leq x_1 < x_2 < x_3 \leq U$, with $L \geq 0$ and $U < \infty$. We will see that the choice of the three stimuli depends on the mathematical behavior of the link function g. Note that our results will be D-optimal when either ϕ is known or otherwise when condition (5) is satisfied. Our main results are formulated in the following three theorems. The proofs are provided in Appendix B.

Theorem 1. *If $g'(\mu) \geq 0$, and $g''(\mu) \leq 0$ holds, then the optimal stimulus x_1^{opt} for x_1 that maximizes determinant $|I_{3\times3}(\boldsymbol{\beta})|$ in (7), is the smallest possible stimulus value, i.e., $x_1^{\text{opt}} = L$.*

Proof. See Appendix B. □

The two conditions on the link function in Theorem 1 indicate that we are studying concave increasing link functions. This is satisfied for the identity link function $g(\mu) = \mu$ with $\mu \in \mathbb{R}$, the log link function $g(\mu) = \log(\mu)$ with $\mu \in (0,\infty)$, the square root link function $g(\mu) = \sqrt{\mu}$ with $\mu \in (0,\infty)$, the negative inverse link function $g(\mu) = -\mu^{-1}$ with $\mu \in (0,\infty)$, and the half negative inverse-square link function $g(\mu) = -0.5\mu^{-2}$ with $\mu \in (0,\infty)$. Thus for these link functions we need to choose the first stimulus x_1 as small as possible if we want to maximize the determinant of the Fisher information matrix. For the logit link function $g(\mu) = \log(\mu/[N-\mu])$ with $\mu \in (0, N)$ the condition $g''(\mu) \leq 0$ is only guaranteed when $\mu \leq N/2$. Thus when $\mu > N/2$, we do not know if stimulus x_1 should be selected as small as possible.

Theorem 2. *If $g'(\mu) \geq 0$, $g''(\mu) \leq 0$, and $g''(\mu)\mu + 2g'(\mu) \geq 0$ holds, then the optimal stimulus x_3^{opt} for x_3 that maximizes determinant $|I_{3\times3}(\boldsymbol{\beta})|$ in (7), is the largest possible value, i.e., $x_3^{\text{opt}} = U$.*

Proof. See Appendix B □.

The third condition $g''(\mu)\mu + 2g'(\mu) \geq 0$ in Theorem 2 for link function g would be satisfied for most of the link functions (e.g., $g(\mu) = \mu$, $g(\mu) = \log(\mu)$, $g(\mu) = \sqrt{\mu}$, $g(\mu) = \log(\mu/[N-\mu])$, and $g(\mu) = -\mu^{-1}$), but it does not hold for the canonical link function $g(\mu) = -0.5\mu^{-2}$, with $\mu \in (0,\infty)$, for the inverse Gaussian distribution. Recall that the canonical link function $g(\mu) = \log(\mu/[N-\mu])$ of the Binomial distribution satisfies condition $g''(\mu) \leq 0$ only when $\mu \leq N/2$. Thus the third stimulus x_3 should be chosen as large as possible for most canonical link functions, but for the inverse Gaussian and Binomial distribution with their canonical link function it may be possible to obtain better designs when we stay away from the boundary value U (see Section 3.1).

Theorem 3. *Assume that $g'(\mu) \geq 0$, and $g''(\mu) \leq 0$ holds and let x_1 and x_3 be given stimuli, then the optimal stimulus x_2^{opt} for x_2 that maximizes determinant $|I_{3\times3}(\boldsymbol{\beta})|$ in (7), is obtained by solving the following equation*

$$\beta_2\beta_3 g''(\mu_2)\left[(x_1x_2)^{\beta_3}\log\left(\frac{x_2}{x_1}\right) - (x_1x_3)^{\beta_3}\log\left(\frac{x_3}{x_1}\right) + (x_2x_3)^{\beta_3}\log\left(\frac{x_3}{x_2}\right)\right]$$

$$+ 2g'(\mu_2)\left[\beta_3 x_1^{\beta_3}\log\left(\frac{x_2}{x_1}\right) + \beta_3 x_3^{\beta_3}\log\left(\frac{x_3}{x_2}\right) + x_1^{\beta_3} - x_3^{\beta_3}\right] = 0. \quad (8)$$

The optimal solution $x_2^{\rm opt}$ *is an element of interval* (x_1, x_3) *and satisfies constraint* $\beta_3\log(x_2^{\rm opt}) \leq [x_3^{\beta_3}\log(x_3^{\beta_3}) - x_1^{\beta_3}\log(x_1^{\beta_3}) - (x_3^{\beta_3} - x_1^{\beta_3})]/[x_3^{\beta_3} - x_1^{\beta_3}]$.

Proof. See Appendix B. □

Theorems 1–3 all formulate optimal choices of only one stimulus, conditionally on the other two stimuli, whether these other two stimuli are chosen optimally or not. The theorems tell us what to do with this one stimulus to maximize determinant $|I_{3\times3}(\boldsymbol{\beta})|$ when the other stimuli are already provided. For Theorem 3 it shows that the best choice for $x_2 \in (x_1, x_3)$ is the value that solves Eq. (8), if we wish to maximize determinant $|I_{3\times3}(\boldsymbol{\beta})|$.

Corollary 1. *If the conditions of Theorem* 3 *hold and* $x_1 = 0$*, then the optimal stimulus* $x_2^{\rm opt}$ *for* x_2 *that maximizes determinant* $|I_{3\times3}(\boldsymbol{\beta})|$ *in* (7)*, is obtained by solving equation*

$$\beta_2\beta_3 g''(\mu_2)\, x_2^{\beta_3}\log\left(\frac{x_3}{x_2}\right) + 2g'(\mu_2)\left[\beta_3\log\left(\frac{x_3}{x_2}\right) - 1\right] = 0 \quad (9)$$

and the optimal solution satisfies $0 < x_2^{\rm opt} \leq x_3\exp\{-\beta_3^{-1}\}$.

Proof. Substituting $x_1 = 0$ in (8) leads directly to Eq. (9), since the value of the third stimulus is always larger than zero (i.e., $x_3 > 0$) to guarantee that we have three different stimuli. Furthermore, substituting $x_1 = 0$ in the boundaries on $x_2^{\rm opt}$ results in the lower and upper boundary 0 and $x_3\exp\{-\beta_3^{-1}\}$, respectively. □

Corollary 1 and Theorem 3 demonstrate that the optimal value $x_2^{\rm opt}$ for the second stimulus depends on the two other stimuli x_1 and x_3, and on the link function g through its derivatives g' and g''. Equations (8) and (9) also show that distributions with the same link function result in the same D-optimal design. Thus the D-optimal designs for a Poisson and Negative Binomial distributed response y are identical, since they both have canonical link function $g(\mu) = \log(\mu)$ and they just differ in the dispersion variable ϕ. Whether all three model parameters $\boldsymbol{\beta}$ are involved in $x_2^{\rm opt}$, depends on the link function (see Section 3.1), but it always involves the power parameter β_3.

3.1. Examples of D-Optimal Designs

We will discuss D-optimal designs for the well-known distribution functions of the exponential family of distributions using their canonical link function. Note that it is common practice to use the canonical link functions in modeling data from measurement reliability studies. As illustration we will assume a measurement system analysis for which the stimuli can range from $L = 0$ to $U = 15$. We will consider six combinations of parameter settings for $\beta_1 \in \{0.5, 1.0\}$, $\beta_2 \in \{0.8, 1.0, 1.2\}$, and $\beta_3 \in \{0.9, 1.0, 1.1\}$.

Gaussian distribution. For the Gaussian distribution with the identity link function, Eq. (8) can be solved explicitly. Theorems 1 and 2 imply that the first and third optimal stimulus should be chosen equal to $x_1^{\rm opt} = L$ and $x_3^{\rm opt} = U$, respectively. Then the optimal second stimulus is equal to

$$x_2^{\rm opt} = \exp\left\{[U^{\beta_3}\log(U) - L^{\beta_3}\log(L)]/[U^{\beta_3} - L^{\beta_3}] - \beta_3^{-1}\right\}, \quad (10)$$

which depends only on the power parameter β_3 (and not on the intercept β_1 and slope β_2). In case the lower boundary L is equal to zero, the optimal second stimulus reduces to $x_2^{\rm opt} = U\exp\{-\beta_3^{-1}\}$, which is equal to the upper bound on $x_2^{\rm opt}$ mentioned in Corollary 1. Table 1 shows the optimal stimulus $x_2^{\rm opt}$ for our illustration. Since $\beta_3 \approx 1$, the stimulus is approximately 36.8% of the upper boundary $U = 15$.

