Brazilian Journal of Probability and Statistics
2014, Vol. 28, No. 4, 483–491
DOI: 10.1214/13-BJPS220

On Bayesian D-optimal design criteria and the General Equivalence Theorem in joint generalized linear models for the mean and dispersion

Edmilson Rodrigues Pinto[a] and Antônio Ponce de Leon

[a]*Federal University of Uberlândia*
[b]*Rio de Janeiro State University*

Abstract. The joint modeling of mean and dispersion has been used to model many problems in statistics, especially in industry, where not only the mean of response, but also the dispersion depends on the covariates. In scientific research, one of the crucial points is the experimental design, which when properly implemented, will create a reliable structure, essential to improve the statistical inference and for the development of the next phases of the experimental process. The theory of optimal design of experiments is a powerful and flexible approach to generate efficient experimental designs. In the context of optimal designs, the General Equivalence Theorem plays a fundamental role, because it permits to check if a design found is optimal. In this article, we investigated the validity of the General Equivalence Theorem for obtaining Bayesian D and D_S optimal designs in joint generalized linear models for the mean and dispersion.

1 Introduction

In many real problems, not only the mean but also, in the more general case, the dispersion depends on a set of explanatory variables. Thus, the analysis of such problems requires the joint modeling of mean and dispersion (JMMD) that usually can be performed by using two interlinked generalized linear models. Joint generalized linear models of the mean and dispersion were introduced by Nelder and Lee (1991) as an alternative to Taguchi's methods in quality-improvement experiment. Further examples appeared in Lee and Nelder (1998) and Lee and Nelder (2003).

As well as modeling, experimentation plays an important part in the scientific method. For experimental data, proper estimation of model parameters depends on well-designed experiments. When conducting an experiment there are many design issues to consider, including deciding which treatments to study, which factors to control, what aspects of an experiment to randomize, how many experimental units are needed, how many replications should be allocated to each treatment, or

Key words and phrases. General Equivalence Theorem, D and D_S-optimal designs, Bayesian designs, joint modeling of mean and dispersion.
Received November 2011; accepted January 2013.

what levels of other input or control variables should be used. Limitations due to costs, time, ethics or human resources are such that sample sizes are usually restricted, therefore, efficient use of available resources is critical. The purpose of optimal experimental design is to improve statistical inference regarding the quantities of interest by optimally selecting the combinations of levels of design factors under the control of the investigator, within, of course, the constraints of available resources (Clyde, 2001).

When we consider obtaining optimal designs for generalized linear or nonlinear models, we come across the problem of dependence of the design on the values of the parameters being estimated. One way to overcome this problem is to consider Bayesian experimental designs. The Bayesian approach to experimental design provides a way to incorporate prior information in the design process. Bayesian approaches to the optimal design of experiments, when the statistical model is nonlinear, have been reviewed by Atkinson et al. (2007) and by Chaloner and Verdinelli (1995). In the theory of optimal designs, the General Equivalence Theorem (GET) proposed by Kiefer and Wolfowitz (1960) and Kiefer (1974) plays an important role. Most algorithms that search for optimal designs make use of the GET given that it provides a way to verify whether a given design is in fact optimal. Without some sort of version of a GET, construction of D-optimal designs reduces to a difficult optimization problem with multiple maxima and no mechanism for assessing when the optimum has been obtained (Atkinson and Cook, 1995). An extension of GET to Bayesian D-optimality is given by Chaloner and Larntz (1989) and, for a more general case, by Firth and Hinde (1997), both following Whittle (1973).

Bayesian optimal designs for JMMD was introduced by Pinto and Ponce de Leon (2004) and Pinto and Ponce de Leon (2007), however, the validity of the General Equivalence Theorem has not been proven anywhere. In this article, we prove the validity of the GET for obtaining Bayesian D and D_S-optimal designs in joint generalized linear models for the mean and dispersion, following Chaloner and Larntz (1989).

In Section 2, we present the Equivalence Theory for nonlinear models and the Whittle's Theorem, that will be the basis of our work. In Section 3, we present the theory of the joint modeling of mean and dispersion and introduce the Bayesian version of the criteria for D and D_S-optimality, as well as the version of the GET for the JMMD. Final considerations are given in Section 4.

2 Equivalence theory

Let $\mathcal{X}$ be a compact set representing the design region. Define Ξ to be the set of all probability measures over $\mathcal{X}$ and consider a continuous design, represented by the measure $\xi \in \Xi$, satisfying the conditions: $\xi(\mathbf{u}) \geq 0$ and $\int_{\mathcal{X}} d\xi(\mathbf{u}) = 1$ for all $\mathbf{u} \in \mathcal{X}$. The total information or the expected Fisher information matrix $\mathbf{M}(\boldsymbol{\theta}|\xi)$ is obtained as the expected Fisher information matrix per observation, with respect to the design measure ξ on $\mathcal{X}$, that is,

$$\mathbf{M}(\boldsymbol{\theta}|\xi) = \int_{\mathcal{X}} \mathbf{I}(\mathbf{u}|\boldsymbol{\theta})\, d\xi(\mathbf{u}). \tag{2.1}$$

