PART III. Optimal design theory
(LECTURE 4)

Shpilev Petr Valerievich
Faculty of Mathematics and Mechanics, SPbU

September, 2025

Санкт-Петербургский государственный университет

Comments

In this lecture we continue the study of optimal design theory by focusing on the equivalence theorems. We begin with the sensitivity function and continue the proof of the Kiefer–Wolfowitz theorem, the fundamental result which provides the key condition for verifying D-optimality of a design. Several examples will illustrate how these conditions can be applied in practice, including discrete designs.

We then extend the equivalence framework to other criteria — E-, L-, and e_k- optimality — covering both nonsingular and singular cases. Special attention is given to extremal polynomials and block-diagonalization techniques, which make the verification of optimality more tractable.

Finally, we discuss analytical solutions of optimal design problems, showing how the theory leads to explicit D-optimal designs in polynomial and trigonometric regression models. This lecture thus connects the abstract equivalence principles with concrete methods for constructing and testing optimal experimental designs.

Lemma 15

For any nonsingular design ξ, the following inequality holds:

$$\sup_x d(x, \xi) \geq m.$$

Proof:

$$\sup_x d(x,\xi) \geq \int d(x,\xi)\,\xi(dx) = \operatorname{tr}\left[M^{-1}(\xi)\int f(x)f^T(x)\,\xi(dx)\right]$$

$$= \operatorname{tr}\left(M^{-1}(\xi)M(\xi)\right) = \operatorname{tr} I_m = m. \quad \square$$

Explanation

- The function $d(x,\xi) = f^T(x)M^{-1}(\xi)f(x)$ is known as the ***sensitivity function***. It represents the variance of the predicted response at point x.
- This lemma establishes a ***fundamental lower bound*** (m) for the maximum variance over the entire design space.
- This result is a crucial component of the ***Kiefer–Wolfowitz Equivalence Theorem***, which uses the condition $\sup d(x,\xi) = m$ as a test for D-optimality.

Comments

This lemma gives us a simple but important inequality: for any nonsingular design, the maximum of the sensitivity function is always at least equal to the number of parameters in the model, denoted by m. The sensitivity function is defined as the transpose of the regression vector at point x multiplied by the inverse of the information matrix times the regression vector again.

In the proof, we use Jensen's inequality for the supremum: the supremum of a function is always greater than or equal to its integral with respect to any probability measure. We apply this to the sensitivity function integrated over the design measure. The integral of the outer product of the regression vector gives us back the information matrix. Then we simply take the trace of the identity matrix, which gives us m.

This inequality tells us that no matter how we choose the design, we cannot reduce the maximum prediction variance below the number of parameters. This is especially important in the context of D-optimal designs. In fact, one of the central results of the Kiefer–Wolfowitz Equivalence Theorem is that D-optimality is achieved when this supremum actually equals m. So this lemma provides a necessary lower bound and builds the foundation for what follows.

Proof of the Kiefer–Wolfowitz Theorem:

The scheme of the proof is: (a) $\Rightarrow$ (c) $\Rightarrow$ (b) $\Rightarrow$ (a)

1. (a) $\Rightarrow$ (c) **:** Let ξ_D be a D-optimal design. Consider the design $\xi_\alpha = (1-\alpha)\xi_D + \alpha\xi_x$, where $\xi_x = \{x, 1\}$ is a design concentrated at point x, and $0 < \alpha < 1$.

Due to the D-optimality of ξ_D, the following inequality holds:

$$\ln\det M(\xi_\alpha) \le \ln\det M(\xi_D) \Leftrightarrow \frac{\ln\det M(\xi_\alpha) - \ln\det M(\xi_D)}{\alpha} \le 0$$

Passing to the limit as $\alpha \to 0+$, we obtain: $\frac{\partial}{\partial\alpha}\ln\det M(\xi_\alpha)\Big|_{\alpha=0+} =$

$$= \operatorname{tr} M^{-1}(\xi_\alpha)\,\frac{\partial M(\xi_\alpha)}{\partial\alpha}\Big|_{\alpha=0+} = \operatorname{tr} M^{-1}(\xi_\alpha)\,\frac{\partial((1-\alpha)M(\xi_D) + \alpha M(\xi_x))}{\partial\alpha}\Big|_{\alpha=0+} =$$

$$= \operatorname{tr} M^{-1}(\xi_D)(f(x)f^T(x) - M(\xi_D)) = d(x,\xi_D) - m \le 0$$

Hence, $d(x,\xi_D) \le m$ for all x, and therefore, $\sup_x d(x,\xi_D) = m$.

Comments

We begin the proof of the Kiefer–Wolfowitz equivalence theorem by showing that condition (a) implies condition (c). The proof is constructed as a cycle of implications, and here we consider the first step.

We assume that the design denoted by the Greek letter ξ with subscript D is optimal in the sense of the D-criterion. This means that it maximizes the logarithm of the determinant of the information matrix over all possible designs.

To test whether this property implies a certain inequality involving the sensitivity function, we perturb the optimal design slightly. We construct a new design by combining the original D-optimal design with a design concentrated at an arbitrary point x. This mixture is controlled by a small positive parameter α.

Since the original design is optimal, any such perturbation should not increase the logarithm of the determinant. Dividing the difference in the log-determinants by α and taking the limit as α approaches zero gives us the directional derivative of the objective function with respect to this perturbation.

Using the formula for the derivative of the logarithm of the determinant and applying the linearity of the matrix derivative, we arrive at an expression for the derivative in terms of the trace of the product of the inverse information matrix and the matrix difference between the outer product of the regression function and the original information matrix.

This expression equals the sensitivity function evaluated at point x minus the number of parameters. Since this must be less than or equal to zero, it follows that the sensitivity function does not exceed the number of parameters anywhere. Therefore, its supremum equals the number of parameters, as claimed in condition (c).

2. (c) $\Rightarrow$ (b)
Suppose the design ξ^* satisfies: $\sup_x d(x, \xi^*) = m$, then by Lemma 15, the design ξ^* is G-optimal.
3. (b) $\Rightarrow$ (a)
We prove by contradiction. Suppose that a G-optimal design exists which is not D-optimal. Let this design be ξ^* and let ξ_D be some D-optimal design.