Poisson and negative binomial distribution. The canonical link function is $g(\mu) = \log(\mu)$, which implies that $x_1^{\rm opt} = L$ and $x_3^{\rm opt} = U$ (Theorems 1 and 2, respectively). A solution for Eq. (8) can only be obtained numerically and this equation contains all three parameters of the Mitscherlich function. When $L = 0$, Eq. (9) reduces to

$$\left[\beta_1 + \beta_2 x_2^{\beta_3}\right]\left[\beta_3\log\left(\frac{U}{x_2}\right) - 2\right] + \beta_1\beta_3\log\left(\frac{U}{x_2}\right) = 0, \quad (11)$$

Table 1. Optimal value for the second stimulus x_2 for different distributions of the exponential family ($L = 0$ and $U = 15$)

Parameters			Gaussian	Poisson	Gamma	Binomial		
β_1	β_2	β_3				$N = 25$	$N = 50$	$N = 100$
0.5	1.2	0.9	4.94	2.24	0.70	2.65	2.41	2.32
0.5	1	1	5.52	2.67	0.90	3.16	2.87	2.76
0.5	0.8	1.1	6.04	3.10	1.14	3.66	3.33	3.20
1.0	1.2	0.9	4.94	2.58	1.12	3.04	2.77	2.67
1.0	1	1	5.52	3.02	1.38	3.57	3.25	3.13
1.0	0.8	1.1	6.04	3.47	1.68	4.08	3.71	3.58

which can not be solved explicitly either and still depends on all three parameters. However, the optimal solution for the second stimulus x_2^{opt} is inside the interval $[U\exp\{-2\beta_3^{-1}\}, U\exp\{-\beta_3^{-1}\}]$. Indeed, the left-hand side in (11) is non-negative when $\beta_3 \log(U/x_2) \geq 2$, which means that $x_2^{\text{opt}} \in [U\exp\{-2\beta_3^{-1}\}, U)$. In addition, the left-hand side in (11) is non-positive if $\beta_3 \log(U/x_2) \leq 1$, which means that $x_2^{\text{opt}} \in (0, U\exp\{-\beta_3^{-1}\}]$, but this was already known from the upper boundary in Corollary 1. When the intercept $\beta_1 = 0$, the optimal value for the second stimulus becomes $x_2^{\text{opt}} = U\exp\{-2\beta_3^{-1}\}$, which is different from the solution of the Gaussian distribution. Table 1 shows that the optimal stimulus x_2^{opt} is substantially lower than the solution of the Gaussian distribution with the identity link function.

Gamma distribution. The canonical link function $g(\mu) = -\mu^{-1}$ together with Theorems 1 and 2, imply that $x_1^{\text{opt}} = L$ and $x_3^{\text{opt}} = U$. The solution of Eq. (8) can be determined numerically and it involves all three parameters β_1, β_2, and β_3 of the Mitscherlich function. If we assume again that $L = 0$, Eq. (9) reduces to

$$\beta_1 + \beta_2 x_2^{\beta_3} + \beta_1\beta_3 \log(x_2) = \beta_1\beta_3 \log(U), \tag{12}$$

which can be solved numerically for different values of β_1, β_2, and β_3. In case $\beta_1 = 0$, the optimal second stimulus becomes equal to $x_2^{\text{opt}} = 0$, but these results are not allowed for a gamma distribution with a positive range. The parameter β_1 should be positive when we allow the stimulus x_1 to be equal to zero. Table 1 shows that the optimal stimulus x_2^{opt} is still close to zero when $\beta_1 > 0$.

Binomial distribution. To solve Eqs. (8) or even (9) for the Binomial distribution with canonical link function $g(\mu) = \log(\mu/[N-\mu])$ is very tedious and does not easily reduce into manageable functions. The solution x_2^{opt} depends on all three parameters β_1, β_2, and β_3 of the Mitscherlich function and numerical approaches should be used to determine the D-optimal design. When $\beta_1 + \beta_2 U^{\beta_3} \leq N/2$ we know that $x_1^{\text{opt}} = L$ and $x_3^{\text{opt}} = U$ based on Theorems 1 and 2. Thus for our illustration with $L = 0$ and $U = 15$ and the six selected combinations of parameter settings $\beta_1 \in \{0.5, 1.0\}$, $\beta_2 \in \{0.8, 1.0, 1.2\}$, and $\beta_3 \in \{0.9, 1.0, 1.1\}$ in Table 1, a sample size of $N = 34$ would be enough to satisfy the condition $g''(\mu) \leq 0$ in Theorems 1–3. Thus for $N = 50$ and $N = 100$, we know that $x_1^{\text{opt}} = L = 0$ and $x_3^{\text{opt}} = U = 15$, but for $N = 25$ we do not know this. Investigating the optimal stimuli $x_i \in [0, 15]$ using numerical calculations (just calculating the determinant for a grid of stimuli using step size $h = 0.01$) still shows that $x_1^{\text{opt}} = 0$ and $x_3^{\text{opt}} = 15$. Table 1 shows the results of the optimal stimulus x_2^{opt} for all settings. The results show that x_2^{opt} is close to the optimal value of the Poisson, which is not a surprise since the Binomial and Poisson distribution are very similar, in particular when the sample size N is increasing. This resemblence between the two distributions may explain why the optimal stimulus x_1^{opt} and x_3^{opt} are still equal to $L = 0$ and $U = 15$ when $N = 25$.

Table 2. D-optimal design for the Inverse Gaussian distribution with canonical link function $g(\mu) = -0.5\mu^{-2}$ and boundaries $L = 0$ and $U = 15$

Parameters			Optimal design*				Equidistant designs, %		
β_1	β_2	β_3	x_1^{opt}	x_2^{opt}	x_3^{opt}	$\lvert I_{3\times 3}(\boldsymbol{\beta})\rvert$	$d = 60$	$d = 30$	$d = 15$
0.5	1.2	0.9	0	0.26	5.21	1.455	70.8	55.4	21.0
0.5	1	1	0	0.36	5.32	1.697	64.5	65.2	32.3
0.5	0.8	1.1	0	0.48	5.58	2.192	51.8	68.8	45.3
1.0	1.2	0.9	0	0.57	11.34	0.045	64.9	73.4	45.9
1.0	1	1	0	0.72	10.65	0.053	51.3	73.7	59.4
1.0	0.8	1.1	0	0.91	10.53	0.068	36.0	66.4	69.8

* We have assumed that the sample sizes are equal to one ($n_1 = n_2 = n_3 = 1$) for calculation of $I_{3\times 3}(\boldsymbol{\beta})$.

Inverse gaussian distribution. Since the canonical link function $g(\mu) = -0.5\mu^{-2}$ does not satisfy the conditions of Theorem 2, i.e., $\mu g''(\mu) + 2g'(\mu) = -\mu^{-3} < 0$ for $\mu \in (0, \infty)$, it is not known how to choose the optimal value for the third stimulus x_3, but we know from Theorem 1 that $x_1^{\text{opt}} = L$. If we assume that $L = 0$, determinant $|I_{3\times 3}(\boldsymbol{\beta})|$ becomes equal to

$$n_1 n_2 n_3 \beta_1^{-3} \beta_2^2 \left[(x_2 x_3)^{\beta_3} \log(\tfrac{x_3}{x_2})\right]^2 \left[\beta_1 + \beta_2 x_2^{\beta_3}\right]^{-3} \left[\beta_1 + \beta_2 x_3^{\beta_3}\right]^{-3}. \tag{13}$$

Given the parameters β_1, β_2, and β_3 a grid search for x_2 and x_3 can be conducted to maximize (13). Table 2 shows the optimal stimuli for a measurement study where the stimuli can range from $L = 0$ to $U = 15$. We used a step size of 0.01 in our grid search. Table 2 shows that x_3^{opt} is far away from the boundary $U = 15$ and x_2^{opt} is relatively close to $L = 0$.

The optimal solution for the second stimulus should satisfy Eq. (9), which can be rewritten in

$$\beta_3 \left[\beta_2 x_2^{\beta_3} - 2\beta_1\right] \log(\tfrac{x_3}{x_2}) + 2\left[\beta_1 + \beta_2 x_2^{\beta_3}\right] = 0. \tag{14}$$

In case $\beta_2 x_2^{\beta_3} - 2\beta_1 > 0$ the left-hand side in (14) is positive, while it is negative when x_2 gets close to zero. This implies that $x_2^{\text{opt}} \in (0, [2\beta_1/\beta_2]^{1/\beta_3})$, illustrating that x_2^{opt} can never be far away from zero (unless β_2 is close to zero). This upper bound on x_2 can be useful in a grid search for maximization of (13), since x_2 should never go beyond $[2\beta_1/\beta_2]^{1/\beta_3}$ and x_3 should never start before $[2\beta_1/\beta_2]^{1/\beta_3}$ when x_2^{opt} reaches this bound.