The design problem consists in finding a measure ξ^* in Ξ that maximizes a criterion function $\Psi(\mathbf{M}(\boldsymbol{\theta}|\xi))$, which for nonlinear models is dependent on the value of parameter $\boldsymbol{\theta}$. For two measures ξ_1 and ξ_2 in Ξ, the Fréchet derivative of Ψ at $\mathbf{M}_1 = \mathbf{M}(\boldsymbol{\theta}|\xi_1)$ in the direction of $\mathbf{M}_2 = \mathbf{M}(\boldsymbol{\theta}|\xi_2)$ is defined, when the limit exists, by

$$F_\Psi(\mathbf{M}_1, \mathbf{M}_2) = \lim_{\varepsilon \to 0^+} \frac{1}{\varepsilon}\{\Psi[(1-\varepsilon)\mathbf{M}_1 + \varepsilon\mathbf{M}_2] - \Psi(\mathbf{M}_1)\} \tag{2.2}$$

with $\Psi : \mathcal{M} \to \mathbb{R}$, where $\mathcal{M} = \{\mathbf{M}(\boldsymbol{\theta}|\xi) : \xi \in \Xi, \boldsymbol{\theta} \in \Theta\}$ and Θ is the parameter space. We denote ξ_u as the measure that puts point mass at a single point $\mathbf{u}$ in $\mathcal{X}$, and further define $D_\Psi(\mathbf{u}, \boldsymbol{\theta}, \xi) = F_\Psi[\mathbf{M}(\boldsymbol{\theta}|\xi), \mathbf{M}(\boldsymbol{\theta}|\xi_u)]$.

The theorem proposed by Whittle (1973) for linear models and used by Chaloner and Larntz (1989) for nonlinear models forms the basis for checking whether particular designs are optimal or not and is reproduced below in our notation.

Theorem 2.1 (Whittle, 1973). *If Ψ is concave, then a Ψ-optimal design ξ^* can be equivalently characterized by any of the three conditions:*

(i) ξ^* *maximizes* Ψ;
(ii) ξ^* *minimizes* $\sup_{\mathbf{u}\in\mathcal{X}} D_\Psi(\mathbf{u}, \boldsymbol{\theta}, \xi)$;
(iii) $\sup_{\mathbf{u}\in\mathcal{X}} D_\Psi(\mathbf{u}, \boldsymbol{\theta}, \xi^*) = 0$.

As pointed out by Chaloner and Larntz (1989), the proof of Whittle's Theorem, for linear design problems, can be applied to general nonlinear problems under the following additional assumptions, namely $\mathcal{X}$ is a compact set, the derivatives exist and are continuous in $\mathbf{u} \in \mathcal{X}$, there is at least one measure in Ξ for which Ψ is finite, and Ψ is such that if $\xi_i \to \xi$ in weak convergence then $\Psi(\mathbf{M}(\boldsymbol{\theta}|\xi_i)) \to \Psi(\mathbf{M}(\boldsymbol{\theta}|\xi))$.

3 General Equivalence Theorem for the joint modeling of mean and dispersion

Let $\mathbf{x}$ be a vector representing the set of factors presumed to influence the response expected value and let $\mathbf{z}$ be a vector representing the set of factors presumed to influence the dispersion. Let $\mathbf{u}$ be a vector containing factors occurring in $\mathbf{x}$ and in $\mathbf{z}$. We allow $\mathbf{z}$ to contain some or all of the elements of $\mathbf{x}$ as well as other elements.

For the joint modeling of the mean and dispersion, let y_u be the response to be observed in the experimental unit $\mathbf{u} \in \mathcal{X}$. Suppose that the distribution of Y is unknown, nevertheless, it is assumed that $E(Y_u) = \mu_u$ and $\text{Var}(Y_u) = \phi_u V(\mu_u)$, where ϕ_u is the dispersion parameter and $V(\cdot)$ is the variance function. Let k be a link function for the mean model, that is, $\eta_u = k(\mu_u) = \mathbf{f}^t(\mathbf{x}_u)\boldsymbol{\beta}$, with $\mathbf{f}^t(\mathbf{x}_u) = [f_1(\mathbf{x}_u), \dots, f_p(\mathbf{x}_u)]$, where $f_j(\mathbf{x}_u)$, for $j = 1, \dots, p$, is a known function of $\mathbf{x}_u$ and $\boldsymbol{\beta}$ is a $p \times 1$ vector of unknown parameters; for the dispersion model, let l be the link function, usually ln, that is, $\tau_u = l(\phi_u) = \mathbf{g}^t(\mathbf{z}_u)\boldsymbol{\gamma}$, with $\mathbf{g}^t(\mathbf{z}_u) = [g_1(\mathbf{z}_u), \dots, g_q(\mathbf{z}_u)]$, where $g_j(\mathbf{z}_u)$, for $j = 1, \dots, q$, is a known function of $\mathbf{z}_u$ and $\boldsymbol{\gamma}$ is a $q \times 1$ vector of unknown parameters. We are considering that $f_1(\mathbf{x}_u) = g_1(\mathbf{z}_u) = 1$.

