Optimal Design of Experiments Subject to Correlated Errors

Pazman, Andrej; Müller, Werner

DOI:
10.57938/e0f36ef1-9cb9-4b44-aead-a7eb391cd0f8

Published: 01/01/2000

Document Version
Publisher's PDF, also known as Version of record

Optimal Design of Experiments Subject to Correlated Errors

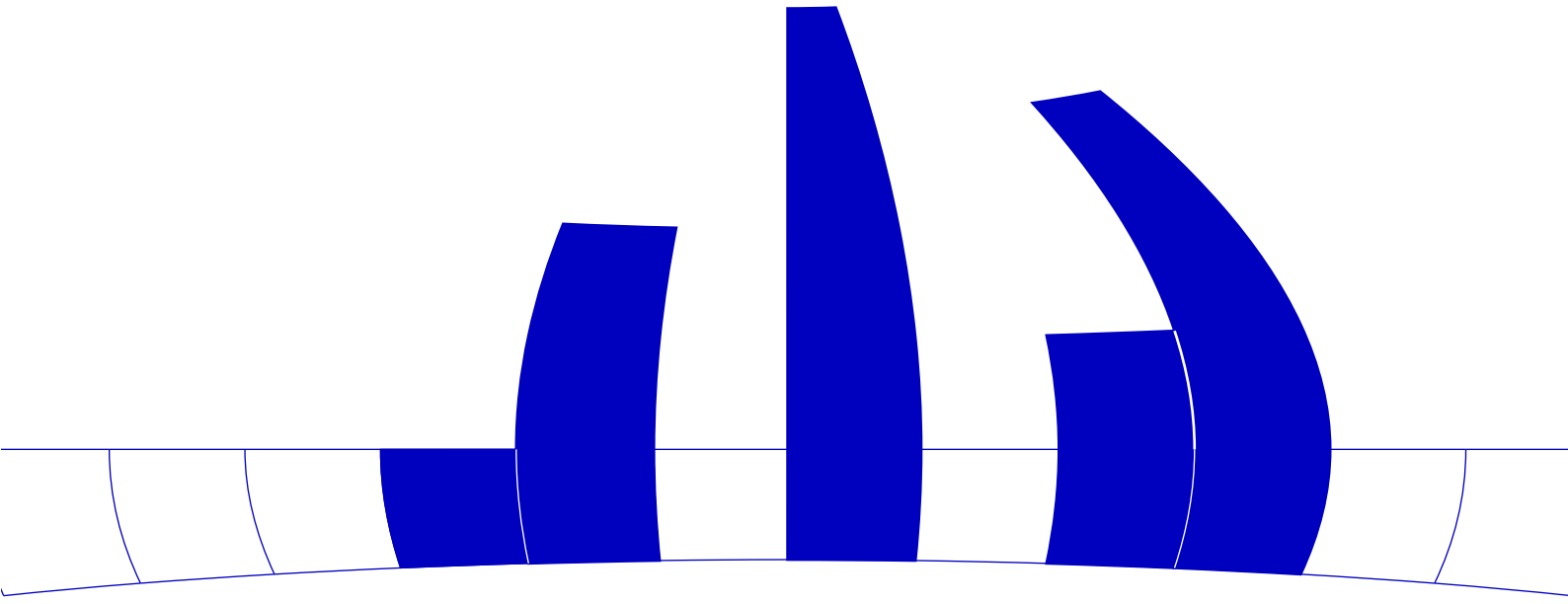
Andrej Pázman, Werner Müller

Institut für Statistik
Wirtschaftsuniversität Wien

Forschungsberichte

Bericht 77
May 2000

http://statmath.wu-wien.ac.at/



Optimal Design of Experiments Subject to Correlated Errors

Andrej Pázman[1] and Werner G. Müller [2]
Comenius University Bratislava, Vienna University of Economics and B.A.

Abstract

In this paper we consider optimal design of experiments in the case of correlated observations, when no replications are possible. This situation is typical when observing a random process or random field with known covariance structure. We present a theorem which demonstrates that the computation of optimum exact designs corresponds to solving minimization problems in terms of design measures.

Keywords: design measure, approximate information matrix, random field.

