Biometrika (2024), **00**, 0, *pp.* 1–5
https://doi.org/10.1093/biomet/asae001
Advance Access publication 18 January 2024

A note on minimax robustness of designs against correlated or heteroscedastic responses

By D. P. WIENS

Department of Mathematical and Statistical Sciences, University of Alberta, Edmonton, Alberta T6G 2G1, Canada
doug.wiens@ualberta.ca

Summary

We present a result according to which certain functions of covariance matrices are maximized at scalar multiples of the identity matrix. This is used to show that experimental designs that are optimal under an assumption of independent, homoscedastic responses can be minimax robust, in broad classes of alternate covariance structures. In particular, it can justify the common practice of disregarding possible dependence, or heteroscedasticity, at the design stage of an experiment.

Some key words: Correlation; Covariance; Induced matrix norm; Loewner ordering; Minimax design; Robustness.

1. Introduction

Experimental designs are typically derived, or chosen, assuming that the responses will be independent and homoscedastic. As well as being simple, this is almost necessary, unless an alternate covariance structure is somehow known. This is frequently a complicating feature of design theory; until a design is constructed and implemented, there are no data that can be used to estimate the model. There is some consolation however if a proposed design, optimal under an assumption of independence and homoscedasticity, is *minimax* against a broad class of alternate structures. By this we mean that the maximum loss in such a class is minimized by this design, which is thus robust against these alternatives. In this note we establish that any function of covariance matrices, possessing a natural monotonicity property, is maximized within such classes at a scalar multiple of the identity matrix. This can be paraphrased by saying that the least favourable covariance structure is that of independence and homoscedasticity. These are eventualities for which the proposed design is optimal; hence, it is minimax.

2. Main result

Suppose that $\|\cdot\|_M$ is a matrix norm, induced by the vector norm $\|\cdot\|_V$, i.e.,

$$\|C\|_M = \sup_{\|x\|_V = 1} \|Cx\|_V. \tag{1}$$

We use the subscript M when referring to an arbitrary matrix norm, but adopt special notation in the following cases.

(i) For the Euclidean norm $\|x\|_V = (x^{\mathrm{T}}x)^{1/2}$, the matrix norm is denoted $\|C\|_E$ and is the spectral radius, i.e., the root of the maximum eigenvalue of $C^{\mathrm{T}}C$. This is the maximum eigenvalue of C if C is a covariance matrix, i.e., is symmetric and positive semidefinite.

(ii) For the sup norm $\|x\|_V = \max_i |x_i|$, the matrix norm $\|C\|_\infty$ is $\max_i \sum_j |c_{ij}|$, the maximum absolute row sum.
(iii) For the 1-norm $\|x\|_V = \sum_i |x_i|$, the matrix norm $\|C\|_1$ is $\max_j \sum_i |c_{ij}|$, the maximum absolute column sum. For symmetric matrices, $\|C\|_1 = \|C\|_\infty$.

See Todd (1977, Ch. 3) for verifications of (i)–(iii).

We require the following properties of induced norms.

Property 1. It holds that $\|I\|_M = 1$.

Property 2. For covariance matrices C, $\|C\|_M \geqslant \|C\|_E$.

Property 1 is immediate from (1). For Property 2, suppose that $\|C\|_M$ is induced by $\|\cdot\|_V$, and that λ_0 is the maximum eigenvalue of C, with eigenvector x_0 normalized so that $\|x_0\|_V = 1$. Then

$$\|C\|_E = \lambda_0 = \|\lambda_0 x_0\|_V = \|Cx_0\|_V \leqslant \sup_{\|x\|_V=1} \|Cx\|_V = \|C\|_M.$$

Now suppose that the loss function in a statistical problem is $\mathcal{L}(C)$, where C is an $N \times N$ covariance matrix and $\mathcal{L}(\cdot)$ is nondecreasing in the Loewner ordering:

$$A \preceq B \implies \mathcal{L}(A) \leqslant \mathcal{L}(B). \tag{2}$$

Here $A \preceq B$ means that $B - A \succeq 0$, i.e., is positive semidefinite.