Using the extended quasi likelihood as a criterion of estimation and supposing the model for dispersion as gamma (see Pinto and Ponce de Leon, 2004), the quasi Fisher information matrix per observation is given by

$$\mathbf{I}_C(\mathbf{u}|\boldsymbol{\theta}) = \begin{bmatrix} w_u \mathbf{f}(\mathbf{x}_u)\mathbf{f}^t(\mathbf{x}_u) & \mathbf{0} \\ \mathbf{0} & v_u \mathbf{g}(\mathbf{z}_u)\mathbf{g}^t(\mathbf{z}_u) \end{bmatrix}, \tag{3.1}$$

where $\boldsymbol{\theta}^t = (\boldsymbol{\beta}^t, \boldsymbol{\gamma}^t)$, $w_u = (\frac{\partial \mu_u}{\partial \eta_u})^2 \frac{1}{\phi_u V(\mu_u)}$ and $v_u = (\frac{\partial \phi_u}{\partial \tau_u})^2 \frac{1}{2\phi_u^2}$.

For a design $\xi \in \Xi$, the quasi information matrix for the JMMD is given by

$$\mathbf{M}_C(\boldsymbol{\theta}|\xi) = \int_{\mathcal{X}} \mathbf{I}_C(\mathbf{u}|\boldsymbol{\theta})\, d\xi(\mathbf{u}) = \begin{bmatrix} \mathbf{C} & \mathbf{0} \\ \mathbf{0} & \mathbf{D} \end{bmatrix}, \tag{3.2}$$

where $\mathbf{C} = \mathbf{C}(\boldsymbol{\theta}|\xi) = \int_{\mathcal{X}} w_u \mathbf{f}(\mathbf{x}_u)\mathbf{f}^t(\mathbf{x}_u)\, d\xi(\mathbf{u})$ and $\mathbf{D} = \mathbf{D}(\boldsymbol{\theta}|\xi) = \int_{\mathcal{X}} v_u \mathbf{g}(\mathbf{z}_u) \times \mathbf{g}^t(\mathbf{z}_u)\, d\xi(\mathbf{u})$.

Note that the quasi information matrix $\mathbf{M}_C(\boldsymbol{\theta}|\xi)$ depends on the value of the parameter $\boldsymbol{\theta}$. In this case, we will use the Bayesian approach as described by Chaloner and Larntz (1989) and Atkinson et al. (2007) for nonlinear models.

3.1 Bayesian D-optimal criterion

The Bayesian D-optimal criterion for the JMMD is defined as

$$\psi(\mathbf{M}_C(\boldsymbol{\theta}|\xi)) = E_\theta[\ln|\mathbf{M}_C(\boldsymbol{\theta}|\xi)|], \tag{3.3}$$

with $\psi : \mathcal{M} \to \mathbb{R}$, where $\mathcal{M} = \{\mathbf{M}_C(\boldsymbol{\theta}|\xi) : \xi \in \Xi, \boldsymbol{\theta} \in \Theta\}$, E_θ refers to expectation with respect to a prior distribution on $\boldsymbol{\theta}$ and $\ln|\mathbf{M}_C(\boldsymbol{\theta}|\xi)|$ represents the natural logarithm of the determinant of the quasi information matrix. In this way, the Bayesian D-optimal problem consists of finding the design measure ξ^*, from a specified class of design measures Ξ, that maximizes ψ.

Proposition 3.1. *The Bayesian criterion function $\psi = E_\theta[\ln|\mathbf{M}_C(\boldsymbol{\theta}|\xi)|]$ is strictly concave over $\mathcal{M}^+$, the subset of $\mathcal{M}$ where ψ is finite.*

Let the quasi information matrix associated to the design ξ_u, which puts point mass at a single point $\mathbf{u}$ in $\mathcal{X}$, be denoted by $\mathbf{M}_C(\boldsymbol{\theta}|\xi_u)$.

Proposition 3.2. *The Fréchet derivative of the criterion function* ψ *at* $\mathbf{M}_C(\boldsymbol{\theta}|\xi)$ *in the direction of* $\mathbf{M}_C(\boldsymbol{\theta}|\xi_u)$ *is given by*

$$D_\psi(\mathbf{u},\xi) = E_\theta\big[\mathrm{tr}\big(\mathbf{M}_{C_u}\mathbf{M}_C^{-1}\big)\big] - t, \tag{3.4}$$

where $\mathbf{M}_C = \mathbf{M}_C(\boldsymbol{\theta}|\xi)$, $\mathbf{M}_{C_u} = \mathbf{M}_C(\boldsymbol{\theta}|\xi_u)$, t *is the number of parameters in the models of the mean and dispersion and* $\mathrm{tr}(\cdot)$ *represents the operator trace.*

In (3.4) $\mathbf{M}_{C_u} = \mathbf{I}_C(\mathbf{u}|\boldsymbol{\theta})$ is given in (3.1), thus

$$D_\psi(\mathbf{u},\xi) = E_\theta\big[w_u\mathbf{f}^t(\mathbf{x}_u)\mathbf{C}^{-1}\mathbf{f}(\mathbf{x}_u) + v_u\mathbf{g}^t(\mathbf{z}_u)\mathbf{D}^{-1}\mathbf{g}(\mathbf{z}_u)\big] - t. \tag{3.5}$$