Consider the design:

$$\xi_\alpha = (1-\alpha)\xi^* + \alpha\xi_D.$$

By Lemma 13, we have:

$$\ln\det M(\xi_\alpha) \geq (1-\alpha)\ln\det M(\xi^*) + \alpha\ln\det M(\xi_D).$$

$$\Rightarrow \quad \ln\det M(\xi_\alpha) - \ln\det M(\xi^*) \geq \alpha\left(\ln\det M(\xi_D) - \ln\det M(\xi^*)\right) > 0,$$

since ξ^* is not D-optimal.

$$\Rightarrow \quad \frac{\ln\det M(\xi_\alpha) - \ln\det M(\xi^*)}{\alpha} > 0.$$

Comments

In the second part of the proof, we show that condition (c) implies condition (b). If the supremum of the sensitivity function for a design equals the number of parameters, then, according to the referenced lemma, this design is optimal in the sense of the G-criterion. That completes the second implication.

Next, we establish the implication from condition (b) to condition (a) by contradiction. Assume there exists a design that is G-optimal but not D-optimal. Denote this G-optimal design by ξ^*, and let ξ_D be a design that is D-optimal.

We construct a new design as a convex combination of ξ^* and ξ_D, controlled by a small positive parameter α.

From the earlier lemma about the strict concavity of the logarithm of the determinant, it follows that the logarithm of the determinant of the information matrix of the combined design is greater than or equal to the convex combination of the logarithms of the determinants of the individual designs.

Subtracting the logarithm corresponding to the G-optimal design and using the assumption that it is not D-optimal, we conclude that the difference is strictly positive. Dividing this difference by α yields a positive quantity.

Proof of the Kiefer–Wolfowitz Theorem, conclusion

From the previous step, we get:

$$\frac{\partial}{\partial \alpha} \ln\det \mathrm{M}(\xi_\alpha)\Big|_{\alpha=0+} = \mathrm{tr}\left(\mathrm{M}^{-1}(\xi^*)\mathrm{M}(\xi_\mathrm{D})\right) - \mathrm{m} > 0, \Rightarrow \mathrm{tr}\left(\mathrm{M}^{-1}(\xi^*)\mathrm{M}(\xi_\mathrm{D})\right) > \mathrm{m}.$$

On the other hand, since the design ξ^* is G-optimal, we have:

$$\mathrm{f}^\mathrm{T}(\mathrm{x})\mathrm{M}^{-1}(\xi^*)\mathrm{f}(\mathrm{x}) \le \mathrm{m} \quad \forall \mathrm{x},$$

which implies:

$$\mathrm{tr}\left(\mathrm{M}^{-1}(\xi^*)\mathrm{M}(\xi_\mathrm{D})\right) = \int \mathrm{f}^\mathrm{T}(\mathrm{x})\mathrm{M}^{-1}(\xi^*)\mathrm{f}(\mathrm{x})\,\xi_\mathrm{D}(\mathrm{dx}) \le \mathrm{m}.$$

This is a contradiction $\Rightarrow$ G-optimal ξ^* is also D-optimal.
Therefore, all three conditions (a), (b), and (c) are equivalent.

- The maximum of $\mathrm{d}(\mathrm{x}, \xi)$ is attained at the support points of an optimal design (by Lemma 15).
- All optimal designs have the same information matrix (by Lemma 13).

The Kiefer–Wolfowitz Theorem is proved. □

Comments

We continue the proof by differentiating the logarithm of the determinant of the information matrix for the convex combination of the G-optimal and D-optimal designs. The derivative at zero from the right turns out to be strictly positive, which implies that the trace of the product of the inverse information matrix of the G-optimal design and the information matrix of the D-optimal design is strictly greater than the number of parameters.

However, because the original design is G-optimal, the sensitivity function is bounded above by the number of parameters for all points in the design space. Integrating this bound with respect to the D-optimal design gives a contradiction: the same trace must be less than or equal to the number of parameters.

This contradiction shows that any G-optimal design must also be D-optimal. As a result, we conclude the equivalence of all three conditions stated in the theorem.

The fact that the maximum of the sensitivity function is attained at the support points of the optimal design (in the case of finite support) follows directly from the earlier lemma, which becomes an equality in this case.

Finally, the fact that all optimal designs have identical information matrices follows from the concavity property of the logarithm of the determinant.

Example: an approximate design

Design and model

- Regression vector: $f(x) = (1,\ x,\ x^2)^T$
- Design space: $\chi = [-1,\ 1]$
- Candidate design:

$$\xi^* = \begin{pmatrix} -1 & 0 & 1 \\ \frac{1}{3} & \frac{1}{3} & \frac{1}{3} \end{pmatrix}$$

Information matrix and its inverse

- Compute $M(\xi^*) = \sum \omega_i f(x_i) f^T(x_i)$ and its inverse:

$$M(\xi^*) = \begin{pmatrix} 1 & 0 & \frac{2}{3} \\ 0 & \frac{2}{3} & 0 \\ \frac{2}{3} & 0 & \frac{2}{3} \end{pmatrix} \qquad M^{-1}(\xi^*) = D(\xi^*) = \begin{pmatrix} 3 & 0 & -3 \\ 0 & \frac{3}{2} & 0 \\ -3 & 0 & \frac{9}{2} \end{pmatrix}$$

Comments

This example demonstrates how we can use the Kiefer–Wolfowitz equivalence theorem to verify whether a specific design is D-optimal. We consider a quadratic regression model, where the regression vector consists of one, x, and x^2. The design space is the closed interval from -1 to 1.

Our candidate design is symmetric, with support points at -1, 0, and 1, each assigned equal weight — one third. We aim to verify whether this design satisfies the condition for D-optimality.

To do this, we start by computing the information matrix for this design. This matrix is obtained as the weighted sum of the outer products of the regression vectors at each design point. The resulting matrix is symmetric, with off-diagonal elements reflecting the mixed terms.

Next, we compute the inverse of this matrix, which we denote as $D(\xi^*)$. This inverse will be used in the next steps, where we evaluate the sensitivity function to test the equivalence condition.

K–W Theorem (proof)
Example: Checking D-optimality
Equivalence theorems
D-optimal designs

D-optimality criterion

To verify D-optimality of ξ^*, check:

$$d(x, \xi^*) = f^T(x)\, D(\xi^*)\, f(x) \le m = 3, \quad \forall x \in [-1, 1]$$

Evaluation of the sensitivity function

$$d(x, \xi^*) = \begin{pmatrix} 1 & x & x^2 \end{pmatrix} \begin{pmatrix} 3 & 0 & -3 \\ 0 & 1.5 & 0 \\ -3 & 0 & 4.5 \end{pmatrix} \begin{pmatrix} 1 \\ x \\ x^2 \end{pmatrix} = 3 + \frac{9}{2}x^2(x^2-1)$$

- The term $\frac{9}{2}x^2(x^2-1)$ is nonpositive for $x \in [-1, 1]$
- Therefore, $d(x, \xi^*) \le 3$ for all x in $[-1, 1]$
- The equality is attained at $x = -1,\ 0,\ 1$
- By the K–W Theorem, ξ^* is D-optimal

Remark

For discrete designs, the equivalence between D-optimality and G-optimality does not generally hold.