Furthermore, in a measurement system analysis it is common to use equidistant stimuli designs, either in the original scale or otherwise in the logarithmic scale. Considering the D-optimal design in Table 2, an equidistant design in the logarithmic scale is closer to the D-optimal design than an equidistant design in its original scale, although the dilution factor varies with the parameters β_1, β_2, and β_3. The efficiency of these so-called dilution designs with respect to the optimal design is provided in Table 2 for different dilution factors d (and taking $x_3 = U$). It is obvious that our D-optimal design is substantially more effcient than a dilution experiment.

4. RELATIONS TO EARLIER WORK AND EXTENSIONS

As we mentioned earlier, D-optimal designs for the Mitscherlich function were already obtained for the normal distribution with homogeneous residual variances [1, 6, 12], but they used different parametrizations of the Mitscherlich function. Here we will show that these parametrizations are irrelevant and that their D-optimal results follow from our work. We will also show that under certain conditions our D-optimal solution can be obtained from minimizing a weighted least squares. However, when we start considering heteroscedastic residual variances, which may be typical in measurement reliability studies, these two approaches for optimal designs can also be different. This shows that weighted least squares solutions and D-optimal designs are not always identical. We finish with a discussion on extending our work to transformations of the Mitscherlich function $\psi(\beta_1 + \beta_2 x_i^{\beta_3})$, making our general results even more general.

4.1. Existing D-Optimal Designs

In our results we formulated the Mitscherlich non-linear function as $\mathbb{E}(y_{ij}|x_i) = \beta_1 + \beta_2 x_i^{\beta_3}$, with x_i a non-negative stimulus, and β_2 and β_3 both positive. However, Box and Lucas [1] introduced the Mitscherlich function as $\mathbb{E}(y_{ij}|z_i) = \beta_1 - \beta_2 \exp\{-\beta_3 z_i\}$, with $\beta_2 > 0$ and $\beta_3 > 0$, Han and Chaloner [12] used $\mathbb{E}(y_{ij}|z_i) = \beta_1 + \beta_2 \exp\{-\beta_3 z_i\}$, with $\beta_2 > 0$ and $\beta_3 > 0$, and Dette et al., [6] used $\mathbb{E}(y_{ij}|z_i) = \beta_1 + \beta_2 \exp\{z_i/\tilde{\beta}_3\}$, with $\beta_2 > 0$ and $\tilde{\beta}_3 > 0$. Due to these reparametrizations, the published D-optimal designs are difficult to compare. Here we will show that it is reasonably straightforward to calculate their D-optimal designs from our results. We will show that their results are just obtained from reformulations of our own form $\mathbb{E}(y_{ij}|x_i) = \beta_1 + \beta_2 x_i^{\beta_3}$ with $\beta_2 > 0$ and $\beta_3 > 0$.

For Dette et al.'s formulation under their assumption of normality with homoscedastic residual variances, we need to both reparametrize β_3 and transform the stimulus. Our stimulus x_i can be taken equal to $x_i = \exp\{z_i\}$ and the power parameter β_3 can be taken equal to $\beta_3 = 1/\tilde{\beta}_3$. Here the transformation for the stimulus is an increasing function of z_i, thus Theorems 1 and 2 indicate we need to choose z_1^{opt} as small as possible and z_3^{opt} as large as possible. In case z_i represents time, the smallest value could potentially be zero, implying that $x_1^{\text{opt}} \geq 1$. Using optimal solution (10) with $L = \exp\{z_1^{\text{opt}}\}$, $U = \exp\{z_3^{\text{opt}}\}$, $\beta_3 = 1/\tilde{\beta}_3$, and $x_2^{\text{opt}} = \exp\{z_2^{\text{opt}}\}$, the optimal solution z_2^{opt} is now equal to

$$z_2^{\text{opt}} = \frac{z_3^{\text{opt}} \exp\{z_3^{\text{opt}}/\tilde{\beta}_3\} - z_1^{\text{opt}} \exp\{z_1^{\text{opt}}/\tilde{\beta}_3\}}{\exp\{z_3^{\text{opt}}/\tilde{\beta}_3\} - \exp\{z_1^{\text{opt}}/\tilde{\beta}_3\}} - \tilde{\beta}_3, \tag{15}$$

which was indeed presented in [6].

The optimal solutions for the other two formulations can be obtained in a similar way. For the formulation of Box and Lucas, we can take $x_i = \exp\{z_i\}$. If one realizes that our proofs of Theorems 1–3 remain correct if both β_2 and β_3 become negative (instead of being both positive), we obtain also $z_1^{\text{opt}} = z_{\min}$, $z_3^{\text{opt}} = z_{\max}$, with $z_{\min}$ and $z_{\max}$ the minimal and maximal allowable value for stimulus z, and z_2^{opt} satisfies (15) with $\tilde{\beta}_3$ replaced by $-\beta_3^{-1}$, which was obtained by [1]. For the formulation of Han and Chaloner, with $x_i = \exp\{-z_i\}$, we obtain $z_1^{\text{opt}} = z_{\max}$, $z_3^{\text{opt}} = z_{\min}$, and

$$z_2^{\text{opt}} = \frac{z_1^{\text{opt}} \exp\{-\beta_3 z_1^{\text{opt}}\} - z_3^{\text{opt}} \exp\{-\beta_3 z_3^{\text{opt}}\}}{\exp\{-\beta_3 z_1^{\text{opt}}\} - \exp\{-\beta_3 z_3^{\text{opt}}\}} + \frac{1}{\beta_3},$$

which was reported by [12]. The order for z_1^{opt} and z_3^{opt} is changed, since the stimulus $x_i = \exp\{-z_i\}$ is now a decreasing function of z_i.

4.2. Weighted Least Squares and Heteroscedasticity

The optimal design for the Mitscherlich function that was proposed by Box and Lucas [1] in 1959, was based on the linearization of the Mitscherlich non-linear function and the maximization of the determinant of the corresponding design matrix (as if they were constructing a D-optimal design for a linear regression problem [24]). Their design matrix was equal to

$$\mathbf{X} = \begin{pmatrix} \frac{\partial \mu_1}{\partial \beta_1} & \frac{\partial \mu_1}{\partial \beta_2} & \frac{\partial \mu_1}{\partial \beta_3} \\ \frac{\partial \mu_2}{\partial \beta_1} & \frac{\partial \mu_2}{\partial \beta_2} & \frac{\partial \mu_2}{\partial \beta_3} \\ \frac{\partial \mu_3}{\partial \beta_1} & \frac{\partial \mu_3}{\partial \beta_2} & \frac{\partial \mu_3}{\partial \beta_3} \end{pmatrix} = \begin{pmatrix} 1 & x_1^{\beta_3} & \beta_2 x_1^{\beta_3} \log(x_1) \\ 1 & x_2^{\beta_3} & \beta_2 x_2^{\beta_3} \log(x_2) \\ 1 & x_3^{\beta_3} & \beta_2 x_3^{\beta_3} \log(x_3) \end{pmatrix} \tag{16}$$

and they maximized determinant $|\mathbf{X}^T\mathbf{X}|$ over $L \leq x_1 < x_2 < x_3 \leq U$, assuming that at each stimulus the same sample size was used. For an imbalanced design the determinant becomes $|\mathbf{X}^T\mathbf{W}\mathbf{X}|$, with $\mathbf{W}$ a 3×3 diagonal matrix with n_1, n_2, and n_3 at the diagonal. Under the assumption of normality with the identity link function, this linearization with an imbalanced design leads to the maximization of $|I_{3\times3}(\boldsymbol{\beta})|$ in (7), see also the variances of the score functions in (4). However, for other distributions, with another canonical link function than the identity, our D-optimal design would deviate from the optimal design of [1] since the linearization should involve the link function.