We further define the sensitivity function as $d_\psi(\mathbf{u},\xi) = E_\theta[\mathrm{tr}(\mathbf{M}_{C_u}\mathbf{M}_C^{-1})]$, thereby the sensitivity function at the point $\mathbf{u}$ is given by

$$d_\psi(\mathbf{u},\xi) = E_\theta\big[w_u\mathbf{f}^t(\mathbf{x}_u)\mathbf{C}^{-1}\mathbf{f}(\mathbf{x}_u) + v_u\mathbf{g}^t(\mathbf{z}_u)\mathbf{D}^{-1}\mathbf{g}(\mathbf{z}_u)\big]. \tag{3.6}$$

Theorem 3.1 (GET). *Let* Ξ *be the class of all probability measures on* $\mathcal{X}$. *Let* $\psi = E_\theta[\ln|\mathbf{M}_C(\boldsymbol{\theta}|\xi)|]$ *be the Bayesian D-optimal criterion for the JMMD. Then a* ψ*-optimal design* ξ^* *can be equivalently characterized by any of the three conditions.*

(a) *The design* ξ^* *maximizes* ψ *over* Ξ;
(b) ξ^* *minimizes* $\sup_{\mathbf{u}\in\mathcal{X}} D_\psi(\mathbf{u},\xi)$;
(c) $\sup_{\mathbf{u}\in\mathcal{X}} d_\psi(\mathbf{u},\xi) = t$.

Where $D_\psi(\mathbf{u},\xi)$ *and* $d_\psi(\mathbf{u},\xi)$ *are given in* (3.5) *and* (3.6), *respectively and* t *is the number of parameters in the models of the mean and dispersion.*

As $\psi = E_\theta[\ln|\mathbf{M}_C(\boldsymbol{\theta}|\xi)|]$ is a concave function by Proposition 3.1, the proof of Theorem 3.1 follows directly from Chaloner and Larntz (1989) using the Theorem 2.1.

3.2 Bayesian D_S-optimal criterion

The theory of D_S-optimality consists in finding optimal designs when we have no interest in the estimation of all parameters in the model. D_S-optimality is a particular case of T-optimality when we are interested in discriminating between two nested models (see Atkinson et al., 2007 and Silvey, 1980).

For the JMMD, we could be interested in a reduced number of parameters both in the mean model as in the dispersion model. Suppose that we are interested in s_m parameters in the mean model and in s_d parameters in the dispersion model, with $1 \le s_m \le p$ and $1 \le s_d \le q$. Thus, we can write $\mathbf{f}^t(\mathbf{x}_u) = (\mathbf{f}_1^t(\mathbf{x}_u), \mathbf{f}_2^t(\mathbf{x}_u))$ and $\mathbf{g}^t(\mathbf{z}_u) = (\mathbf{g}_1^t(\mathbf{z}_u), \mathbf{g}_2^t(\mathbf{z}_u))$ with $\mathbf{f}_1^t(\mathbf{x}_u) = (f_1(\mathbf{x}_u), \dots, f_{s_m}(\mathbf{x}_u))$, $\mathbf{f}_2^t(\mathbf{x}_u) =$

$(f_{s_{m+1}}(\mathbf{x}_u), \ldots, f_p(\mathbf{x}_u))$, $\mathbf{g}_1^t(\mathbf{z}_u) = (g_1(\mathbf{z}_u), \ldots, g_{s_d}(\mathbf{z}_u))$ and $\mathbf{g}_2^t(\mathbf{z}_u) = (g_{s_{d+1}}(\mathbf{z}_u), \ldots, g_q(\mathbf{z}_u))$. The quasi information matrix for the JMMD, given in (3.2), can be written as follows

$$\mathbf{M}_C(\boldsymbol{\theta}|\xi) = \begin{bmatrix} \mathbf{C}_{11} & \mathbf{C}_{12} & \mathbf{0} & \mathbf{0} \\ \mathbf{C}_{21} & \mathbf{C}_{22} & \mathbf{0} & \mathbf{0} \\ \mathbf{0} & \mathbf{0} & \mathbf{D}_{11} & \mathbf{D}_{12} \\ \mathbf{0} & \mathbf{0} & \mathbf{D}_{21} & \mathbf{D}_{22} \end{bmatrix}, \tag{3.7}$$

where $\mathbf{C}_{ij} = \mathbf{C}_{ij}(\boldsymbol{\theta}|\xi) = \int_{\mathcal{X}} w_u \mathbf{f}_i(\mathbf{x}_u)\mathbf{f}_j^t(\mathbf{x}_u)\, d\xi(\mathbf{u})$ and $\mathbf{D}_{ij} = \mathbf{D}_{ij}(\boldsymbol{\theta}|\xi) = \int_{\mathcal{X}} v_u \mathbf{g}_i(\mathbf{z}_u)\mathbf{g}_j^t(\mathbf{z}_u)\, d\xi(\mathbf{u})$, for $i, j = 1, 2$. Thus, considering the Bayesian approach, Pinto and Ponce de Leon (2007) proposed the Bayesian criterion function for the D_S-optimality as follows