Comments

To check D-optimality of the design, we apply the equivalence theorem. According to it, we need to verify that the sensitivity function — that is, the transpose of the regression vector multiplied by the inverse information matrix and then by the regression vector again — does not exceed the number of parameters in the model. In our case, the number of parameters is three.

We compute this function explicitly. The result is a constant term three plus an additional term equal to $\frac{9}{2}$ times x^2 times the difference between x^2 and one. Since this second term is nonpositive over the interval from -1 to 1, the whole expression is maximized when x is -1, 0, or 1. At each of these points, the expression equals three.

Therefore, the function never exceeds three, and the equivalence condition is satisfied. This confirms that the design we are testing is indeed D-optimal.

Finally, a brief remark: for discrete designs, the equivalence between D-optimality and G-optimality does not generally hold. That is, a D-optimal design may fail to be G-optimal if it is discrete.

Example: a discrete design

Model setup

- Regression vector: $f(x) = (1\ x)^T$
- Design space: $\chi = [-1, 1]$
- Number of trials: $N = 3$

Candidate design structure

$$\xi^* = \begin{pmatrix} x_1 & x_2 & x_3 \\ \frac{1}{3} & \frac{1}{3} & \frac{1}{3} \end{pmatrix}$$

- At least two distinct support points to avoid degeneracy (by Theorem 7)

Information matrix

$$M(\xi^*) = \sum_{i=1}^{3} \frac{1}{3} f(x_i) f^T(x_i) = \begin{pmatrix} 1 & \frac{x_1+x_2+x_3}{3} \\ \frac{x_1+x_2+x_3}{3} & \frac{x_1^2+x_2^2+x_3^2}{3} \end{pmatrix}$$

Comments

In this example, we work with a simple linear model where the regression vector consists of one and x. The design space is the interval from −1 to 1. We are allowed to perform three measurements, and therefore we look for a D-optimal design with three equally weighted support points.

To avoid degeneracy of the information matrix, we assume that at least two of the support points are distinct. This requirement comes from the earlier theorem on properties of the information matrix, which states that when the number of a design support points is fewer than the number of parameters, its information matrix becomes rank-deficient.

Let us compute the information matrix corresponding to this design. It takes a compact symmetric form with the average of the support points appearing in the off-diagonal entries, and the average of the squared points appearing in the lower-right element. This matrix characterizes the total information collected under this design and will be the basis for our optimization.

Example (continued): a discrete design

Determinant of the information matrix

$$\det(M(\xi^*)) = \frac{2}{9}(x_1^2 + x_2^2 + x_3^2 - x_1x_2 - x_2x_3 - x_1x_3) = \\ = \frac{1}{9}\left[(x_1 - x_2)^2 + (x_1 - x_3)^2 + (x_2 - x_3)^2\right]$$

Optimization over design points

- Without loss of generality: $x_1 \le x_2 \le x_3$
- Maximum over x_1 attained at $x_1 = -1$; over x_3 — at $x_3 = 1$
- Resulting quadratic in x_2 achieves maximum at endpoints: $x_2 = -1$ or $x_2 = 1$

D-optimal designs

$$\xi_1^* = \begin{pmatrix} -1 & 1 \\ \frac{2}{3} & \frac{1}{3} \end{pmatrix}, \quad \xi_2^* = \begin{pmatrix} -1 & 1 \\ \frac{1}{3} & \frac{2}{3} \end{pmatrix}, \quad \det(M) = \frac{8}{9}$$

Comments

To identify the optimal design, we maximize the determinant of the information matrix. The expression for the determinant can be transformed into a sum of squared differences between the design points. This shows that the determinant grows when the design points are spread farther apart.

Assuming the design points are ordered from smallest to largest, we analyze the derivative of the determinant with respect to the outer points. It turns out the maximum is reached when the first point is -1 and the last point is 1.

Substituting these values reduces the optimization to a function of the middle point. Since this function is quadratic and concave, the maximum occurs at either endpoint, which means the second point must also be either -1 or 1.

This gives us two distinct but equivalent D-optimal designs, each supported at the endpoints of the interval but with different weight distributions. The determinant of the information matrix is the same for both designs and equals $\frac{8}{9}$.

Example (final): a discrete design

D-optimal designs: matrices and d-functions

$$M(\xi_1^*) = \begin{pmatrix} 1 & -\frac{1}{3} \\ -\frac{1}{3} & 1 \end{pmatrix}, \quad D(\xi_1^*) = \begin{pmatrix} \frac{9}{8} & \frac{3}{8} \\ \frac{3}{8} & \frac{9}{8} \end{pmatrix}$$

$$d(x, \xi_1^*) = f^T(x) D(\xi_1^*) f(x) = \frac{1}{8}(3x+1)^2 + 1$$

$$\max_{x \in [-1,1]} d(x, \xi_1^*) = 3, \quad \max_{x \in [-1,1]} d(x, \xi_2^*) = \frac{1}{8}(3x-1)^2 + 1 = 3$$

Comparison with a uniform plan

$$\overline{\xi} = \begin{pmatrix} -1 & 0 & 1 \\ \frac{1}{3} & \frac{1}{3} & \frac{1}{3} \end{pmatrix}, \quad M(\overline{\xi}) = \begin{pmatrix} 1 & 0 \\ 0 & \frac{2}{3} \end{pmatrix}, \quad D(\overline{\xi}) = \begin{pmatrix} 1 & 0 \\ 0 & \frac{3}{2} \end{pmatrix}$$

$$d(x, \overline{\xi}) = 1 + \frac{3}{2}x^2, \quad \max_{x \in [-1,1]} d(x, \overline{\xi}) = 2.5 < 3$$

Thus the discrete D-optimal design ξ_1^* and ξ_2^* are not G-optimal.

Comments

To analyze whether the D-optimal designs are also G-optimal, we compute the information matrix and its inverse for one of the D-optimal plans. Then we compute the sensitivity function d, which is the quadratic form involving the inverse information matrix and the regression vector. For the first D-optimal design, this function is a quadratic polynomial in the design variable, and its maximum over the interval equals three. The same maximum is obtained for the second D-optimal design.

We then compare these values with those of an alternative, uniformly weighted design supported at three equally spaced points. In this case, the information matrix and its inverse yield a simpler d-function, which is also quadratic. However, the maximum value of this function is only two and a half, strictly less than three.