To generalize the approach of [1], we should change $\mathbf{X}^T\mathbf{X}$ such that it represents the variances in (4). Thus if we would choose weight matrix $\mathbf{W}$ by

$$\mathbf{W} = \begin{pmatrix} n_1 g'(\mu_1)/a(\phi) & 0 & 0 \\ 0 & n_2 g'(\mu_2)/a(\phi) & 0 \\ 0 & 0 & n_3 g'(\mu_3)/a(\phi) \end{pmatrix} \qquad (17)$$

and maximize determinant $|\mathbf{X}^T\mathbf{W}\mathbf{X}|$, we would maximize determinant $|I_{3\times3}(\boldsymbol{\beta})|$ in (7). Thus by changing the least squares approach of [1] to a weighted least squares approach, we obtain the D-optimal designs for the Mitscherlich non-linear function for any of the distributions in the exponential family that satisfy condition (5). It should be noted that the weight $w_i = n_i a^{-1}(\phi) g'(\mu_i)$ in (17) is equal to $w_i = n_i[\mathsf{VAR}(y_{ij})]^{-1}$, the typical weights used in a weighted linear regression approach.

If we return to the normal distribution again and assume that the dispersion parameter ϕ depends on the mean μ, i.e., $y_{ij} \sim N(\mu_i, \sigma^2\varphi(\mu_i))$, with φ a positive function that is twice differentiable, and σ^2 known, the Fisher information matrix $I_{3\times3}(\boldsymbol{\beta})$ becomes equal to (Appendix C):

$$\sum_{i=1}^{3} n_i h(\mu_i) \begin{pmatrix} 1 & x_i^{\beta_3} & \beta_2 x_i^{\beta_3}\log(x_i) \\ x_i^{\beta_3} & x_i^{2\beta_3} & \beta_2 x_i^{2\beta_3}\log(x_i) \\ \beta_2 x_i^{\beta_3}\log(x_i) & \beta_2 x_i^{2\beta_3}\log(x_i) & [\beta_2 x_i^{\beta_3}\log(x_i)]^2 \end{pmatrix} \qquad (18)$$

with $h(\mu) = 0.5[\varphi'(\mu)/\varphi(\mu)]^2 + [\sigma^2\varphi(\mu)]^{-1}$, φ' the first derivative of φ, and with the summation in (18) taken element wise. This Fisher information matrix has strong similarities with the Fisher information matrix for constructing optimal designs for the Michaelis–Menten curve studied in [7]. If we now take the weight $w_i = n_i[\mathsf{VAR}(y_{ij})]^{-1}$ and consider $\mathbf{X}^T\mathbf{W}\mathbf{X}$, we obtain matrix (18) with $h(\mu_i)$ equal to $[\sigma^2\varphi(\mu_i)]^{-1}$. Thus under heteroscedasticity, the usual inverse variance weight $w_i = n_i[\mathsf{VAR}(y_{ij})]^{-1}$ does not lead to a D-optimal design, but if we choose the weight $w_i = n_i h(\mu_i)$, with $h(\mu_i)$ as defined in (18), $\mathbf{X}^T\mathbf{W}\mathbf{X}$ becomes equal to (18).

The determinant of $I_{3\times3}(\boldsymbol{\beta})$ in (18) becomes equal to (7) with $g'(\mu_i)$ replaced by $h(\mu_i)$, making use of Matlab. Thus the solutions x_1, x_2, and x_3 that maximize $|I_{3\times3}(\boldsymbol{\beta})|$ are determined by our Theorems 1, 2, and 3 when the following three conditions are satisfied $h(\mu) \geq 0$, $h'(\mu) \leq 0$, and $\mu h'(\mu) + 2h(\mu) \geq 0$. If we would assume that the dispersion parameter is a power function of the mean, i.e., $\varphi(\mu) = \mu^p$, with $p > 0$, we obtain that $h(\mu) = 0.5p^2\mu^{-2} + \sigma^{-2}\mu^{-p}$, $h'(\mu) = -p[p\mu^{-3} + \sigma^{-2}\mu^{-p-1}]$, and $\mu h'(\mu) + 2h(\mu) = [2-p]\sigma^{-2}\mu^{-p}$. Thus conditions $h(\mu) \geq 0$ and $h'(\mu) \leq 0$ are always satisfied when $p > 0$, but condition $\mu h'(\mu) + 2h(\mu) \geq 0$ is only satisfied when $0 < p \leq 2$, implying that we would only choose x_3 equal to its maximum value when $p \leq 2$. When we assume that $x_1^{\text{opt}} = 0$ and $p \in (0, 2]$, the optimal solution x_2^{opt} follows from (9) with $g' = h$, and should satisfy equation

$$\beta_3 \left[\frac{(2-p)}{\sigma^2\mu_2^p} + \frac{p\beta_1}{\sigma^2\mu_2^{p+1}} + \frac{p^2\beta_1}{\mu_2^3}\right] \log\left(\frac{x_3}{x_2}\right) = \frac{p^2}{\mu_2^2} + \frac{2}{\sigma^2\mu_2^p}$$

with $\mu_2 = \beta_1 + \beta_2 x_2^{\beta_3}$.

Unfortunately, when σ^2 would be unknown, the D-optimal design is not determined by determinant $I_{3\times3}(\boldsymbol{\beta})$ anymore, even though condition (5) is still satisfied. Indeed, the variance of the score function with respect to σ is independent of $\boldsymbol{\beta}$, since $\mathsf{VAR}(\ell'_\sigma) = 0.5n/\sigma^4$ (Appendix C). The issue is that the covariances between the score functions with respect to $\boldsymbol{\beta}$ and the score function with respect to σ are no longer equal to zero (Appendix C).

4.3. Transformations of the Mitscherlich Function

Our results provide D-optimal designs for $\mathbb{E}(y_{ij}|x_i) = \beta_1 + \beta_2 x_i^{\beta_3}$, with $\beta_1 \geq 0$, $\beta_2, \beta_3 > 0$, $x_i \geq 0$, and y_{ij} having a distribution in the canonical exponential family with $\theta_i = g(\beta_1 + \beta_2 x_i^{\beta_3})$. If we wish to study the non-linear function $\mathbb{E}(y_{ij}|x_i) = \psi(\beta_1 + \beta_2 x_i^{\beta_3})$, with canonical link function g, the log-likelihood function in (2) and determinant $|I_{3\times 3}(\boldsymbol{\beta})|$ in (7) both change due to the transformation ψ. The determinant becomes

$$\beta_2^2 \left[M(\mathbf{x}|\boldsymbol{\beta})\right]^2 \prod_{i=1}^{3} \left[n_i g'(\psi(\mu_i)) \left\{\psi'(\mu_i)\right\}^2\right],$$

with $M(\mathbf{x}|\boldsymbol{\beta}) = (x_1x_2)^{\beta_3}\log(x_2/x_1) - (x_1x_3)^{\beta_3}\log(x_3/x_1) + (x_2x_3)^{\beta_3}\log(x_3/x_2)$, $\mu_i = \beta_1 + \beta_2 x_i^{\beta_3}$, and ψ' the derivative of ψ. If we now define the function $\tilde{g}$ through its derivative $\tilde{g}'(\mu) = g'(\psi(\mu))[\psi'(\mu)]^2$, the second derivative of function $\tilde{g}$ would become $\tilde{g}''(\mu) = g''(\psi(\mu))[\psi'(\mu)]^3 + 2g'(\psi(\mu))\psi''(\mu)\psi'(\mu)$. Based on the proof of Theorem 1, x_1^{opt} should be chosen equal to L when $\tilde{g}'(\mu) = g'(\psi(\mu))[\psi'(\mu)]^2$ is non-negative and decreasing in μ (i.e., $g''(\psi(\mu))[\psi'(\mu)]^3 + 2g'(\psi(\mu))\psi''(\mu)\psi'(\mu) \leq 0$). If we also have that $\mu\tilde{g}''(\mu) + 2\tilde{g}'(\mu) \geq 0$, x_3^{opt} should be chosen equal to U. The optimal stimuli x_2^{opt} should satisfy Eq. (8) with $g'(\mu_2)$ and $g''(\mu_2)$ replaced by $\tilde{g}'(\mu_2)$ and $\tilde{g}''(\mu_2)$, respectively.