1 Introduction.

Designing an experiment aims at obtaining more effective parameter estimators or model selection procedures by an adequate selection of trials. Many results in design of experiments are for uncorrelated observations, where consequently independent replications of observations are allowed for. Much of this is considered as classical today (see for instance the recent book by Pukelsheim, 1993). These results rely very much on the fact that instead of an 'exact' experimental design we can use a so-called design measure which is proportional to the number of independent replications. The information which is obtained from the experiment is then expressed through an information matrix, which is a simple function of the design measure. The situation is very different when the observations are correlated according to a given covariance function. Typically, such a situation arises when we observe the values of a random process or a random field at some (time-) points. When the mean of the process contains unknown parameters, we have again a regression model, but typically without the possibility of replications, because just one realization of the process is allowed for, and the experimental design consists of an adequate choice of observation points (times).

[1]The first autor is supported by the grant No. 1/7295/20 of the Slovak Grant Agency VEGA.

[2]*Address for correspondence:* Department of Statistics, University of Economics and Business Administration, Augasse 2-6, A-1090 Vienna, Austria. Email: werner.mueller@wu-wien.ac.at

The most elegant results in this domain have been obtained in a series of papers started by Sacks and Ylvisaker (1966), whose approach is asymptotical in the sense that the rate of increase of information is used as the design criterion. Another suggestion goes back to Brimkulov et al. (1980), who heuristically extend ideas from well-known algorithmic approaches to the correlated case. For a recent detailed survey of the above and some alternative methods see Chapter 5 of Müller (1998). However, until recently there was no approach to the problem which utilizes design measures, although the results from the classical (uncorrelated) theory make this concept so attractive. The aim of the present paper is to show that it is possible to associate some 'approximate' information matrices based on an approach introduced by Pázman and Müller (1998) to each design measure. Uniform design measures correspond then to exact designs, and in this case the approximate information matrix coincides with the standard one.

2 The setup.

Let $\mathcal{X}$ denote the design space (a finite set of potential trials). At each $x \in \mathcal{X}$ one can observe a random variable

$$y(x) = f^T(x)\theta + \epsilon(x).$$

Here $f^T(x)\theta$ is the response function at x containing m unknown parameters $\theta = (\theta_1, \ldots, \theta_m)^T \in \mathbb{R}^m$. The noise $\epsilon(x)$ is supposed to have zero mean and a covariance function

$$\mathrm{Cov}[\epsilon(x), \epsilon(z)] = \sigma^2 C(x, z),$$

with $C(x, z)$ known, and σ^2 unknown. In this setup an exact N_A-point design is a set $A = \{x_1, \ldots, x_{N_A}\} \subset \mathcal{X}$. Replications are not allowed, i.e. $x_i \neq x_j$ if $i \neq j$.

For any $A \subset \mathcal{X}$ we denote by $C(A)$ a matrix with entries $C(x, z); x \in A, z \in A$. Hence $\sigma^2 C(A)$ is the covariance matrix of the observed vector, when the design A is used. We shall suppose that the matrix $C(\mathcal{X})$ is regular; consequently $C(A)$ is regular for every $A \subset \mathcal{X}$. The information matrix corresponding to the exact design A is

$$M(A) = \sigma^{-2} \sum_{x \in A} \sum_{z \in A} f(x)[C^{-1}(A)]_{x,z} f^T(z). \tag{1}$$

The optimum N-point design is a solution of the minimization problem

$$\min_{A \subset \mathcal{X}, N_A \leq N} \Phi[M(A)], \tag{2}$$

where N_A denotes the number of points in the design A, and $\Phi[.]$ is a given criterion function. In particular if $\Phi[M] = -\ln\det M$, we have the criterion of D-optimality, if $\Phi[M] = \mathrm{tr}M^{-1}$ (if M is regular) or $\Phi[M] = +\infty$ (if M is singular), we have the criterion of A-optimality, etc.

3 Perturbation of observations by a design depending supplementary noise.

Here we consider to solve the discontinuous problem (2) by smoothing the involved functions. For this purpose we introduce design measures on $\mathcal{X}$, a concept well known from the classical theory (for uncorrelated observations with replications see Kiefer, 1959), however with a clearly different interpretation.

By definition, a design measure ξ is any probability measure on $\mathcal{X}$, supported by a finite set (the so-called support of ξ)

$$S_\xi = \{x : \xi(x) > 0\}.$$

Hence, $\xi(x) \geq 0$ for every $x \in \mathcal{X}$, and $\sum_{x\in\mathcal{X}} \xi(x) = 1$.

In Kiefer's setup the value of $\xi(x)$ is proportional to the number of independent and uncorrelated replications of observations in the same point $x \in \mathcal{X}$. For situations without replications the interpretation must be quite different, but two features from Kiefer's approach need to remain:

a) $\xi(x) = 0$ means that there is no information from observations at x;

b) by increasing $\xi(x)$ we increase the amount of information from the observation at x.