This shows that although the D-optimal designs provide maximum information in terms of determinant, they do not minimize the maximum prediction variance across the design space. Hence, they are not G-optimal.

Equivalence theorem for the E-optimality criterion

- There are many versions of the Kiefer–Wolfowitz theorem for various optimality criteria.
- They are usually called equivalence theorems.
- Such theorems reduce the extremum problem to verifying a specific conditions for a function, simplifying the task and enabling design optimality checking.

Equivalence theorem for the E-criterion

Let $\mathcal{A}$ be the class of all positive semidefinite matrices A with $\operatorname{tr} A = 1$.
A design $\xi^* \in \Xi$ is E-optimal if and only if there exists $A^* \in \mathcal{A}$ such that for all $x \in \chi$,

$$f^T(x)A^*f(x) \le \lambda_{\min}(M(\xi^*)).$$

Moreover,

$$\min_{A\in\mathcal{A}} \max_{x\in\chi} f^T(x)Af(x) = \max_{\xi} \lambda_{\min}(M(\xi)),$$

$$f^T(x_i^*)A^*f(x_i^*) = \lambda_{\min}(M(\xi^*)),$$

where $x_i^*, i = 1, \ldots, n$ are the support points of the E-optimal design.

This result is given in Melas V.B., *E-optimal experimental designs*, SPbU, 1997.

Comments

There exists a general family of results known as equivalence theorems, which play a central role in the theory of optimal experimental designs. These theorems, stemming from the work of Kiefer and Wolfowitz, establish necessary and sufficient conditions for a design to be optimal under a given criterion. Rather than directly maximizing or minimizing the criterion, they reformulate the task as checking for certain conditions related to a specific function, which are called "extremal polynomials" in the literature on experimental design. This approach not only simplifies the optimization problem but also provides a powerful way to check whether a candidate design is indeed optimal.

We will consider several versions of the equivalence theorem for different optimality criteria, but without proofs. The logic and structure of the proof is similar in all cases; only technical details differ. We start with the E-optimality criterion, which minimizes the length of the longest axis of the confidence ellipsoid for the estimated parameters — equivalently, it maximizes the smallest eigenvalue of the information matrix.

The theorem states that an E-optimal design exists if and only if there is a matrix with unit trace, which majorizes all values of the corresponding quadratic form over the design space, and matches the minimum eigenvalue at the support points.

Setup for L-optimality criterion

Let $L = \sum_{i=1}^{k} l_i l_i^T$, where $l_i \in \mathbb{R}^m$. Define the class Ξ_L as the set of all designs for which the linear combinations $l_i^T\theta$, $i = 1, \dots, k$, are estimable.

Definition: class Ξ_L^*

A approximate design η belongs to the class Ξ_L^* if $\eta \in \Xi_L$ and for any approximate design ξ the limit

$$\lim_{\alpha \to 0} f^T(t)M^+(\xi_\alpha)LM^+(\xi_\alpha)f(t) = f^T(t)M^+(\eta)LM^+(\eta)f(t)$$

exists, where $\xi_\alpha = (1-\alpha)\eta + \alpha\xi$, $\alpha \in [0,1]$, and $M^+(\xi_\alpha)$ is the Moore–Penrose pseudoinverse of $M(\xi_\alpha)$.

Definition: L-optimal design

An approximate design $\xi^* \in \Xi_L^*$ is L-optimal if

$$\xi^* = \arg\min_{\xi \in \Xi_L^*} \mathrm{tr}(LM^+(\xi))$$

where L is a fixed nonnegative definite matrix. If ξ^* is nonsingular, then $M^+(\xi^*) = M^{-1}(\xi^*)$.

Comments

We now introduce the definitions necessary to formulate the equivalence theorem for the L-optimality criterion. First, we assume a symmetric nonnegative definite matrix L, defined as the sum of outer products of given vectors — that is, L equals the sum from $i = 1$ to k of the vector l_i times its transpose. The set Ξ_L consists of all continuous experimental designs for which the linear combinations of the parameters, namely $l_i^T\theta$, are estimable for all i.

Next, we define an extended class of designs, denoted Ξ_L^*. A design η belongs to this class if it is in Ξ_L and satisfies a specific continuity property: for any other design ξ, the limit as α goes to zero of the function $f^T(t)M^+(\xi_\alpha)LM^+(\xi_\alpha)f(t)$ exists and equals the corresponding expression evaluated at η. Here, the pseudoinverse is understood in the Moore–Penrose sense.

Finally, we define an approximate L-optimal design as one that minimizes the trace of L times the pseudoinverse of the information matrix over the class Ξ_L^*. If the design is nondegenerate, then the pseudoinverse coincides with the usual inverse of the information matrix.

We will separately consider the nonsingular and singular cases.

Comments

- Assume the existence of a nonsingular L-optimal design ξ^*
- If $\operatorname{rank} L = m$, then all L-optimal designs are nonsingular

Equivalence theorem for L-optimality

The following statements are equivalent:

(a) $\xi^* = \arg\min_{\xi \in \Xi_L} \operatorname{tr} LD(\xi)$

(b) $\max_{x \in \Xi_L} q(x, \xi^*) = \operatorname{tr} LD(\xi^*)$, where $q(x, \xi) = f^\top(x) D(\xi) LD(\xi) f(x)$.

Moreover, equality holds for all support points x_i of ξ^*:

$$q(x_i, \xi^*) = \operatorname{tr} LD(\xi^*)$$

This result can be found in Ermakov S.M., Zhiglyavsky A.A., Mathematical Theory of Optimal Experiment, Nauka, 1987. (Theorem 2.4, p. 112).

Comments

This theorem is a version of the equivalence theorem tailored for the L-optimality criterion, under the assumption that the optimal design is nonsingular. The assumption is reasonable whenever the rank of L equals m, the number of parameters in the model, since in that case all optimal designs are automatically nonsingular.

The equivalence theorem states that two conditions are equivalent: first, that the design minimizes the trace of L times the inverse of the information matrix, and second, that the maximum value of the function q, defined as the transpose of f times the inverse information matrix times L times the inverse again times f, over all design points equals that trace. Furthermore, this maximum is attained exactly at the support points of the optimal design.

It is important to highlight a major difference from the D-optimality criterion. In the D-optimal case, we always require nonsingular designs, since all parameters must be estimable. In contrast, for L-optimality, the focus is on estimating only certain linear combinations of the parameters. Therefore, we may allow singular designs, where fewer measurements than parameters are used. This flexibility makes the theory more subtle. That is why we will treat the nonsingular and singular cases separately: the singular case requires a more advanced treatment and leads to complications not present in the nonsingular setting.