LEMMA 1. *For $\eta^2 > 0$, covariance matrix C and induced norm $\|C\|_M$, define*

$$\mathcal{C}_M = \{C \mid C \succeq 0 \text{ and } \|C\|_M \leqslant \eta^2\}.$$

For the norm $\|\cdot\|_E$, an equivalent definition is

$$\mathcal{C}_E = \{C \mid 0 \preceq C \preceq \eta^2 I_N\}. \tag{3}$$

Then

(i) *in any such class $\mathcal{C}_M$,* $\max_{\mathcal{C}_M} \mathcal{L}(C) = \mathcal{L}(\eta^2 I_N)$*,*
(ii) *if $\mathcal{C}' \subseteq \mathcal{C}_M$ and $\eta^2 I_N \in \mathcal{C}'$,* $\max_{\mathcal{C}'} \mathcal{L}(C) = \mathcal{L}(\eta^2 I_N)$.

Proof. We first establish the equivalence of (3). If the condition there holds then, by Weyl's monotonicity theorem (Bhatia, 1997, p. 63), all eigenvalues of C are dominated by η^2; hence, $\|C\|_E \leqslant \eta^2$. Conversely, if $\|C\|_E \leqslant \eta^2$ then all eigenvalues of C are dominated by η^2; hence, all those of $\eta^2 I_N - C$ are nonnegative and so the condition in (3) holds.

By (2) and (3), $\max_{\mathcal{C}_E} \mathcal{L}(C) = \mathcal{L}(\eta^2 I_N)$. Then, by Property 1 followed by Property 2, $\eta^2 I_N \in \mathcal{C}_M \subseteq \mathcal{C}_E$, and so the maximizer in the larger class is a member of the smaller class; hence, a fortiori the maximizer there. This proves (i). The proof of (ii) uses the same idea; the maximizer in the larger class $\mathcal{C}_M$ is a member of the smaller class $\mathcal{C}'$. □

Remark 1. An interpretation of Lemma 1 is as follows. Suppose that one has derived a technique under an assumption of uncorrelated, homoscedastic errors, i.e., a covariance matrix $\sigma^2 I_N$, which is optimal in the sense of minimizing $\mathcal{L}$, for any $\sigma^2 > 0$. Now suppose that one is concerned that the covariance matrix might instead be a member C of $\mathcal{C}_M$, and that $\mathcal{L}$ is monotonic in the sense described above. Then the technique minimizes $\max_{\mathcal{C}_M} \mathcal{L}(C) = \mathcal{L}(\eta^2 I_N)$, i.e., is minimax in $\mathcal{C}_M$.

Remark 2. In Remark 1 we implicitly assume that $\eta^2 \geqslant \sigma^2$; otherwise, $\mathcal{C}_M$ does not contain $\sigma^2 I_N$. An argument for taking $\eta^2 > \sigma^2$ arises if one assumes homoscedasticity and writes $C = \sigma^2 P$, where

P is a correlation matrix. Then in $\mathcal{C}_1$, $\eta^2 \geqslant \|C\|_1 = \sigma^2\|P\|_1 \geqslant \sigma^2$, with the final inequality being an equality if and only if $P = I_N$. Thus, take $\eta^2 > \sigma^2$. Then an intuitive explanation of the lemma is that, in determining a least favourable covariance structure, one can alter the correlations in some manner that increases $\|C\|_M$, or one can merely increase the variances. In fact, one should always just increase the variances.

Remark 3. A version of Lemma 1 was used by Wiens & Zhou (2008) in a maximization problem related to the planning of field experiments. It was rediscovered by Welsh & Wiens (2013) in a study of model-based sampling procedures. This note seems to be the first systematic study of the design implications of the lemma.