The main idea how to use a design measure ξ for experiments in the correlated case is based upon a suggestion by the authors in Pàzman and Müller (1998), namely perturbing the observed process $y(x); x \in \mathcal{X}$ by a supplementary 'design depending' noise. It is presented in the following diagram:

$$f^T(x)\theta \longrightarrow \overset{\epsilon(x)}{\bigoplus} \rightarrow y(x) = f^T(x)\theta + \epsilon(x) \rightarrow \overset{\epsilon_{\xi,\gamma}(x)}{\bigoplus} \rightarrow y_\xi(x) = f^T(x)\theta + \epsilon(x) + \epsilon_{\xi,\gamma}(x).$$

To the original (truly observed) process $y(x); x \in \mathcal{X}$ we add a supplementary white noise $\epsilon_{\xi,\gamma}(x); x \in \mathcal{X}$ with

$$\sigma^{-2}\mathrm{Var}[\epsilon_{\xi,\gamma}(x)] \;=\; \gamma \ln\left[\frac{\xi_{\max}}{\xi(x)}\right] \qquad (3)$$

$$= \begin{cases} +\infty & \text{if } \xi(x) = 0, \\ 0 & \text{if } \xi(x) = \xi_{\max}, \\ \to 0 & \text{if } \gamma \to 0 \text{ and } \xi(x) > 0. \end{cases}$$

Here $\gamma > 0$ is a perturbation parameter which needs to be small. The effect of the supplementary noise is then the reduction of the information from points, for which the measure ξ is small, thereby fulfilling the properties a) and b) from above. Further $\xi_{\max} = \max_{x \in \mathcal{X}} \xi(x)$.

4 The information matrix of the perturbed process and its properties.

Evidently, the information matrix of the perturbed process in the diagram has the form

$$I^{(\gamma)}(\xi) = \sigma^{-2} \sum_{x \in S_\xi} \sum_{z \in S_\xi} f(x) \left[C(S_\xi) + \gamma \text{diag} \left\{ \ln \frac{\xi_{\max}}{\xi(x)} \right\} \right]^{-1}_{x,z} f^T(z). \qquad (4)$$

To prove the continuity of the mapping $\xi \to I^{(\gamma)}(\xi)$, we use the following well known matrix lemma (refer e.g. to Kubáček et al. (1995)):

LEMMA 1. Let $U = \begin{pmatrix} U_I & U_{II} \\ U_{II}^T & U_{III} \end{pmatrix}$ be a symmetric, positive definite matrix, then

$$U^{-1} = \begin{pmatrix} U_1 & U_2 \\ U_2^T & U_3 \end{pmatrix},$$

with $U_! = (U_I - U_{II} U_{III}^{-1} U_{II}^T)^{-1}$, $U_2 = -(U_I - U_{II} U_{III}^{-1} U_{II}^T)^{-1} U_{II} U_{III}^{-1}$, and $U_3 = U_{III}^{-1} + U_{III}^{-1} U_{II}^T (U_I - U_{II} U_{III}^{-1} U_{II}^T)^{-1} U_{II} U_{III}^{-1}$. If U is positive semidefinite then each inverse can be substituted by a g-inverse.

PROPERTY 1. Let ξ_n, $n = 1, 2, \ldots$ be a sequence of design measures having the same support $S = S_{\xi_n}$ such that it converges to a design measure μ, i.e.

$$\lim_{n \to \infty} \xi_n(x) = \mu(x); \quad x \in \mathcal{X}.$$

Then for each $\gamma > 0$

$$\lim_{n \to \infty} I^{(\gamma)}(\xi_n) = I^{(\gamma)}(\mu).$$

Proof. Denote $S = \{x_1, \ldots, x_s\}$. We have $S_\mu \in S$, say $S_\mu = \{x_1, \ldots, x_a\}$, where $a \leq s$. If $a = s$, the proof is evident. If $a < s$, we shall do a step by step proof using the lemma from above. Take $U = C(S) + W^{(\gamma)}(\xi_n)$ with

$W^{(\gamma)}(\xi) = \text{diag}\left\{\ln \frac{\xi_{\max}}{\xi(x)}\right\}$ and denote U_I the upper left $(s-1)\times(s-1)$ block of this matrix. Then U_{III}^{-1} is a number

$$U_{III}^{-1} = \left[C(x_s, x_s) + \gamma \ln\left(\frac{\xi_{\max}}{\xi_n(x_s)}\right)\right]^{-1},$$

which tends to zero since $\xi_n(x_s) \to 0$. From Lemma 1 it follows that

$$\lim_{n\to\infty}[C(S) + W^{(\gamma)}(\xi_n)]^{-1} = \begin{pmatrix} \lim_{n\to\infty}\left[C(S-\{x_s\}) + [W^{(\gamma)}(\xi_n)]_I\right]^{-1} & 0 \\ 0 & 0 \end{pmatrix}.$$