Model setup

- Let $L = \sum_{i=1}^{k} l_i l_i^T$ with fixed $l_i \in \mathbb{R}^m$
- Assume an optimal design $\xi^* \in \Xi_L^*$ exists

Equivalence theorem for L-optimality (singular case)

For model (1o):

1) A design ξ belongs to Ξ_L if and only if for all $i = 1, \dots, k$:

$$l_i^T M^-(\xi) M(\xi) = l_i^T$$

2) A design $\xi^* \in \Xi_L^*$ is L-optimal if and only if:

$$\text{(a)} \quad \max_{t \in \mathcal{X}} \varphi(t, \xi^*) = \operatorname{tr}(LM^+(\xi^*)), \quad \varphi(t, \xi) = f^T(t) M^+(\xi) L M^+(\xi) f(t)$$

$$\text{(b)} \quad \varphi(t_i, \xi^*) = \operatorname{tr}(LM^+(\xi^*)), \quad t_i \in \operatorname{supp}(\xi^*)$$

This and the following result can be found in Shpilev P.V. *Equivalence Theorem for Singular L-Optimal Designs*, Vestnik St. Petersburg University. Mathematics, Vol. 48, No. 1, pp. 29–34, (2015).

Comments

We now turn to the equivalence theorem for the singular case. This is a more general and technically more complex situation, in which the information matrix may be singular. Such cases naturally arise when we are interested not in estimating all parameters, but only certain linear combinations — for instance, specific contrasts in regression.

We start with the structure of L, which is assumed to be the sum of outer products of fixed vectors l_i. These vectors specify the combinations of parameters that we care about. Under this structure, a design belongs to the class Ξ_L if and only if the identity $l_i^T M^-(\xi) M(\xi) = l_i^T$ holds for all i. This ensures that the design provides enough information to estimate the desired linear combinations.

The second part of the theorem characterizes L-optimal designs within the narrowed class Ξ_L^*. Such a design is optimal if and only if the maximum value of the function φ, which is the transpose of f times the pseudoinverse of M times L times the same pseudoinverse again times f, equals the trace of L times the pseudoinverse of the information matrix. Moreover, this maximum must be attained at all support points of the design.

This formulation generalizes the equivalence result to situations where the information matrix may be singular, and only certain linear estimators are required. It allows for designs with fewer support points than parameters, which makes the class of admissible designs significantly richer.

Notation for perturbed designs

Let $\xi \in \Xi_L$ be a singular design and let $\xi_\alpha = \alpha\eta + (1-\alpha)\xi$ with any design η such that ξ_α is nonsingular. Define:

$$\tilde{D}_{ij}(\xi) := \begin{cases} y_{ij}, & \text{if } 0 < |\lim\limits_{\alpha\to 0} M_{ij}^{-1}(\xi_\alpha)| < \infty \\ 0, & \text{otherwise} \end{cases}$$

$$\tilde{D}_{ij}^{+}(\xi) := \begin{cases} M_{ij}^{+}(\xi), & \text{if } \lim\limits_{\alpha\to 0} M_{ij}^{-1}(\xi_\alpha) = M_{ij}^{+}(\xi) \\ \tilde{D}_{ij}(\xi), & \text{otherwise} \end{cases}$$

Matrix $\tilde{D}^{+}(\xi)$ is symmetric, y_{ij} is unknow variables.

Theorem 8(Extremal polynomial in singular case)

Let $L \in \mathbb{R}^{m\times m}$ be fixed and positive semidefinite, and let $\xi^* \in \Xi_L \setminus \Xi_L^*$. Then the coefficients of the extremal polynomial $\overline{\varphi}(t,\xi^*) = f^T(t)\tilde{D}^{+}(\xi^*)L\tilde{D}^{+}(\xi^*)f(t)$ are the solution to the system:

$$\begin{cases} \overline{\varphi}'(t_i,\xi^*) = 0 \\ \overline{\varphi}(t_i,\xi^*) = \operatorname{tr}(LM^{+}(\xi^*)) \end{cases} \quad \text{for } t_i \in \operatorname{supp}(\xi^*)$$

Comments

We now address the most technically challenging scenario: when the optimal design is singular and does not belong to Ξ_L^*. This means that for design ξ_α from definition some specific elements of the Moore–Penrose pseudoinverse of the information matrix — namely, those involved in computing the coefficients of the extremal polynomial — do not coincide with the limiting values of the corresponding elements in the Moore–Penrose pseudoinverse for optimal design. Or in other words, relevant elements of pseudoinverse matrix are not uniquely determined by the limiting process.

To handle this ambiguity, we define a matrix with entries denoted by $\tilde{D}_{ij}(\xi)$. Each such entry equals y_{ij} — a symbolic unknown — if the absolute value of the limit of the ij element of the inverse of ξ_α exists, is nonzero, and finite, where ξ_α is defined as $\alpha\eta + (1-\alpha)\xi$, with α between zero and one, and η any design for which ξ_α is nonsingular. Otherwise, the corresponding entry is set to zero. The symbols y_{ij} are symmetric by construction, meaning y_{ij} equals y_{ji}.

We then define another matrix, $\tilde{D}^{+}(\xi)$, which selects either the corresponding element of the Moore–Penrose pseudoinverse or falls back to $\tilde{D}$ depending on whether the limiting value exists and matches. With this definition, we construct a generalized extremal polynomial — denoted by $\overline{\varphi}(t)$ — as the quadratic form $f^T(t)\tilde{D}^{+}(\xi^*)L\tilde{D}^{+}(\xi^*)f(t)$. The coefficients of this polynomial are to be determined as the solution of a system of equations evaluated at the support points of ξ^*.

K–W Theorem (proof)
Example: Checking D-optimality
Equivalence theorems
D-optimal designs

Remark

Let $\xi^* \in \Xi_L \setminus \Xi_L^*$. Then ξ^* is L-optimal if and only if the extremal polynomial $\overline{\varphi}(t, \xi^*)$ (as defined in Theorem 8) satisfies:

(a) $\max_{t \in \mathcal{X}} \overline{\varphi}(t, \xi^*) = \mathrm{tr}(LM^+(\xi^*))$

(b) $\overline{\varphi}(t_i, \xi^*) = \mathrm{tr}(LM^+(\xi^*)), \quad t_i \in \mathrm{supp}(\xi^*)$

The trigonometric regression model

Let us consider the third-order trigonometric regression model on $[-\pi, \pi]$:

$$y = \beta^T f(t) = \beta_0 + \beta_1 \sin(t) + \beta_2 \cos(t) + \beta_3 \sin(2t) + \beta_4 \cos(2t) + \\ + \beta_5 \sin(3t) + \beta_6 \cos(3t) + \varepsilon$$