To illustrate these results, let's assume we would like to study the square root transformation $\psi(\mu) = \sqrt{\mu}$ of the Mitscherlich function $\mu = \beta_1 + \beta_2 x^{\beta_3}$ and assume that the canonical link function is equal to $g(\mu) = \log(\mu)$, with $\mu > 0$. Then the function $\tilde{g}'(\mu) = 0.25\mu^{-3/2}$ is a positive decreasing function and $\tilde{g}''(\mu) = -0.375\mu^{-5/2}$ is negative for all $\mu > 0$. Furthermore, the condition $\mu\tilde{g}''(\mu) + 2\tilde{g}'(\mu)$ is equal to $0.125\mu^{-3/2}$ and positive for all $\mu > 0$. Thus the combination $g(\mu) = \log(\mu)$ and $\psi(\mu) = \sqrt{\mu}$ results into $x_1^{\text{opt}} = L$, $x_3^{\text{opt}} = U$, and x_2^{opt} can be obtained from

$$-0.75\beta_2\beta_3 \left[(Lx_2)^{\beta_3}\log\left(fracx_2L\right) - (LU)^{\beta_3}\log\left(\frac{U}{L}\right) + (x_2U)^{\beta_3}\log\left(\frac{U}{x_2}\right)\right]$$
$$+\mu_2\left[\beta_3 L^{\beta_3}\log\left(\frac{x_2}{L}\right) + \beta_3 U^{\beta_3}\log\left(\frac{U}{x_2}\right) + L^{\beta_3} - U^{\beta_3}\right] = 0.$$

On the other hand, when we would like to study the exponential transformation $\psi(\mu) = \exp\{\mu\}$ and the canonical link function is still $g(\mu) = \log(\mu)$, with $\mu > 0$, we do not satisfy the criteria. The function $\tilde{g}'(\mu) = \exp\{\mu\}$ is still positive, but it is clearly not a decreasing function, since the derivative $\tilde{g}''(\mu) = \exp\{\mu\}$ is always positive. Thus D-optimal designs for the combination $\psi(\mu) = \exp\{\mu\}$ and $g(\mu) = \log(\mu)$ may be different from what Theorems 1–3 seem to indicate. Thus our theorems only apply to certain combinations.

5. SUMMARY AND DISCUSSION

In this paper, we determined D-optimal designs for responses having their distribution in the exponential family and their mean equal to the three-parameter Mitscherlich non-linear function, Transformations of the Mitscherlich function are possible too, but only under certain conditions. The D-optimal criterion is independent of estimation of the dispersion parameter if the precision of the estimation of the dispersion parameter is independent of the parameters of the Mitscherlich function, a condition that holds for all well-known distribution functions. It would be interesting to know if there exists an example within our formulation of the exponential family for which this condition does not hold.

It was demonstrated that the canonical link function plays an important role in selecting the optimal values of the three stimuli. For most distribution functions we should choose the first stimulus as small as possible and the third stimulus as large as possible, but for the inverse Gaussian distribution the third stimulus can be substantially smaller than the maximum allowable stimulus. The middle stimulus depends on the parameters, the optimal first and third stimulus, and the canonical link function. For the Binomial distribution, the conditions on the canonical (logit) link function in our theory may also not always be satisfied. Hence, the canonical link function has a strong effect on how to choose the stimuli (as we illustrated).

We showed that our results are an extension of earlier results [1], and that their approach of linearization of the Mitscherlich function can be extended easily by including weights. Our D-optimality criterion is identical to the D-optimality criterion of a weighted linear regression problem, where the weights are the traditional inverse variances of the response at the selected stimuli. However, when the residual variance would be heterogeneous, linearization of the Mitscherlich function does not lead to a D-optimal design anymore.

We believe that the Mitscherlich non-linear function has not received enough attention, while we believe it is a very useful stimulus-response function for validation studies of measurement systems. The Mitscherlich function is an extension of the two-parameter linear or log-linear regression function and therefore useful to investigate linearity of the system. Our results may help formulate an optimal design for maximizing the precision of the estimators of the parameters of the Mitscherlich function and then evaluate linearity of the measurement system.

Altough we were able to formulate D-optimal designs for the Mitscherlich function with a large class of probability distributions, testing for linearity may require alternative optimal designs, since two different models are being compared that would not have the same D-optimal design. Secondly, it would be of interest to determine the optimal settings in case more than three stimuli are being selected, for instance to test the goodness-of-fit of the Mitscherlich function. Thirdly, more work is needed to understand the optimal designs for transformations that may not satisfy our conditions. Finally, we believe that our work may be extended to other non-linear functions that have similar characteristics as the Mitscherlich function.

Appendix A

Here we will determine the Fisher information matrix for the parameters $\boldsymbol{\beta}$, using the second derivatives of the log likelihood function. An explicit formula for determinant $|I_{3\times 3}(\boldsymbol{\beta})|$ in (7) can then be determined using for instance Matlab.

The second derivative $\partial^2\ell(\boldsymbol{\beta},\phi|\mathbf{y})/(\partial\beta_k)^2$ of the log likelihood function in (2) with respect to β_k is given by

$$-\frac{1}{a(\phi)}\sum_{i=1}^{m}\left[n_i g'(\mu_i)\left(\frac{\partial\mu_i}{\partial\beta_k}\right)^2-\left\{g''(\mu_i)\left(\frac{\partial\mu_i}{\partial\beta_k}\right)^2+g'(\mu_i)\frac{\partial^2\mu_i}{(\partial\beta_k)^2}\right\}(y_{i.}-n_i\mu_i)\right].$$

Using [16, Theorem 1.1, p. 406], we obtain that the variance of the score function ℓ'_{β_k} is equal to $\mathsf{VAR}(\ell'_{\beta_k})=-\mathbb{E}[\partial^2\ell(\boldsymbol{\beta},\phi|\mathbf{y})/(\partial\beta_k)^2]$. This leads to the first variance in (4), since $\mathbb{E}[y_{i.}-n_i\mu_i]=0$.

The second derivative $\partial^2\ell(\boldsymbol{\beta},\phi|\mathbf{y})/(\partial\phi)^2$ of the log likelihood function in (2) with respect to ϕ is given by

$$-\sum_{i=1}^{m}\sum_{j=1}^{n_i}\left[\tfrac{a''(\phi)a(\phi)-2[a'(\phi)]^2}{a^3(\phi)}\left(y_{ij}g(\mu_i)-b(g(\mu_i))\right)-c''(y_{ij},\phi)\right],$$

with $c''(\cdot,\cdot)$ the second derivative with respect to the second argument. Using again [16, Theorem 1.1, p. 406], the variance of the score function ℓ'_ϕ is equal to $\mathsf{VAR}(\ell'_\phi)=-\mathbb{E}[\partial^2\ell(\boldsymbol{\beta},\phi|\mathbf{y})/(\partial\phi)^2]$. Since $\mathbb{E}c'(y_{ij},\phi)=a'(\phi)\mathbb{E}[y_{ij}g(\mu_i)-b(g(\mu_i))]/a^2(\phi)$, with $c'(y_{ij},\phi)=\partial c(y_{ij},\phi)/\partial\phi$, the variance $\mathsf{VAR}(\ell'_\phi)$ becomes equal to the second equation in (4).

The second derivative $\partial^2\ell(\boldsymbol{\beta},\phi|\mathbf{y})/(\partial\beta_r\partial\beta_s)$ of the log likelihood function in (2) with respect to β_r and β_s, with $r\neq s$, is given by

$$\frac{1}{a(\phi)}\sum_{i=1}^{m}\left[-n_i g'(\mu_i)\frac{\partial\mu_i}{\partial\beta_r}\frac{\partial\mu_i}{\partial\beta_s}+(y_{i.}-n_i\mu_i)\left(g''(\mu_i)\frac{\partial\mu_i}{\partial\beta_r}\frac{\partial\mu_i}{\partial\beta_s}+g'(\mu_i)\frac{\partial^2\mu_i}{\partial\beta_r\partial\beta_s}\right)\right].$$

Using the same [16 , Theorem 1.1, p. 406], the covariance of the score functions ℓ'_{β_r} and ℓ'_{β_s} is equal to $\mathsf{COV}(\ell'_{\beta_r}, \ell'_{\beta_s}) = -\mathbb{E}[\partial^2 \ell(\boldsymbol{\beta}, \phi|\mathbf{y})/(\partial\beta_r \partial\beta_s)]$. Using again that $\mathbb{E}[y_{i.} - n_i\mu_i] = 0$, the covariance $\mathsf{COV}(\ell'_{\beta_r}, \ell'_{\beta_s})$ becomes equal to the third equation in (4).