We use the same procedure $s-a$ times to obtain finally

$$\lim_{n\to\infty}[C(S) + W^{(\gamma)}(\xi_n)]^{-1} = \begin{pmatrix} [C(S_\mu) + W^{(\gamma)}(\mu)]^{-1} & 0 \\ 0 & 0 \end{pmatrix}.$$

Note that $W^{(\gamma)}(\mu)$ is an $a \times a$ matrix. The proof is finished by using the definition of $I^{(\gamma)}(\mu)$. □

One can write equivalently with (4)

$$I^{(\gamma)}(\xi) = \sigma^{-2}\sum_{x\in\mathcal{X}}\sum_{z\in\mathcal{X}} f(x)\left[C(S_\xi) + \gamma\,\text{diag}\left\{\ln\frac{\xi_{\max}}{\xi(x)}\right\}\right]^{-1}_{x,z} f^T(z), \qquad (5)$$

since according to Property 1, any point x with measure $\xi(x) = 0$ is effectively excluded from the expression (5), which therefore gives (4).

PROPERTY 2. For every $\gamma > 0$ and every design A we have

$$I^{(\gamma)}(\xi_A) = M(A),$$

where ξ_A is a design measure which is uniform on A and zero outside.

Proof. To show this property just verify that $\xi_A(x) = (\xi_A)_{\max}$ for every $x \in A$ and use expression (4). □

PROPERTY 3. For every $\gamma > 0$ and every design measure ξ we have

$$I^{(\gamma)}(\xi) \le M(S_\xi)$$

in the Loewner ordering.

Proof. The proof follows from the fact that $M(S_\xi) = I^{(0)}(\xi)$, and that the noise $\epsilon_{\xi,\gamma}(.)$ diminishes the amount of information. □

The basic reason for introducing the scheme in the diagram is given by the following theorem, which states that we can use design measures ξ and the matrices $I^{(\gamma)}(\xi)$ instead of using exact designs A and the information

matrices $M(A)$ for solving (2). We denote by Ξ_N the set of all design measures supported by no more than N points. We employ a criterion function $\Phi[.]$ which is nonincreasing, i.e. $\Phi[M_1] \leq \Phi[M_2]$ if $M_1 \geq M_2$ in the Loewner ordering (the D-optimality criterion $\Phi[M] = -\ln\det M$ is a well-known example).

Theorem 1. Let N and $\gamma \geq 0$ be fixed and suppose that $\min_{A:N_A \leq N} \Phi[M(A)] < \infty$.

Then

a) If

$$\xi^{\#} \in \arg\min_{\xi \in \Xi_N} \Phi[I^{(\gamma)}(\xi)] \tag{6}$$

and $S = S_{\xi^{\#}}$, then also ξ_S solves (6) and

$$S \in \arg\min_{A:N_A \leq N} \Phi[M(A)]. \tag{7}$$

b) If $A^{\#}$ solves (7) then $\xi_{A^{\#}}$ solves (6).

Proof. If (6) holds then for any exact design A with $N_A \leq N$, we have

$$\Phi[M(S)] = \Phi[I^{(\gamma)}(\xi_S)] \leq \Phi[I^{(\gamma)}(\xi^{\#})] \leq \Phi[I^{(\gamma)}(\xi_A)] = \Phi[M(A)],$$

where we used Properties 2 and 3 - so (7) holds, and ξ_S solves (6).

Conversely, suppose that $A^{\#}$ solves (7). Then for every $\xi \in \Xi_N$ we have by Properties 2 and 3

$$\Phi[I^{(\gamma)}(\xi)] \geq \Phi[M(S_\xi)] \geq \Phi[M(A^{\#})] = \Phi[I^{(\gamma)}(\xi_{A^{\#}})],$$

so $\xi_{A^{\#}}$ is optimal in the sense of (6). □

Corollary 1. There is a uniform design measure (namely ξ_S), which solves (6), and its support (namely S) is a solution of (2).

Solving problem (6) presented in Theorem 1 still requires a minimization under the constraint $N_{S_\xi} \leq N$, which is discontinuous. To avoid this constraint we modify the variance of the supplementary noise in (3) by 'cutting' the design measure ξ at the level $\frac{1}{N}$. Then problem (7) corresponds to a minimization without constraints.