Example: Singular design optimal for estimating the linear combination $\beta_2 + \beta_3$

The candidate to optimal design

Let us consider
a symmetric design: $\xi_\alpha = \begin{pmatrix} -\pi + x & -x & 0 & x & \pi - x \\ \frac{1-\alpha}{4} & \frac{1-\alpha}{4} & \alpha & \frac{1-\alpha}{4} & \frac{1-\alpha}{4} \end{pmatrix}, \quad \alpha \in [0, 1]$

Comments

In the singular case, when the design ξ^* belongs to the set Ξ_L but not to Ξ_L^*, the equivalence theorem remains applicable via the generalized extremal polynomial from Theorem 8. In this situation, the extremal polynomial defined using the matrix $\tilde{D}^+(\xi^*)$ must satisfy the same two conditions that appear in the equivalence theorem for the nonsingular case: its maximum over the design space must be equal to the trace of L times M^+, and it must attain this maximum at all support points of the design.

Below the remark, we begin an example that illustrates how this works in practice. We consider a third-order trigonometric regression model on the interval from $-\pi$ to π. The regression function includes the constant term, the sine and cosine of t, the sine and cosine of 2t, and the sine and cosine of 3t, making a total of seven parameters.

We consider an optimal design within a class of symmetric designs dependent on parameter α. The design's support points are symmetric about zero, with the central point assigned weight α, while the two outer pairs equally share the remaining weight. Since this design comprises only five support points, whereas the regression model has seven parameters, the information matrix remains singular for any value of α.

Structure of the information matrix

From the definition of the information matrix, it follows that for the design ξ_α we have:

$$M(\xi_\alpha) = \begin{pmatrix} 1 & 0 & 0 & 0 & m_{0,4} & 0 & 0 \\ 0 & m_{1,1} & 0 & 0 & 0 & m_{1,5} & 0 \\ 0 & 0 & m_{2,2} & 0 & 0 & 0 & m_{2,6} \\ 0 & 0 & 0 & m_{3,3} & 0 & 0 & 0 \\ m_{0,4} & 0 & 0 & 0 & m_{4,4} & 0 & 0 \\ 0 & m_{1,5} & 0 & 0 & 0 & m_{5,5} & 0 \\ 0 & 0 & m_{2,6} & 0 & 0 & 0 & m_{6,6} \end{pmatrix}$$

where the nonzero entries are given by:

$$m_{i,j} = \sum_{k=1}^{5} f_i(t_k) f_j(t_k) \omega_k$$

for t_k, ω_k being the support points and weights of the design ξ_α.

Comments

From the definition of the information matrix, we observe that due to the symmetry of the design, many cross-products of basis functions vanish when integrated with respect to the design measure. In particular, the basis function corresponding to the constant term interacts only with itself and with the second cosine term, that is, cosine of double argument. This is a direct consequence of the orthogonality relations between the trigonometric basis functions over symmetric support.

The resulting information matrix has a block-sparse structure, in which nonzero elements appear only for specific pairs of basis functions. The positions of the nonzero elements reflect which function pairs have nonzero scalar products on the design support. These scalar products are explicitly given by a weighted sum over the support points of the design: the nonzero entry in position i, j is the sum of the product of the ith and jth basis functions evaluated at those points, each weighted by the corresponding design weight.

The sparsity and structure of this matrix reflect the properties of the design and play a crucial role in computing its generalized inverse and the associated extremal polynomial.

Optimality criterion

To estimate β_2 and β_3, the optimal design minimizes $\operatorname{tr} \mathrm{L}\mathrm{M}^+(\xi)$

$$\text{with} \quad \mathrm{L} = \begin{pmatrix} 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 1 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 1 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 \end{pmatrix}$$

Block-diagonalization

The information matrix $\mathrm{M}(\xi_\alpha)$ can be block-diagonalized using a nonsingular symmetric matrix P:

$$\mathrm{P}\mathrm{M}(\xi_\alpha)\mathrm{P} = \begin{pmatrix} \bar{\mathrm{M}}(\xi_\alpha) & 0 \\ 0 & \bar{\mathrm{M}}_1(\xi_\alpha) \end{pmatrix}, \ \bar{\mathrm{M}}(\xi_\alpha) = \begin{pmatrix} \mathrm{m}_{2,2} & 0 & \mathrm{m}_{2,6} \\ 0 & \mathrm{m}_{3,3} & 0 \\ \mathrm{m}_{2,6} & 0 & \mathrm{m}_{6,6} \end{pmatrix}$$

Comments

The goal is to estimate the parameters β_2 and β_3. According to the general theory, the optimal design for this purpose minimizes the trace of the product of a selection matrix L and the Moore–Penrose inverse of the information matrix. Matrix L is constructed to extract the sum of variances of the least-squares estimators for β_2 and β_3, which correspond to the third and fourth diagonal elements of the pseudoinverse.

The specific structure of the information matrix corresponding to the design ξ_α makes it possible to construct a nonsingular symmetric matrix P that transforms the information matrix into a block-diagonal form. In this representation, the upper-left block is a three-by-three matrix that includes the elements corresponding to the third, fourth, and seventh columns and rows of the original matrix. These entries relate to the basis functions sine of t, sine of 2t, and cosine of 3t. These basis functions are precisely the ones involved in the estimation of β_2 and β_3, as well as any interactions between them and other components.

The remaining part of the information matrix, after transformation, corresponds to the components of the model that are not involved in the estimation of β_2 and β_3. As a result, they do not contribute to the trace of the selection matrix L times the pseudoinverse.

This block-diagonalization allows us to reduce the dimensionality of the optimization problem by focusing only on the relevant submatrix.

Equivalence theorem test for the design ξ^*

- For the symmetric design ξ_α, the optimization problem $\min\limits_{\xi} \operatorname{tr}(\mathrm{L}\mathrm{M}^+(\xi))$

$$\text{reduces to the problem } \min_{\xi} \operatorname{tr}(\bar{\mathrm{L}}\bar{\mathrm{M}}^+(\xi_\alpha)),\ \ \bar{\mathrm{L}} = \begin{pmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 0 \end{pmatrix}$$

- Test the design

$$\xi^* = \begin{pmatrix} -\frac{5\pi}{6} & -\frac{\pi}{6} & \frac{\pi}{6} & \frac{5\pi}{6} \\ \frac{1}{4} & \frac{1}{4} & \frac{1}{4} & \frac{1}{4} \end{pmatrix}$$

for optimality using the equivalence theorem.