The second derivative $\partial^2\ell(\boldsymbol{\beta}, \phi|\mathbf{y})/(\partial\beta_k\partial\phi)$ of the log likelihood function in (2) with respect to β_k and ϕ is given by

$$\frac{-a'(\phi)}{a^2(\phi)} \sum_{i=1}^{m} \left[(y_{i.} - n_i\mu_i)\, g'(\mu_i) \frac{\partial\mu_i}{\partial\beta_k} \right].$$

Using [16 , Theorem 1.1, p. 406], the covariance of the score functions ℓ'_{β_k} and ℓ'_{ϕ} is equal to $\mathsf{COV}(\ell'_{\beta_k}, \ell'_{\phi}) = -\mathbb{E}[\partial^2\ell(\boldsymbol{\beta}, \phi|\mathbf{y})/(\partial\beta_k\partial\phi)]$. Since $\mathbb{E}y_{i.} = n_i\mu_i$, the covariance $\mathsf{COV}(\ell'_{\beta_k}, \ell'_{\phi})$ becomes equal to zero.

Appendix B

Here, we will provide the proofs of Theorems 1–3.

Proof of Theorem 1

We may assume that $n_1 = n_2 = n_3 = 1$ without loss of generality and introduce $z_i = x_i^{\beta_3}$ with $\beta_3 > 0$. The determinant $|I_{3\times3}(\boldsymbol{\beta})|$ can now be written as

$$|I_{3\times3}(\boldsymbol{\beta})| = \beta_3^{-2}\beta_2^2 g'(\beta_1 + \beta_2 z_1) g'(\beta_1 + \beta_2 z_2) g'(\beta_1 + \beta_2 z_3) [h(z_1|z_2, z_3)]^2,$$

with function $h(z_1|z_2, z_3) = z_1 z_2 \log(\frac{z_2}{z_1}) - z_1 z_3 \log(\frac{z_3}{z_1}) + z_2 z_3 \log(\frac{z_3}{z_2})$. We will demonstrate that the determinant $|I_{3\times3}(\boldsymbol{\beta})|$ is a decreasing function of $z_1 \in [0, z_2)$ for any value of $z_2 > 0$ and $z_3 > z_2$, which proves that we should choose x_1 as small as possible.

Since we assumed that $g''(\mu) \leq 0$, the function $g'(\mu)$ is a non-increasing function in μ. Since μ_1 is an increasing function in z_1 ($\beta_2 > 0$), we have demonstrated that $g'(\mu_1)$ is a non-increasing function in z_1 and thus in x_1. We will now demonstrate that $h(z_1|z_2, z_3)$ is decreasing in z_1 by showing that $\partial h(z_1|z_2, z_3)/\partial z_1 \leq 0$ for all $z_1 \in [0, z_2)$.

The derivative of $h(z_1|z_2, z_3)$ with respect to z_1 is given by

$$h'(z_1|z_2, z_3) = (z_3 - z_2)\log(z_1) + z_2(\log(z_2) - 1) - z_3(\log(z_3) - 1),$$

which is an increasing function in z_1. It is clear that for z_1 approaching 0 from above, $\lim_{z_1\downarrow 0} h'(z_1|z_2, z_3) = -\infty$, that $h'(z_1|z_2, z_3) = 0$ at $z_1 = z_1^0$ with

$$\log(z_1^0) = \frac{z_3(\log(z_3) - 1) - z_2(\log(z_2) - 1)}{z_3 - z_2},$$

and that $h'(z_1|z_2, z_3)$ is negative for $z_1 \in [0, z_1^0)$. If we can show that $\log(z_1^0) \geq \log(z_2)$, then we can conclude that $h(z_1|z_2, z_3)$ is a decreasing function on the interval $[0, z_2)$. Inequality $\log(z_1^0) \geq \log(z_2)$ is identical to $z_3(\log(z_3) - \log(z_2)) \geq z_3 - z_2$, using standard algebra. If we now choose $z_2 = az_3$, with $0 < a < 1$, inequality $\log(z_1^0) \geq \log(z_2)$ results in inequality $a - \log(a) \geq 1$. Since $a - \log(a)$ is a decreasing function for $a \in (0, 1)$ and $a - \log(a)$ is equal to one for $a = 1$, we have demonstrated that $z_1^0 > z_2$ and that $h(z_1|z_2, z_3)$ is a decreasing function in z_1.

Furthermore, $h(z_2|z_2, z_3)$ is equal to zero, which means that $h(z_1|z_2, z_3) > 0$ for $z_1 \in [0, z_2)$ and hence $[h(z_1|z_2, z_3)]^2$ is a decreasing function in z_1 on the interval $[0, z_2)$. This implies that $g'(\beta_1 + \beta_2 z_1)[h(z_1|z_2, z_3)]^2$ is a decreasing function in z_1 on the interval $[0, z_2)$ and thus $|I_{3\times3}(\boldsymbol{\beta})|$ is a decreasing function in x_1 on the interval $[0, x_2)$.

Proof of Theorem 2

Again we assume that $n_1 = n_2 = n_3 = 1$ and introduce $z_i = x_i^{\beta_3}$ with $\beta_3 > 0$. The determinant $|I_{3\times3}(\boldsymbol{\beta})|$ can now be written as

$$|I_{3\times3}(\boldsymbol{\beta})| = \beta_3^{-2}\beta_2^2 g'(\beta_1+\beta_2 z_1)g'(\beta_1+\beta_2 z_2)g'(\beta_1+\beta_2 z_3)[h(z_3|z_1,z_2)]^2,$$

with function $h(z_3|z_1,z_2) = z_1z_2\log(\frac{z_2}{z_1}) - z_1z_3\log(\frac{z_3}{z_1}) + z_2z_3\log(\frac{z_3}{z_2})$. We will demonstrate that determinant $|I_{3\times3}(\boldsymbol{\beta})|$ is an increasing function in $z_3 \in (z_2,\infty)$, under the stated conditions of Theorem 2 for any value of $z_1 \geq 0$ and $z_2 > z_1$, which proves that we should choose x_3 as large as possible. To do this, we may just study the product function $D(z_3) = g'(\beta_1+\beta_2 z_3)[h(z_3|z_1,z_2)]^2$, since all other elements in $|I_{3\times3}(\boldsymbol{\beta})|$ are positive constants with respect to z_3.

If we denote $h'(z_3|z_1,z_2)$ as the derivative of $h(z_3|z_1,z_2)$ with respect to z_3 and define $C(\mu) = g''(\mu)\mu + 2g'(\mu)$, the derivative of $D(z_3)$ with respect to z_3 can be written as

$$\begin{aligned}\frac{\partial D(z_3)}{\partial z_3} &= \beta_2 g''(\mu_3)\,[h(z_3|z_1,z_2)]^2 + 2g'(\mu_3)h(z_3|z_1,z_2)h'(z_3|z_1,z_2)\\ &= \frac{[h(z_3|z_1,z_2)]^2}{z_3}\left[g''(\mu_3)\beta_2 z_3 + 2g'(\mu_3)\frac{z_3h'(z_3|z_1,z_2)}{h(z_3|z_1,z_2)}\right]\\ &= \frac{[h(z_3|z_1,z_2)]^2}{z_3}\left[g''(\mu_3)\mu_3 - \beta_1 g''(\mu_3) + 2g'(\mu_3)\frac{z_3h'(z_3|z_1,z_2)}{h(z_3|z_1,z_2)}\right]\\ &= \frac{[h(z_3|z_1,z_2)]^2}{z_3}\left[C(\mu_3) - \beta_1 g''(\mu_3) + 2g'(\mu_3)\left(\frac{z_3h'(z_3|z_1,z_2)}{h(z_3|z_1,z_2)} - 1\right)\right].\end{aligned} \tag{19}$$

Based on the assumptions of Theorem 2, we have that $C(\mu_3) \geq 0$, $-\beta_1 g''(\mu_3) \geq 0$, and $2g'(\mu_3) \geq 0$. Furthermore, the term $[h(z_3|z_1,z_2)]^2/z_3$ is non-negative for all z_3. If we can now demonstrate that the term $z_3h'(z_3|z_1,z_2)/h(z_3|z_1,z_2) \geq 1$, we have demonstrated that the derivative $\partial D(z_3)/\partial z_3$ is non-negative, indicating that $D(z_3)$ is increasing.