3. Applications

3.1. *Experimental design in the linear model*

Consider the linear model $y = X\theta + \varepsilon$. Suppose that the random errors ε have covariance matrix $C \in \mathcal{C}_M$. If C is known then the best linear unbiased estimate is $\hat{\theta}_{\text{BLUE}} = (X^{\mathrm{T}}C^{-1}X)^{-1}X^{\mathrm{T}}C^{-1}y$ and there is an extensive design literature; see Dette et al. (2015) for a review. In the more common case that the covariances are only vaguely known, or perhaps only suspected, it is more usual to use the ordinary least-squares estimate $\hat{\theta}_{\text{OLS}}$, design as though the errors are uncorrelated and hope for the best. An implication of the results of this section is that, in a minimax sense, that approach can be sensible.

In the classical alphabetic design problems, one seeks to minimize a function Φ of the covariance matrix of the regression estimates. Let ξ_0 be the minimizing design, under the possibly erroneous assumption of uncorrelated, homoscedastic errors. Assume that, under ξ_0, the moment matrix $X^{\mathrm{T}}X$ is nonsingular. Then the covariance matrix of $\hat{\theta}_{\text{OLS}}$ is

$$\operatorname{cov}(\hat{\theta}_{ols} \mid C) = (X^{\mathrm{T}}X)^{-1}X^{\mathrm{T}}CX(X^{\mathrm{T}}X)^{-1}. \tag{4}$$

Suppose that $0 \preceq C_1 \preceq C_2$, so that $C_2 - C_1 = A^{\mathrm{T}}A$ for some A. Then

$$\operatorname{cov}(\hat{\theta}_{ols} \mid C_2) - \operatorname{cov}(\hat{\theta}_{ols} \mid C_1) = B^{\mathrm{T}}B \succeq 0$$

for $B = AX(X^{\mathrm{T}}X)^{-1}$; hence, $\operatorname{cov}(\hat{\theta}_{ols} \mid C_1) \preceq \operatorname{cov}(\hat{\theta}_{ols} \mid C_2)$. It follows that if Φ is nondecreasing in the Loewner ordering then $\mathcal{L}(C) = \Phi\{\operatorname{cov}(\hat{\theta}_{ols} \mid C)\}$ is also nondecreasing and the conclusions of the lemma hold. Then, as in Remark 1, ξ_0 is a minimax design; it minimizes the maximum loss as the covariance structure varies over $\mathcal{C}_M$.

Again by Weyl's monotonicity theorem, if $0 \preceq \Sigma_1 \preceq \Sigma_2$ then the ith largest eigenvalue λ_i of Σ_2 dominates that of Σ_1 for all i. It follows that Φ is nondecreasing in the Loewner ordering if

(i) $\Phi(\Sigma) = \operatorname{tr}(\Sigma) = \sum_i \lambda_i(\Sigma)$, corresponding to A-optimality;
(ii) $\Phi(\Sigma) = \det(\Sigma) = \prod_i \lambda_i(\Sigma)$, corresponding to D-optimality;
(iii) $\Phi(\Sigma) = \max_i \lambda_i(\Sigma)$, corresponding to E-optimality;
(iv) $\Phi(\Sigma) = \operatorname{tr}(L\Sigma)$ for $L \succeq 0$, corresponding to L-optimality and including I-optimality, minimizing the integrated variance of the predictions. Thus, the designs optimal under any of these criteria are minimax.

Example 1 (*moving-average-of-order*-1 *errors*). As a particular case, assume first that the random errors are homoscedastic, but are possibly serially correlated, following a moving-average-of-order-1, MA(1), model with $\operatorname{corr}(\varepsilon_i, \varepsilon_j) = \rho I(|i-j| = 1)$ and with $|\rho| \leqslant \rho_{\max}$. Then, under this structure, C varies over the subclass $\mathcal{C}'$ of $\mathcal{C}_\infty$ defined by $c_{ij} = 0$ if $|i-j| > 1$ and $\|C\|_\infty \leqslant \sigma^2(1 + 2\rho_{\max})$, which we define to be η^2. Since $\eta^2 I_N \in \mathcal{C}'$, Lemma 1(ii) applies and it follows that ξ_0 is a minimax design in $\mathcal{C}'$ and with respect to any of the alphabetic criteria above. If the errors are instead heteroscedastic then σ^2 is replaced by the maximum of the variances.