$$\varphi[\mathbf{M}_C(\boldsymbol{\theta}|\xi)] = E_\theta\left\{\ln\left[\frac{|\mathbf{C}|}{|\mathbf{C}_{22}|}\right] + \ln\left[\frac{|\mathbf{D}|}{|\mathbf{D}_{22}|}\right]\right\}, \tag{3.8}$$

with $\varphi : \mathcal{M} \to \mathbb{R}$, where $\mathcal{M} = \{\mathbf{M}_C(\boldsymbol{\theta}|\xi) : \xi \in \Xi, \boldsymbol{\theta} \in \Theta\}$, E_θ refers to expectation with respect to a prior distribution on $\boldsymbol{\theta}$, ln represents the natural logarithm and $|\mathbf{B}|$ represents the determinant of the matrix $\mathbf{B}$. In this way, the Bayesian D_S-optimal problem consists of finding the design measure ξ^*, from a specified class of design measures Ξ, that maximizes φ.

Proposition 3.3. *The Bayesian criterion function φ is strictly concave over $\mathcal{M}^+$, the subset of $\mathcal{M}$ where φ is finite.*

The Fréchet derivative for D_S-optimality of φ at $\mathbf{M}_{C_1}$ in the direction of $\mathbf{M}_{C_2}$, where $\mathbf{M}_{C_1} = \mathbf{M}_C(\boldsymbol{\theta}|\xi_1)$, $\mathbf{M}_{C_2} = \mathbf{M}_C(\boldsymbol{\theta}|\xi_2)$ and $\xi_1, \xi_2 \in \Xi$, is given by

$$\begin{aligned} F_\varphi(\mathbf{M}_{C_1}, \mathbf{M}_{C_2}) = E_\theta[&\mathrm{tr}(\mathbf{C}_2\mathbf{C}_1^{-1}) - \mathrm{tr}(\mathbf{C}_{2,22}\mathbf{C}_{1,22}^{-1}) \\ &+ \mathrm{tr}(\mathbf{D}_2\mathbf{D}_1^{-1}) - \mathrm{tr}(\mathbf{D}_{2,22}\mathbf{D}_{1,22}^{-1})] - s, \end{aligned} \tag{3.9}$$

where $s = s_m + s_d$, $\mathbf{C}_j = \mathbf{C}(\boldsymbol{\theta}|\xi_j)$, $\mathbf{D}_j = \mathbf{D}(\boldsymbol{\theta}|\xi_j)$, $\mathbf{C}_{j,22} = \mathbf{C}_{22}(\boldsymbol{\theta}|\xi_j)$ and $\mathbf{D}_{j,22} = \mathbf{D}_{22}(\boldsymbol{\theta}|\xi_j)$, for $j = 1, 2$. The proof of (3.9), given by Pinto and Ponce de Leon (2007), is conducted showing that

$$\begin{aligned} F_\varphi(\mathbf{M}_{C_1}, \mathbf{M}_{C_2}) = &F_\psi(\mathbf{C}_1, \mathbf{C}_2) - F_\psi(\mathbf{C}_{1,22}, \mathbf{C}_{2,22}) \\ &+ F_\psi(\mathbf{D}_1, \mathbf{D}_2) - F_\psi(\mathbf{D}_{1,22}, \mathbf{D}_{2,22}), \end{aligned} \tag{3.10}$$

where $F_\psi(\mathbf{M}_1, \mathbf{M}_2) = E_\theta[\mathrm{tr}(\mathbf{M}_2\mathbf{M}_1^{-1})] - m$ is the Fréchet derivative for the Bayesian D-optimality of ψ at $\mathbf{M}_1 = \mathbf{M}(\boldsymbol{\theta}|\xi_1)$ in direction of $\mathbf{M}_2 = \mathbf{M}(\boldsymbol{\theta}|\xi_2)$, as shown by Proposition 3.2. Here, $m = \dim(\mathbf{M}_1) = \dim(\mathbf{M}_2)$, where $\dim(\mathbf{A})$ represents the dimension of the square matrix $\mathbf{A}$.

Thus, using (3.9), the Fréchet derivative of φ at $\mathbf{M}_C = \mathbf{M}_C(\boldsymbol{\theta}|\xi)$ in the direction of $\mathbf{M}_{Cu} = \mathbf{M}_C(\boldsymbol{\theta}|\xi_u)$ is given by

$$\begin{aligned} D_\varphi(\mathbf{u}, \xi) &= F_\varphi(\mathbf{M}_C, \mathbf{M}_{C_u}) \\ &= E_\theta[\operatorname{tr}(\mathbf{C}_u\mathbf{C}^{-1}) - \operatorname{tr}(\mathbf{C}_{u,22}\mathbf{C}_{22}^{-1}) \\ &\quad + \operatorname{tr}(\mathbf{D}_u\mathbf{D}^{-1}) - \operatorname{tr}(\mathbf{D}_{u,22}\mathbf{D}_{22}^{-1})] - s, \end{aligned} \tag{3.11}$$