If we rewrite $h(z_3|z_1,z_2)$ into $h(z_3|z_1,z_2) = A_1 + A_2z_3\log(z_3) + A_3z_3$, with $A_1 = z_1z_2\log(\frac{z_2}{z_1})$, $A_2 = z_2 - z_1$, and $A_3 = z_1\log(z_1) - z_2\log(z_2)$, the derivative of $h(z_3|z_1,z_2)$ with respect to z_3 is given by $h'(z_3|z_1,z_2) = A_2 + A_2\log(z_3) + A_3$. Since $A_2 > 0$, $h'(z_3|z_1,z_2)$ is an increasing function with $\lim_{z_3\to\infty} h'(z_3|z_1,z_2) = \infty$. The function $h'(z_3|z_1,z_2)$ is equal to zero in $z_3 = z_3^0$, with

$$\log(z_3^0) = -\frac{A_3 + A_2}{A_2} = \frac{z_2\log(z_2) - z_1\log(z_1)}{z_2 - z_1} - 1.$$

If we can demonstrate that $\log(z_3^0) \leq \log(z_2)$, we would obtain that $h'(z_3|z_1,z_2) > 0$ for $z_3 \in (z_2,\infty)$, and thus $h(z_3|z_1,z_2)$ is an increasing function. Since $h(z_2|z_1,z_2) = 0$, $h(z_3|z_1,z_2)$ would also be positive on (z_2,∞).

If we now choose $z_1 = az_2$, with $a \in [0,1)$, the solution $\log(z_3^0)$ is equal to $[\log(z_2) - a\log(z_2) - a\log(a) - 1 + a]/[1-a]$. Then inequality $\log(z_3^0) \leq \log(z_2)$ results in $a - a\log(a) \leq 1$. The function $a - a\log(a)$ is an increasing function in $a \in [0,1)$, since its derivative is equal to $-\log(a)$, and it is equal to one when $a = 1$. Thus inequality $\log(z_3^0) \leq \log(z_2)$ is guaranteed.

Knowing that $h(z_3|z_1,z_2) > 0$ for $z_3 \in (z_2,\infty)$, we can see that inequality $z_3h'(z_3|z_1,z_2)/h(z_3|z_1,z_2) \geq 1$ is equivalent to the following inequality

$$\begin{aligned} z_3h'(z_3|z_1,z_2)/h(z_3|z_1,z_2) \geq 1 &\iff A_2z_3 - A_1 \geq 0\\ &\iff z_3 \geq z_1z_2\log(z_2/z_1)/[z_2 - z_1].\end{aligned}$$

If we can prove that $z_1z_2\log(z_2/z_1)/[z_2-z_1]$ is smaller than or equal to z_2, we have demonstrated that $z_3h'(z_3|z_1,z_2)/h(z_3|z_1,z_2) \geq 1$ holds for $z_3 \in (z_2,\infty)$. If we again take $z_1 = az_2$, with $a \in [0,1)$, we obtain that $z_1z_2\log(z_2/z_1)/[z_2-z_1] = -az_2\log(a)/[1-a]$ and this function is smaller or equal to z_2 when $-a\log(a)/[1-a] \leq 1$. This results again in $a - a\log(a) \leq 1$, which we already demonstrated to be true.

Thus we have finally shown that $\partial D(z_3)/\partial z_3 > 0$ under the stated conditions of Theorem 2, making the determinant $|I_{3\times3}(\boldsymbol{\beta})|$ an increasing function in z_3 on the interval (z_2,∞).

Proof of Theorem 3

We start again with the assumption that $n_1 = n_2 = n_3 = 1$ and we introduce $z_i = x_i^{\beta_3}$ with $\beta_3 > 0$. The determinant $|I_{3\times 3}(\boldsymbol{\beta})|$ can now be written as

$$|I_{3\times 3}(\boldsymbol{\beta})| = \beta_3^{-2}\beta_2^2 g'(\beta_1 + \beta_2 z_1)g'(\beta_1 + \beta_2 z_2)g'(\beta_1 + \beta_2 z_3)[h(z_2|z_1, z_3)]^2,$$

with function $h(z_2|z_1, z_3) = z_1 z_2 \log(\frac{z_2}{z_1}) - z_1 z_3 \log(\frac{z_3}{z_1}) + z_2 z_3 \log(\frac{z_3}{z_2})$. We will now study $D(z_2) = g'(\beta_1 + \beta_2 z_2)[h(z_2|z_1, z_3)]^2$ as function of z_2 in the interval (z_1, z_3), since the remaining part of the determinant is just a constant.

Function $h(z_2|z_1, z_3)$ is rewritten into $h(z_2|z_1, z_3) = -A_1 z_2 \log(z_2) + A_2 z_2 - A_3$, with $A_1 = z_3 - z_1$, $A_2 = z_3 \log(z_3) - z_1 \log(z_1)$, and $A_3 = z_1 z_3 \log(z_3/z_1)$. The terms A_1 and A_3 are both positive when $z_3 > z_1 \geq 0$, but the term A_2 is only positive when we may assume that $z_3 \geq 1$. Note that $h(z_1|z_1, z_3) = h(z_3|z_1, z_3) = 0$ and the derivative $h'(z_2|z_1, z_3) = \partial h(z_2|z_1, z_3)/\partial z_2$ is equal to $A_2 - A_1 - A_1 \log(z_2)$, which is a decreasing function in z_2. The solution z_2^0 of $h'(z_2|z_1, z_3) = 0$ is unique and satisfies $\log(z_2^0) = [A_2 - A_1]/A_1$. If we can demonstrate that $\log(z_1) < \log(z_2^0) < \log(z_3)$, we would know that $h(z_2|z_1, z_3)$ is increasing on interval $z_2 \in (z_1, z_2^0)$ and decreasing on interval $z_2 \in (z_2^0, z_3)$ and thus always positive on $z_2 \in (z_1, z_3)$.

Inequality $\log(z_1) < \log(z_2^0)$ is equivalent to inequality $z_3 - z_1 < z_3[\log(z_3) - \log(z_1)]$. If we choose $z_1 = az_3$, with $a \in [0, 1)$, the inequality becomes $1 - a + \log(a) < 0$, which holds for $a \in [0, 1)$, since $1 - a + \log(a)$ is an increasing function on interval $a \in [0, 1)$ with $\lim_{a\to 1}[1 - a + \log(a)] = 0$. Inequality $\log(z_2^0) < \log(z_3)$ is equivalent to inequality $z_1[\log(z_3) - \log(z_1)] < (z_3 - z_1)$. If we choose $z_1 = az_3$, with $a \in [0, 1)$, the inequality becomes $1 - a + a\log(a) > 0$, which holds for $a \in [0, 1)$, since $1 - a + a\log(a)$ is a decreasing function on interval $a \in [0, 1)$ with $\lim_{a\to 1}[1 - a + a\log(a)] = 0$. Thus this proves $\log(z_1) < \log(z_2^0) < \log(z_3)$.

Maximizing $D(z_2)$ in z_2, can be done by setting the derivative equal to zero. Since $h(z_2|z_1, z_3)$ is positive on $z_2 \in (z_1, z_3)$, setting the derivative $\partial D(z_2)/\partial z_2$ equal to zero leads to the following equality

$$\beta_2 g''(\mu_2)h(z_2|z_1, z_3) + 2g'(\mu_2)h'(z_2|z_1, z_3) = 0, \tag{20}$$

which is identical to Eq. (8). We now need to demonstrate that Eq. (20) has at least one solution for a value of $z_2 \in (z_1, z_3)$. Rewriting Eq. (20), leads to $\beta_2 g''(\mu_2)/g'(\mu_2) = -2h'(z_2|z_1, z_3)/h(z_2|z_1, z_3)$. The left-hand side is smaller or equal to zero for any z_2, while the right-hand side is negative for $z_2 \in (z_1, z_2^0)$, zero at $z_2 = z_2^0$, and positive for $z_2 \in (z_2^0, z_3)$, using the results on $h'(z_2|z_1, z_3)$ and $h(z_2|z_1, z_3)$ above. Since we also have that $\lim_{z_2\downarrow z_1} -2h'(z_2|z_1, z_3)/h(z_2|z_1, z_3) = -\infty$ when z_2 goes down to z_1, we now know that a solution must occur for $z_2 \in (z_1, z_2^0]$, thereby satisfying the constraint in Theorem 3.