- Let $\xi^*_\alpha = (1-\alpha)\xi^* + \alpha\eta$ where $\eta = \begin{pmatrix} 0 \\ 1 \end{pmatrix}$. Then

$$\lim_{\alpha\to 0} \mathrm{f}^{\mathrm{T}}(\mathrm{t})\bar{\mathrm{M}}^+(\xi^*_\alpha)\bar{\mathrm{L}}\bar{\mathrm{M}}^+(\xi^*_\alpha)\mathrm{f}(\mathrm{t}) = \frac{16}{9}\cos^2(\mathrm{t}) - \frac{32}{9}\cos(\mathrm{t})\cos(3\mathrm{t}) +$$

$$+\frac{16}{9}\sin^2(2\mathrm{t}) + \frac{16}{9}\cos^2(3\mathrm{t}) \neq \mathrm{f}^{\mathrm{T}}(\mathrm{t})\bar{\mathrm{M}}^+(\xi^*)\bar{\mathrm{L}}\bar{\mathrm{M}}^+(\xi^*)\mathrm{f}(\mathrm{t}) = \frac{16}{9}\cos^2(\mathrm{t}) + \frac{16}{9}\sin^2(2\mathrm{t})$$

- Therefore, $\xi^* \notin \Xi^*_{\mathrm{L}} \Rightarrow$ we cannot apply the equivalence theorem to check its optimality directly.

Comments

Having reduced the criterion to a trace expression involving only the relevant block of the information matrix, we now test whether a given symmetric design is optimal. The candidate design ξ^* is supported at four symmetric points with equal weights. To verify the optimality of the candidate design ξ^*, we construct a perturbed design ξ_α by mixing it with a point mass at zero. The aim of this construction is to ensure that the reduced information matrix remains nonsingular. This allows us to replace the Moore–Penrose pseudoinverse with the usual inverse and compute the limiting form of the extremal polynomial uniquely.

A comparison of the limiting extremal polynomial with the expression obtained by directly substituting ξ^* into the pseudoinverse matrix formula shows that the two results do not coincide. This discrepancy means that the design ξ^* does not belong to the class Ξ^*_{L}. As a result, we cannot apply the equivalence theorem to assess its optimality directly. We need to use theorem 8 first.

Matrices involved

For the candidate design ξ^*, the matrices $\tilde{D}^+(\xi^*)$ and $\bar{M}^+(\xi^*)$ are:

$$\tilde{D}^+(\xi^*) = \begin{pmatrix} \frac{4}{3} & 0 & y_{31} \\ 0 & \frac{4}{3} & 0 \\ y_{31} & 0 & 0 \end{pmatrix}, \quad \bar{M}^+(\xi^*) = \begin{pmatrix} \frac{4}{3} & 0 & 0 \\ 0 & \frac{4}{3} & 0 \\ 0 & 0 & 0 \end{pmatrix}$$

Extremal polynomial

The extremal polynomial from Theorem 8: $\overline{\varphi}(t, \xi^*) = f^T(t)\tilde{D}^+(\xi^*)\bar{L}\tilde{D}^+(\xi^*)f(t)$

$$= \frac{16}{9}\cos^2(t) - \frac{8}{3}y_{31}\cos(t)\cos(3t) + \frac{16}{9}\sin^2(2t) + y_{31}^2\cos^2(3t)$$

- From the condition $\overline{\varphi}'(\frac{\pi}{6}, \xi^*) = 0$ we find $y_{31} = \frac{2}{9}$.
- Then $\overline{\varphi}(t)$ achieves its maximum value $\frac{8}{3}$ at all support points of ξ^*:

$$\max_{t \in [-\pi, \pi]} \overline{\varphi}(t, \xi^*) = \overline{\varphi}\left(\pm\frac{\pi}{6}, \xi^*\right) = \overline{\varphi}\left(\pm\frac{5\pi}{6}, \xi^*\right) = \mathrm{tr}(\bar{L}\,\bar{M}^+(\xi^*)) = \frac{8}{3}$$

- Hence, ξ^* satisfies conditions (a) and (b) of the equivalence theorem and is L-optimal.

Comments

Now let us examine how we can still apply the equivalence theorem in this case. For this purpose, we use Theorem 8 and construct the extremal polynomial $\overline{\varphi}$ using the general inverse matrix $\tilde{D}^+(\xi^*)$. This matrix differs from the standard pseudoinverse in that it may contain symbolic variables—in our case, a single unknown variable denoted y_{31}. Using the system from Theorem 8, we determine the value of this variable. We find that y_{31} equals $\frac{2}{9}$.

It remains to verify that the constructed polynomial satisfies conditions (a) and (b) of the equivalence theorem i.e. that the resulting polynomial attains its maximum at all support points of the design and that this maximum matches the trace of the matrix L multiplied by the Moore–Penrose pseudoinverse of the reduced information matrix. Both conditions are fulfilled: the extremal polynomial peaks at $\pm\frac{\pi}{6}$ and $\pm\frac{5\pi}{6}$, and the maximum value equals $\frac{8}{3}$. Thus, the equivalence theorem confirms that this design is indeed L-optimal—even though its information matrix is singular.

Therefore, the design ξ^* is L-optimal.

Example: Extremal polynomial (Figure)

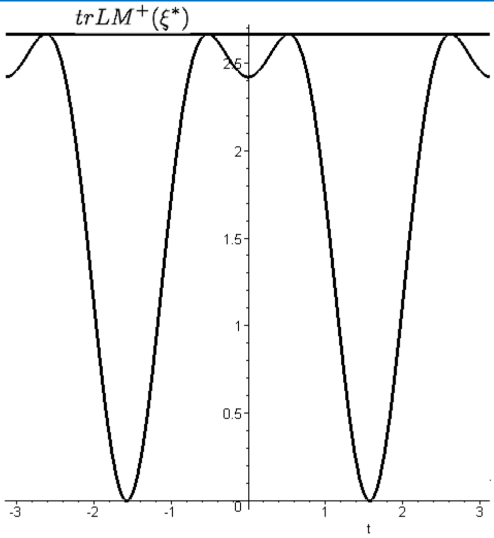

Figure: The behavior of the extremal polynomial $\overline{\varphi}(\mathrm{t}, \xi^*)$ on the interval $[-\pi, \pi]$.

Comments

Here we visualize the extremal polynomial for the optimal design ξ^*. The horizontal axis represents the design space — in this case, the interval from $-\pi$ to π. The vertical axis shows the value of the extremal polynomial over this interval.

Now, according to the general equivalence theorem, a design is optimal if and only if the extremal polynomial achieves its maximum exactly at the support points of the design. In our case, the support points are $\pm\frac{\pi}{6}$ and $\pm\frac{5\pi}{6}$. We can immediately spot these four points on the graph — they are precisely where the curve touches its highest level.