Example 2 (*autoregressive-of-order*-1 *errors*). It is known that the eigenvalues of an autoregressive-of-order-1, AR(1), autocorrelation matrix with autocorrelation parameter ρ are bounded, and that the maximum eigenvalue $\lambda(\rho)$ has $\lambda^* = \max_\rho \lambda(\rho) > \lambda(0) = 1$. See, for instance, Trench (1999, p. 182). Then, again under homoscedasticity, the covariance matrix C has $\|C\|_E \leqslant \sigma^2\lambda^*$, and a design optimal when $\rho = 0$ is minimax in the subclass of $\mathcal{C}_E$ defined by the autocorrelation structure and $\eta^2 = \sigma^2\lambda^*$.

3.2. *Designs robust against model misspecifications*

Working in finite design spaces χ and with p-dimensional regressors $f(x)$, Wiens (2018) derived minimax designs for possibly misspecified regression models

$$Y(x) = f^{\mathrm{T}}(x)\theta + \psi(x) + \varepsilon, \tag{5}$$

with the unknown contaminant ψ ranging over a class Ψ and satisfying, for identifiability of θ, the orthogonality condition

$$\sum_{x\in\chi} f(x)\psi(x) = 0_{p\times 1}. \tag{6}$$

For designs ξ placing mass ξ_i on $x_i \in \chi$, he took $\hat\theta = \hat\theta_{\text{OLS}}$,

$$I(\psi,\xi) = \sum_{x\in\chi} E[f^{\mathrm{T}}(x)\hat\theta - E\{Y(x)\}]^2, \tag{7}$$

$$D(\psi,\xi) = [\det E\{(\hat\theta-\theta)(\hat\theta-\theta)^{\mathrm{T}}\}]^{1/p}, \tag{8}$$

and found designs minimizing the maximum, over ψ, of these loss functions. The random errors ε_i were assumed to be independent and identically distributed; now suppose that they instead have covariance matrix $C \in \mathcal{C}_M$.

Consider first (7). Using (6), and with $d_\psi = \{E(\hat\theta) - \theta\}$, which does not depend on the covariance structure, this decomposes as

$$\mathcal{I}(\psi,\xi,C) = \sum_{x\in\chi} f^{\mathrm{T}}(x)\mathrm{cov}(\hat\theta \mid C) f(x) + \sum_{x\in\chi}\{f^{\mathrm{T}}(x)d_\psi d_\psi^{\mathrm{T}} f(x) + \psi^2(x)\}. \tag{9}$$

The first sum above does not depend on ψ; the second depends on ψ, but not on the covariance structure. Then an extended minimax problem is to find designs ξ minimizing

$$\max_{\psi,C}\mathcal{I}(\psi,\xi,C) = \max_{C\in\mathcal{C}_M}\sum_{x\in\chi} f^{\mathrm{T}}(x)\mathrm{cov}(\hat\theta \mid C) f(x) + \max_{\psi\in\Psi}\sum_{x\in\chi}\{f^{\mathrm{T}}(x)d_\psi d_\psi^{\mathrm{T}} f(x) + \psi^2(x)\}.$$

As in § 3.1, and taking $L = \sum_{x\in\chi} f(x)f^{\mathrm{T}}(x)$ in (iv) of that section, for $C \in \mathcal{C}_M$, the first sum is maximized by a multiple η^2 of the identity matrix, and then the remainder of the minimax problem is that which was solved by Wiens (2018). The minimax designs, termed I-robust designs, obtained there do not depend on the value of η^2, and so enjoy the extended property of minimizing $\max_{\psi,C}\mathcal{I}(\psi,\xi,C)$ for $C \in \mathcal{C}_M$.