where $\mathbf{C}_u = w_u\mathbf{f}(\mathbf{x}_u)\mathbf{f}^t(\mathbf{x}_u)$, $\mathbf{C}_{u,22} = w_u\mathbf{f}_2(\mathbf{x}_u)\mathbf{f}_2^t(\mathbf{x}_u)$, $\mathbf{D}_u = v_u\mathbf{g}(\mathbf{z}_u)\mathbf{g}^t(\mathbf{z}_u)$, $\mathbf{D}_{u,22} = v_u\mathbf{g}_2(\mathbf{z}_u)\mathbf{g}_2^t(\mathbf{z}_u)$, thus

$$\begin{aligned} D_\varphi(\mathbf{u}, \xi) = E_\theta[&w_u\mathbf{f}^t(\mathbf{x}_u)\mathbf{C}^{-1}\mathbf{f}(\mathbf{x}_u) + v_u\mathbf{g}^t(\mathbf{z}_u)\mathbf{D}^{-1}\mathbf{g}(\mathbf{z}_u) \\ &- w_u\mathbf{f}_2^t(\mathbf{x}_u)\mathbf{C}_{22}^{-1}\mathbf{f}_2(\mathbf{x}_u) - v_u\mathbf{g}_2^t(\mathbf{z}_u)\mathbf{D}_{22}^{-1}\mathbf{g}_2(\mathbf{z}_u)] - s. \end{aligned} \tag{3.12}$$

As before, the sensitivity function at the point $\mathbf{u}$ is defined as

$$\begin{aligned} d_\varphi(\mathbf{u}, \xi) = E_\theta\{&w_u[\mathbf{f}^t(\mathbf{x}_u)\mathbf{C}^{-1}\mathbf{f}(\mathbf{x}_u) - \mathbf{f}_2^t(\mathbf{x}_u)\mathbf{C}_{22}^{-1}\mathbf{f}_2(\mathbf{x}_u)] \\ &+ v_u[\mathbf{g}^t(\mathbf{z}_u)\mathbf{D}^{-1}\mathbf{g}(\mathbf{z}_u) - \mathbf{g}_2^t(\mathbf{z}_u)\mathbf{D}_{22}^{-1}\mathbf{g}_2(\mathbf{z}_u)]\}. \end{aligned} \tag{3.13}$$

Theorem 3.2 (GET). *Let Ξ denote the class of all probability measures on $\mathcal{X}$. Let φ be the Bayesian D_S-optimal criterion for the JMMD. Then a φ-optimal design ξ^* can be equivalently characterized by any of the three conditions.*

(a) *The design ξ^* maximizes φ over Ξ;*
(b) *ξ^* minimizes $\sup_{\mathbf{u}\in\mathcal{X}} D_\varphi(\mathbf{u}, \xi)$;*
(c) $\sup_{\mathbf{u}\in\mathcal{X}} d_\varphi(\mathbf{u}, \xi) = s$.

Where $D_\varphi(\mathbf{u}, \xi)$ and $d_\varphi(\mathbf{u}, \xi)$ are given in (3.12) *and* (3.13), *respectively and $s = s_m + s_d$.*

Again, as φ is a concave function by Proposition 3.3, the proof of Theorem 3.2 follows directly from Chaloner and Larntz (1989) using the Theorem 2.1.

4 Final considerations

The attainment of optimal designs for joint models of mean and dispersion is very important, specially in quality-improvement experiments, because it promotes the reduction of the number of points to be considered in the experiment, resulting in savings of both time and money. The optimal design ensures the model parameters will be estimated as best as possible. The General Equivalence Theorem enables optimal designs to be found, giving a form to verify if in fact the design at hand is optimal. For more details and examples about optimal designs for the JMMD see Pinto and Ponce de Leon (2004) and Pinto and Ponce de Leon (2007).

Appendix

A.1 Proof of the Proposition 3.1

We consider $\mathbf{M}_{C_1} = \mathbf{M}_C(\boldsymbol{\theta}|\xi_1)$, $\mathbf{M}_{C_2} = \mathbf{M}_C(\boldsymbol{\theta}|\xi_2)$ two information matrices based on the designs ξ_1 and ξ_2, respectively. Let be $0 < \alpha < 1$ a real number. Then for $\psi[\mathbf{M}_C(\boldsymbol{\theta}|\xi)] = E_\theta[\ln|\mathbf{M}_C(\boldsymbol{\theta}|\xi)|]$ we have that $\psi[\alpha\mathbf{M}_{C_1} + (1-\alpha)\mathbf{M}_{C_2}] = E_\theta[\ln|\alpha\mathbf{M}_{C_1} + (1-\alpha)\mathbf{M}_{C_2}|] > E_\theta[\ln|\mathbf{M}_{C_1}|^\alpha + \ln|\mathbf{M}_{C_2}|^{(1-\alpha)}] = \alpha E_\theta \ln|\mathbf{M}_{C_1}| + (1-\alpha)E_\theta \ln|\mathbf{M}_{C_2}| = \alpha\psi(\mathbf{M}_{C_1}) + (1-\alpha)\psi(\mathbf{M}_{C_2})$.