Let us denote $\kappa = \frac{1}{N}$, $\tilde{\xi} = \max\{0, \xi(x) - \kappa\}$ and $\hat{\xi}$ the normed $\tilde{\xi}$, i.e. $\hat{\xi}(x) = \frac{\tilde{\xi}(x)}{\sum_{z \in X} \tilde{\xi}(z)}$ for all nonzero measures $\tilde{\xi}$.

Then each $\hat{\xi}$ is in Ξ_N, and $S_{\tilde{\xi}} = S_{\hat{\xi}}$. We define $I^{(\gamma)}(\tilde{\xi})$ as in (4), but with $\tilde{\xi}$ instead of ξ and the variance $\sigma^{-2}\mathrm{Var}[\epsilon_{\xi,\gamma}(x)] = \gamma \ln\left[\frac{\max\{\tilde{\xi}_{\max}, \varepsilon\}}{\tilde{\xi}(x)}\right]$. Here $\varepsilon > 0$

is a small number for regularization. Note that

$$\lim_{\varepsilon \to 0} \gamma \ln \left[\frac{\max\left\{\tilde{\xi}_{\max}, \varepsilon\right\}}{\tilde{\xi}(x)} \right] = \begin{cases} \gamma \ln \left[\frac{\tilde{\xi}_{\max}}{\tilde{\xi}(x)} \right] & \tilde{\xi}_{\max} > 0 \\ \infty & \tilde{\xi}_{\max} = 0 \end{cases},$$

which means that designs with $\tilde{\xi}_{\max} = 0$ are excluded form the consideration. Consequently, we have $I^{(\gamma)}(\tilde{\xi}) = I^{(\gamma)}(\hat{\xi})$.

Corollary 2 of Theorem 1. If

$$\xi^* \in \arg\min_{\xi \in \Xi} \Phi[I^{(\gamma)}(\tilde{\xi})] \tag{8}$$

and $S = S_{\tilde{\xi}^*}$, then also the uniform design measure ξ_S solves (8), and S is a solution of (7). Conversely, if A^* solves (7), then ξ_{A^*} solves (8).

Proof. Reformulate Theorem 1 in terms of $\hat{\xi}$ instead of ξ. Note that Property 2 still holds for the relevant case that the number of support points for ξ is less than N. We do not need the constraint on $\hat{\xi}$ in (6), since it is fulfilled automatically. Now rewrite the result in terms of $\tilde{\xi}$ instead of $\hat{\xi}$. Since $I^{(\gamma)}(\tilde{\xi}) = I^{(\gamma)}(\hat{\xi})$ and since $S_{\tilde{\xi}} = S_{\hat{\xi}}$ for every ξ, we obtain the required result. □

Theorem 1 and Corollary 2 are justifying a numerical computation of optimum designs by continuous methods (e.g. gradient algorithms), as was until now only available in the uncorrelated case. The remaining difficulty in the construction of such procedures is that the operation of maximum appearing in $\xi_{\max}$ and $\tilde{\xi}$ must be approximated in a smooth (differentiable) way. For the situation covered by Corollary 2 (or rather a minor modification of it) a working version of an algorithm can be found in Müller and Pázman (1999).

References

BRIMKULOV, U., KRUG, G. & SAVANOV, V. (1980). Numerical construction of exact experimental designs when the measurements are correlated (in Russian). *Zavodskaya Laboratoria (Industrial Laboratory)* **36**, 435–442.

KIEFER, J. (1959). Optimal experimental designs (with discussion). *Journal of the Royal Statistical Society* **B**, 272–319.

KUBÁČEK, L., KUBÁČKOVA, L. AND VOLAUFOVÁ, J. (1995). *Statistical Models with Linear Structures.* Veda Publishing House.

MÜLLER, W.G. (1998). *Collecting Spatial Data.* Contributions to Statistics. Physica Verlag, Heidelberg.

MÜLLER, W.G. & PÁZMAN, A. (1999). An algorithm for the computation of optimum designs under a given covariance structure. *Computational Statistics* **14**, 197–211.

PÁZMAN, A. & MÜLLER, W.G. (1998). A new interpretation of design measures. In Atkinson, A., Pronzato, L. & Wynn, H., editors, *Model-Oriented Data Analysis 5*, 239–246, Physica Verlag, Heidelberg.

PUKELSHEIM, F. (1993). *Optimal Design of Experiments.* John Wiley & Sons, Inc., New York.

SACKS, J. & YLVISAKER, D. (1966). Design for regression problems with correlated errors. *Annals of Mathematical Statistics* **37**, 66–89.