Now consider (8). The analogue of (9) is

$$\mathcal{D}(\psi,\xi,C) = [\det\{\mathrm{cov}(\hat\theta \mid C) + d_\psi d_\psi^{\mathrm{T}}\}]^{1/p}.$$

Since $\mathrm{cov}(\hat\theta \mid C) + d_\psi d_\psi^{\mathrm{T}}$, hence its determinant, is nondecreasing in the Loewner ordering, $\mathcal{D}(\psi,\xi,C)$ is maximized for $C \in \mathcal{C}_M$ by a multiple of the identity matrix. The rest of the argument is identical to that in the preceding paragraph, and so the D-robust designs obtained by Wiens (2018) also minimize $\max_{\psi,C}\mathcal{D}(\psi,\xi,C)$ for $C \in \mathcal{C}_M$.

Remark 4. Results in the same vein as those above have been obtained in cases that do not seem to be covered by Lemma 1. For instance, Wiens & Zhou (1996) sought minimax designs for the misspecification model (5), under conditions on the spectral density of the error process. They stated that '… a design which is asymptotically (minimax) optimal for uncorrelated errors retains its optimality under autocorrelation if the design points are a random sample, or a random permutation, of points …', with details in their Theorems 2.4 and 2.5.

3.3. *Designs for nonlinear regression models*

In the nonlinear regression model $y = f(x;\theta) + \varepsilon$, the goal of a design is typically the minimization of some function of the covariance matrix of $\hat{\theta}$, after a linearization of the model. When the errors have covariance C, the target of this minimization continues to be given by the right-hand side of (4), but with matrix X replaced by the gradient $F(\theta) = \{\partial f(x_i;\theta)/\partial\theta_j\}_{i,j}$, and with $F(\theta)$ evaluated at an initial parameter estimate, or at a previous estimate in an iterative estimation scheme. Then the results of § 3.1 continue to hold, and a design optimal for the loss functions given there, under independence and homoscedasticity, is minimax. A caveat, however, is that optimal design theory for nonlinear models is much less well developed, and more model specific, than that for linear models. See Bates & Watts (1988) and Hamilton & Watts (1985) for discussions.

Acknowledgement

This work was carried out with the support of the Natural Sciences and Engineering Research Council of Canada. It has benefited from the helpful comments of two referees.

References

Bates, D. & Watts, D. (1988). *Nonlinear Regression Analysis and Its Applications*. New York: John Wiley.

Bhatia, R. (1997). *Matrix Analysis*. Berlin: Springer.

Dette, H., Pepelyshev, A. & Zhigljavsky, A. (2015). Designs for linear regression models with correlated errors. In *Handbook of Design and Analysis of Experiments*, Ed. A. Dean, M. Morris, J. Stufken and D. Bingham, pp. 236–76. Boca Raton, FL: CRC Press.

Hamilton, D. C. & Watts, D. G (1985). A quadratic design criterion for precise estimation in nonlinear regression models. *Technometrics* **27**, 241–50.

Todd, J. (1977). *Basic Numerical Mathematics*, vol. 2. Stuttgart: Birkhäuser.

Trench, W. F. (1999). Asymptotic distribution of the spectra of a class of generalized Kac-Murdock-Szegö matrices. *Lin. Algeb. Appl.* **294**, 181–92.

Welsh, A. H. & Wiens, D. P. (2013). Robust model-based sampling designs. *Statist. Comp.* **23**, 689–701.

Wiens, D. P. (2018). I-robust and D-robust designs on a finite design space. *Statist. Comp.* **28**, 241–58.

Wiens, D. P. & Zhou, J. (1996). Minimax regression designs for approximately linear models with autocorrelated errors. *J. Statist. Plan. Infer.* **55**, 95–106.

Wiens, D. P. & Zhou, J. (2008). Robust estimators and designs for field experiments. *J. Statist. Plan. Infer.* **138**, 93–104.

[*Received on* 30 *September* 2023. *Editorial decision on* 2 *January* 2024]