A.2 Proof of the Proposition 3.2

The Fréchet derivative of the function ψ at $\mathbf{M}_C = \mathbf{M}_C(\boldsymbol{\theta}|\xi)$ in the direction of $\mathbf{M}_{C_u} = \mathbf{M}_C(\boldsymbol{\theta}|\xi_u)$ is given by $F_\psi(\mathbf{M}_C, \mathbf{M}_{C_u}) = \lim_{\varepsilon\to 0^+} \frac{1}{\varepsilon}\{\psi[(1-\varepsilon)\mathbf{M}_C + \varepsilon\mathbf{M}_{C_u}] - \psi(\mathbf{M}_C)\}$.

Our proof is based on Silvey (1980, page 21). The Bayesian criterion for the D-optimality is $\psi = E_\theta[\ln|\mathbf{M}_C(\boldsymbol{\theta}|\xi)|]$. We start by calculating $F_\psi(\mathbf{M}_C, \mathbf{M}_{C_u})$ for nonsingular $\mathbf{M}_C$ and the Fréchet derivative via the Gâutex derivative given by $G_\psi(\mathbf{M}_C, \mathbf{M}_{C_u}) = \lim_{\varepsilon\to 0^+} \frac{1}{\varepsilon}\{\psi[\mathbf{M}_C + \mathbf{M}_{C_u}] - \psi(\mathbf{M}_C)\}$. Then, $G_\psi(\mathbf{M}_C, \mathbf{M}_{C_u}) = \lim_{\varepsilon\to 0^+} \frac{1}{\varepsilon}\{E_\theta \ln|(\mathbf{M}_C + \varepsilon\mathbf{M}_{C_u})| - E_\theta \ln|\mathbf{M}_C|\} = \lim_{\varepsilon\to 0^+} \frac{1}{\varepsilon}\{E_\theta \ln[|(\mathbf{M}_C + \varepsilon\mathbf{M}_{C_u})||\mathbf{M}_C|^{-1}]\} = \lim_{\varepsilon\to 0^+} \frac{1}{\varepsilon}\{E_\theta \ln[|(\mathbf{I}_t + \varepsilon\mathbf{M}_{C_u}\mathbf{M}_C^{-1}|]\} = \lim_{\varepsilon\to 0^+} \frac{1}{\varepsilon}\{E_\theta \ln[1 + \varepsilon\,\mathrm{tr}(\mathbf{M}_{C_u}\mathbf{M}_C^{-1}) + O(\varepsilon^2)]\} = \lim_{\varepsilon\to 0^+} \frac{1}{\varepsilon}E_\theta\{\varepsilon\,\mathrm{tr}(\mathbf{M}_{C_u}\mathbf{M}_C^{-1}) + O(\varepsilon^2)]\} = E_\theta \lim_{\varepsilon\to 0^+} \frac{1}{\varepsilon}[\varepsilon\,\mathrm{tr}(\mathbf{M}_{C_u}\mathbf{M}_C^{-1}) + O(\varepsilon^2)] = E_\theta[\mathrm{tr}(\mathbf{M}_{C_u}\mathbf{M}_C^{-1})]$. Here, $\mathbf{I}_t$ is the identity matrix $t \times t$.

It is known that (see Silvey, 1980) $F_\psi(\mathbf{M}_C, \mathbf{M}_{C_u}) = G_\psi(\mathbf{M}_C, \mathbf{M}_{C_u} - \mathbf{M}_C) = E_\theta\{\mathrm{tr}[(\mathbf{M}_{C_u} - \mathbf{M}_C)\mathbf{M}_C^{-1})]\} = E_\theta\{\mathrm{tr}[(\mathbf{M}_{C_u}\mathbf{M}_C^{-1} - \mathbf{I}_t))]\} = E_\theta\{\mathrm{tr}[\mathbf{M}_{C_u}\mathbf{M}_C^{-1}]\} - t$.