Appendix C

Here, we will assume that $y_{ij} \sim N(\mu_i, \sigma^2\varphi(\mu_i))$ is normally distributed with a mean μ_i equal to the Mitscherlich function $\mathbb{E}(y_{ij}|x_i) \equiv \mu_i = \beta_1 + \beta_2 x_i^{\beta_3}$, and with $\beta_i > 0$. We will provide the elements of the Fisher information matrix for estimation of $\boldsymbol{\theta}^T = (\beta_1, \beta_2, \beta_3, \sigma^2)$. The log likelihood function is given by

$$\ell(\boldsymbol{\theta}|\mathbf{y}) = -\frac{1}{2}\sum_{i=1}^{m}\sum_{j=1}^{n_i}\left[\log(2\pi) + \log(\sigma^2) + \log(\varphi(\mu_i)) + (y_{ij} - \mu_i)^2/(\sigma^2\varphi(\mu_i))\right]$$

with $\mathbf{y} = (\mathbf{y}_1, \mathbf{y}_2, ..., \mathbf{y}_m)^T$ and $\mathbf{y}_i = (y_{i1}, y_{i2}, ..., y_{in_i})^T$. The four score functions $\ell'_{\beta_k} = \partial\ell(\boldsymbol{\theta}|\mathbf{y})/\partial\beta_k$, $k \in \{1, 2, 3\}$, and $\ell'_{\sigma^2} = \partial\ell(\boldsymbol{\theta}|\mathbf{y})/\partial(\sigma^2)$ are now given by

$$\ell'_{\beta_k} = -\frac{1}{2}\sum_{i=1}^{m}\sum_{j=1}^{n_i}\left[\frac{\varphi'(\mu_i)}{\varphi(\mu_i)} - 2\frac{y_{ij} - \mu_i}{\sigma^2\varphi(\mu_i)} - \frac{\varphi'(\mu_i)(y_{ij} - \mu_i)^2}{[\sigma\varphi(\mu_i)]^2}\right]\frac{\partial\mu_i}{\partial\beta_k},$$

$$\ell'_{\sigma^2} = -\frac{1}{2}\sum_{i=1}^{m}\sum_{j=1}^{n_i}\left[\frac{1}{\sigma^2} - \frac{(y_{ij}-\mu_i)^2}{\sigma^4\varphi(\mu_i)}\right].$$

After tedious algebraic calculations using the score functions directly, the variances and covariances of the score functions can be calculated. They are given by

$$\mathsf{VAR}(\ell'_{\beta_k}) = \sum_{i=1}^{m} n_i\left[\frac{1}{2}\left(\frac{\varphi'(\mu_i)}{\varphi(\mu_i)}\right)^2 + \frac{1}{\sigma^2\varphi(\mu_i)}\right]\left(\frac{\partial\mu_i}{\partial\beta_k}\right)^2,$$

$$\mathsf{VAR}(\ell'_{\sigma^2}) = \frac{1}{2\sigma^4}\sum_{i=1}^{m} n_i,$$

$$\mathsf{COV}(\ell'_{\beta_r}, \ell'_{\beta_s}) = \sum_{i=1}^{m} n_i\left[\frac{1}{2}\left(\frac{\varphi'(\mu_i)}{\varphi(\mu_i)}\right)^2 + \frac{1}{\sigma^2\varphi(\mu_i)}\right]\frac{\partial\mu_i}{\partial\beta_r}\frac{\partial\mu_i}{\partial\beta_s},$$

$$\mathsf{COV}(\ell'_{\beta_k}, \ell'_{\sigma^2}) = \frac{1}{2\sigma^2}\sum_{i=1}^{m} n_i\left[\frac{\varphi'(\mu_i)}{\varphi(\mu_i)}\right]\frac{\partial\mu_i}{\partial\beta_k}.$$

Since we have $\partial\mu_i/\partial\beta_1 = 1$, $\partial\mu_i/\partial\beta_2 = x_i^{\beta_3}$, and $\partial\mu_i/\partial\beta_3 = \beta_2 x_i^{\beta_3}\log(x_i)$, we obtain that the matrix $I_{3\times 3}(\boldsymbol{\beta})$ is given by (18).

FUNDING

This work is part of the research program Rapid Micro Statistics with project no. 15990, which is (partly) financed by the Netherlands Organization for Scientific Research (NWO). The authors gratefully acknowledge the support of the user committee and the funding organization.

REFERENCES

1. G. E. Box and H. L. Lucas, “Design of experiments in non-linear situations,” Biometrika **46** (1/2), 77–90 (1959).
2. D. R. Clark and C. A. Thayer, “A primer on the exponential family of distributions,” Casualty Actuarial Society Spring Forum, 117–148 (2004).
3. G. M. Cordeiro and P. McCullagh, “Bias correction in generalized linear models,” J. R. Stat. Soc. Series B Stat. Methodol. **53** (3), 629–643 (1991).
4. M. Davidian and D. M. Giltinan, “Nonlinear models for repeated measurement data: an overview and update,” J. Agric. Biol. Environ. Stat. **8** (4), 387–419 (2003).
5. H. Dette and S. Biedermann, “Robust and efficient designs for Michaelis–Menten model,” J. Amer. Statist. Assoc. **98** (463), 679–686 (2003).
6. H. Dette, C. Kiss, M. Bevanda, and F. Bretz, “Optimal designs for the EMAX, log-linear and exponential models,” Biometrika **97** (2), 513–518 (2010).
7. H. Dette and W. K. Wong, “Optimal designs when the variance is a function of the mean,” Biometrics **55** (3), 925–929 (1999).
8. V. V. Fedorov and S. L. Leonov, “Optimal design of dose response experiments: a model-oriented approach,” Drug Inf. J. **35** (4), 1373–1383 (2001).
9. V. V. Fedorov and S. L. Leonov, *Optimal Design for Nonlinear Response Models* (Chapman and Hall/CRC, New York, 2013).
10. I. Ford, D. M. Titterington, and C. P. Kitsos, “Recent advances in nonlinear experimental design,” Technometrics **31** (1), 49–60 (1989).
11. I. Ford, B. Torsney, and C. J. Wu, “The use of a canonical form in the construction of locally optimal designs for non-linear problems,” J. R. Stat. Soc. Series B Stat. Methodol. **54** (2), 569–583 (1992).
12. C. Han and K. Chaloner, “D- and c-optimal designs for exponential regression models used in viral dynamics and other applications,” J. Statist. Plann. Inference **115** (2), 585–601 (2003).
13. A. S. Hedayat, B. Yan, and J. M. Pezzuto, “Modeling and identifying optimum designs for fitting dose-response curves based on raw optical density data,” J. Amer. Statist. Assoc. **92** (439), 1132–1140 (1997).
14. R. S. John and N. R. Draper, “D-optimality for regression designs: a review,” Technometrics **17** (1), 15–23 (1975).

15. A. I. Khuri, “A note on D-optimal designs for partially nonlinear regression models,” Technometrics **26** (1), 59–61 (1984).
16. E. L. Lehmann, *Theory of Point Estimation* (John Wiley, New York, 1983).
17. G. Li and N. Balakrishnan, “Optimal designs for tumor regrowth models,” J. Statist. Plann. Inference **141** (2), 644–654 (2011).
18. G. Li and D. Majumdar, “D-optimal designs for logistic models with three and four parameters,” J. Statist. Plann. Inference **138** (7), 1950–1959 (2008).
19. G. Li and D. Majumdar, “Some results on D-optimal designs for nonlinear models with applications,” Biometrika **96** (2), 487–493 (2009).
20. A. Maloney, U. S. Simonsson, and M. Schaddelee, “D-optimal designs for three poisson dose-response models,” J. Pharmacokinet. Pharmacodyn. **40** (2), 201–211 (2013).
21. P. McCullagh and J. A. Nelder, *Generalized Linear Models* (Chapman and Hall, 2nd ed., London, 1989).
22. S. Minkin, “Experimental design for clonogenic assays in chemotherapy,” J. Amer. Statist. Assoc. **88** (422), 410–420 (1993).
23. E. A. Mitscherlich, “The law of the minimum and the law of diminishing soil productivity,” Landwirtschaftliche Jahrbuecher **38**, 537–552 (1909).
24. D. A. Ratkowsky and G. R. Dolby, “Taylor series linearization and scoring for parameters in nonlinear regression,” J. R. Stat. Soc. Ser. C Appl. Stat. **24** (1), 109–111 (1975).
25. C. RodrMguez-Torreblanca and J. M. RodrMguez-DMaz, “Locally D- and c-optimal designs for poisson and negative binomial regression models,” Metrika **66** (2), 161–172 (2007).
26. K. G. Russell, D. C. Woods, S. M. Lewis, and J. A. Eccleston, “D-optimal designs for poisson regression models,” Statistica Sinica **19**, 721–730 (2009).
27. A. Vølund, “Application of the four-parameter logistic model to bioassay: comparison with slope ratio and parallel line model,” Biometrics **34**, 357–365 (1978).
28. Y. Wang, R. H. Myers, E. P. Smith, and K. Ye, “D-optimal designs for poisson regression models,” J. Statist. Plann. Inference **136** (8), 2831–2845 (2006).
29. B. C. Wei, *Exponential Family Nonlinear Models* (Springer, Singapore, 1998).