A.3 Proof of the Proposition 3.3

We Consider the matrix $\mathbf{M}_C = \mathbf{M}_C(\boldsymbol{\theta}|\xi)$ partitioned as the equation (3.7), let be $\mathbf{M}_{C_1} = \mathbf{M}_C(\boldsymbol{\theta}|\xi_1)$, $\mathbf{M}_{C_2} = \mathbf{M}_C(\boldsymbol{\theta}|\xi_2)$, also partitioned as the equation (3.7), two quasi information matrices based on the designs ξ_1 and ξ_2, respectively. The Bayesian criterion for the D_S-optimality is $\varphi[\mathbf{M}_C] = E_\theta\{\ln[\frac{|\mathbf{C}|}{|\mathbf{C}_{22}|}] + \ln[\frac{|\mathbf{D}|}{|\mathbf{D}_{22}|}]\}$. Let be $0 < \alpha < 1$ a real number, then $\varphi[\alpha\mathbf{M}_{C_1} + (1-\alpha)\mathbf{M}_{C_2}] = E_\theta\{\ln[\frac{|\alpha\mathbf{C}_1+(1-\alpha)\mathbf{C}_2|}{|\alpha\mathbf{C}_{1,22}+(1-\alpha)\mathbf{C}_{2,22}|}] + \ln[\frac{|\alpha\mathbf{D}_1+(1-\alpha)\mathbf{D}_2|}{|\alpha\mathbf{D}_{1,22}+(1-\alpha)\mathbf{D}_{2,22}|}]\} = E_\theta\{\ln|\alpha\mathbf{C}_1 + (1-\alpha)\mathbf{C}_2| + \ln|\alpha\mathbf{C}_{1,22} + (1-\alpha)\mathbf{C}_{2,22}|^{-1} + \ln|\alpha\mathbf{D}_1 + (1-\alpha)\mathbf{D}_2| + \ln|\alpha\mathbf{D}_{1,22} + (1-\alpha) \times \mathbf{D}_{2,22}|^{-1}\} > E_\theta\{\ln[|\mathbf{C}_1|^\alpha|\mathbf{C}_2|^{(1-\alpha)}] + \ln[|\mathbf{C}_{1,22}|^{-\alpha}|\mathbf{C}_{2,22}|^{-(1-\alpha)}] + \ln[|\mathbf{D}_1|^\alpha \times |\mathbf{D}_2|^{(1-\alpha)}] + \ln[|\mathbf{D}_{1,22}|^{-\alpha}|\mathbf{D}_{2,22}|^{-(1-\alpha)}]\} = E_\theta\{\alpha[\ln(\frac{|\mathbf{C}_1|}{|\mathbf{C}_{1,22}|}) + \ln(\frac{|\mathbf{D}_1|}{|\mathbf{D}_{1,22}|})] + (1-\alpha)[\ln(\frac{|\mathbf{C}_2|}{|\mathbf{C}_{2,22}|}) + \ln(\frac{|\mathbf{D}_2|}{|\mathbf{D}_{2,22}|})]\} = \alpha\varphi(\mathbf{M}_{C_1}) + (1-\alpha)\varphi(\mathbf{M}_{C_2})$. The properties about determinants can be found in Fedorov (1972).

Acknowledgments

The authors would like to express their gratitude to the Editor and especially thank the referees for carefully reading the paper. Their suggestions and constructive criticisms led us to improve the paper substantially.

References

Atkinson, A. C. and Cook, R. D. (1995). D-optimal designs for heteroscedastic linear models. *Journal of the American Statistical Association* **90**(429), 204–212. MR1325128

Atkinson, A. C., Donev, A. N. and Tobias, R.D. (2007). *Optimum Experimental Design, with SAS*. Oxford: Oxford Univ. Press. MR2323647

Chaloner, K. and Larntz, K. (1989). Optimal Bayesian design applied to logistic regression experiments. *Journal of the Statistical Planning and Inference* **21**, 191–208. MR0985457

Chaloner, K. and Verdinelli, I. (1995). Bayesian experimental design: A review. *Statistical Science* **10**, 273–304. MR1390519

Clyde, M. A. (2001). Experimental design: A Bayesian perspective. In *International Encyclopedia of the Social & Behavioral Sciences* (N. J. Smelser and P. B. Baltes, eds.) 5075–5081. Oxford: Pergamon.

Fedorov, V. V. (1972). *Theory of Optimal Experiments*. New York: Academic Press. MR0403103

Firth D. and Hinde, J.P. (1997). On Bayesian D-optimum criteria and the Equivalence Theorem in non-linear models. *Journal of the Royal Statistical Society*, *Ser. B* **59**(4), 793–797. MR1483214

Kiefer, J. (1974). General Equivalence Theorem for optimum designs (approximate theory). *The Annals of Statistics* **2**, 849–879. MR0356386

Kiefer, J. and Wolfowitz, J. (1960). The equivalence of two extremum problems. *Canadian Journal of Statistics* **12**, 363–366. MR0117842

Lee, Y. and Nelder, J. A. (1998). Generalized linear models for the analysis of quality-improvement experiments. *Canadian Journal of Statistics* **26**, 95–105. MR1624385

Lee, Y. and Nelder, J.A. (2003). Robust design via generalized linear models. *Journal of Quality Technology* **35**, 2–12.

Nelder, J.A. and Lee, Y. (1991). Generalized linear models for the analysis of Taguchi-type experiments. *Applied Stochastic Models and Data Analysis* **7**, 107–120.

Pinto, E. R. and Ponce de Leon, A. (2004). Bayesian D-optimum designs for generalized linear models with varying dispersion parameter. In *Contributions to Statistics* (A. Di Bucchianico, H. Läuter and H. P. Wynn, eds.). mODa7-Advances in Model Oriented Data Analysis, 127–134. New York: Springer.

Pinto, E. R. and Ponce de Leon, A. (2007). Bayesian D_S-optimum designs for generalized linear models with varying dispersion parameter. In *Contributions to Statistics* (J. López-Fidalgo, J. M. Rodríguez-Días and B. Torsney, eds.). mODa8-Advances in Model Oriented Data Analysis, 181–188. New York: Springer. MR2404081

Silvey, S. D. (1980). *Optimal Design*. London: Chapman & Hall. MR0606742

Whittle, P. (1973). Some general points in the theory of optimal experimental design. *Journal of the Royal Statistical Society*, *Ser. B*, **35**, 123–130. MR0356388

Faculdade de Matemática
Universidade Federal de Uberlândia
Uberlândia-MG
Brazil
E-mail: edmilson@famat.ufu.br

Instituto de Medicina Social
Universidade do Estado do Rio de Janeiro
Rio de Janeiro-RJ
Brazil
E-mail: ponce@ims.uerj.br