In practical terms, this plot provides visual confirmation that the design ξ^* is indeed optimal under the L-criterion, even though its information matrix is singular. This makes the extremal polynomial a powerful diagnostic tool — it translates an abstract matrix condition into a concrete and interpretable shape on the real line.

Equivalence theorem for the e_k-criterion

Let $\bar{f}_k(t)$ denote the vector obtained by removing $f_k(t)$ from $f(t) = (f_1(t), \dots, f_m(t))^T$, for some $k = 1, \dots, m$. Then a design ξ^* is optimal for estimating θ_k in model (1o) under assumptions (a)–(e) if and only if there exist $h > 0$ and $q \in \mathbb{R}^{m-1}$ such that the extremal polynomial

$$\varphi(t) = f_k(t) - q^T \bar{f}_k(t)$$

satisfies:

1) $h\varphi^2(t) \le 1$ for all $t \in \chi$;
2) $\operatorname{supp}(\xi^*) \subset \{t \in \chi \mid h\varphi^2(t) = 1\}$;
3) $\displaystyle\int_\chi \varphi(t)\bar{f}_k(t)\,\xi^*(dt) = 0 \in \mathbb{R}^{m-1}$.

In this case, $h = e_k^T M^-(\xi^*) e_k$.

Source: Dette, Melas, Pepelyshev (2004), *Optimal designs for estimating individual coefficients in polynomial regression – a functional approach.* JSP&I, 118, 201–219.

K–W Theorem (proof)
Example: Checking D-optimality
Equivalence theorems
D-optimal designs

Comments

We now present the equivalence theorem for the so-called e_k-criterion. This criterion focuses on the precision of estimating a single component of the parameter vector — namely, the k-th coordinate of θ. Instead of optimizing the estimation of all parameters together, as in most common design criteria, this approach focuses on just one parameter and tries to estimate it as precisely as possible.

The theorem provides necessary and sufficient conditions for a design ξ^* to be optimal for estimating θ_k. It states that such a design exists if and only if we can construct a function $\varphi(t)$, which is a linear combination of the components of $f(t)$, such that three conditions are met.

First, the function $h\varphi^2(t)$ must stay below or equal to one throughout the entire design space. Second, the support of the design must be concentrated exactly where this function reaches its maximum value of one. And third, a certain orthogonality condition must be satisfied — namely, the weighted average of $\varphi(t)$ times $\bar{f}_k(t)$ over the design must be zero.

The scalar h that appears in this condition turns out to be exactly the k-th diagonal entry of the generalized inverse information matrix. That is, h equals $e_k^T M^- e_k$. The function φ is called the extremal polynomial and plays the same certifying role as in the classical case.

This theorem offers a constructive way to verify optimality and to visualize it using extremal polynomials — especially in polynomial regression, where this form becomes explicit.

Comments

- In many classical models, optimal designs can be found analytically.
- We will now look at how to construct a D-optimal design for two widely used models:
  - Polynomial regression
  - Trigonometric regression

Polynomial regression model:

$$y_j = \theta^T f(x_j) + \varepsilon_j, \quad j = 1, \ldots, N, \qquad f(x) = (1, x, \ldots, x^{m-1})^T, \quad \chi = [-1, 1]$$

Theorem 9

For polynomial model on $[-1, 1]$, an unique approximate D-optimal design exists.

- It is supported with equal weights on m points.
- These points are the roots of the polynomial $(x^2 - 1)P'_{m-1}(x)$, where $P_{m-1}(x)$ is the Legendre polynomial of degree $m - 1$.

Comments

In many practical applications, equivalence theorems play a crucial role in simplifying the search for optimal designs. Instead of exploring the entire space of possible approximate designs, a more efficient approach is to focus on a well-structured subclass. For example, in the case we examined earlier, we restricted our search to 5-point symmetric designs as potential candidates. This constraint dramatically reduced the problem's complexity and made the optimization process more manageable. By confirming that the resulting design meets the equivalence conditions, we can be certain that it represents a global optimum—not just a local one.

Equivalence theorems often enable analytical derivation of optimal designs. As an illustration, consider D-optimal design construction for two commonly used models: polynomial and trigonometric regression. Let's begin with the polynomial case.

In this model, the regression functions consist of monomials: one, x, x^2, and so on, up to x^{m-1}. Under standard error assumptions, the D-optimal design on the interval is unique and has several remarkable properties. First, it is an approximate design supported at exactly m points. Second, these support points coincide with the roots of the polynomial $(x^2 - 1)P'_{m-1}(x)$ where $P_{m-1}(x)$ is the Legendre polynomial of degree $m - 1$. Finally, all m points receive equal weights.

This elegant result provides researchers with a powerful analytical tool for constructing optimal designs in polynomial regression settings.

Proof: D-optimal design for polynomial regression

Proof:

- Existence: Follows from the continuity of regressors and the compactness of the set of information matrices.
- Kiefer-Wolfowitz Theorem: We use this theorem to prove the remaining statements about the design.

Kiefer–Wolfowitz condition: the function $d(x, \xi^*)$ must satisfy:

$$\max_{x\in[-1,1]} d(x, \xi^*) = f^T(x)M^-1(\xi^*)f(x) = \sum_{i,j=1}^{m} d_{ij}(\xi^*)x^{i+j-2} = m$$

- $d(x, \xi^*)$ is a polynomial of degree $2m - 2$.
- The maxima occur at the support points of the optimal design.
- The design must have exactly m support points to avoid a singular information matrix.

The candidate for D-optimal design has the form:

$$\xi^* = \begin{pmatrix} x_1 & x_2 & \dots & x_m \\ w_1 & w_2 & \dots & w_m \end{pmatrix}$$

where $-1 = x_1 < x_2 < \ldots < x_m = 1$ and $\sum_i w_i = 1,\ w_i > 0$ for all i.

K–W Theorem (proof)

Example: Checking D-optimality

Equivalence theorems

D-optimal designs

Comments

Let's prove this theorem. The existence of a D-optimal design is guaranteed by the continuity of the regressors and the compactness of the design space. To characterize the structure of the optimal design, we apply the Kiefer–Wolfowitz equivalence theorem.

According to this theorem, a design ξ^* is D-optimal if and only if the maximum of the extremal polynomial $d(x, \xi^*)$, which is also known as a sensitivity function in the literature, does not exceed m, the number of model parameters, and this maximum is attained exactly at the support points of the design.

Due to the fact that the vector of regression functions $f(x)$ has the form: one, x, x^2, and so on, up to x^{m-1}, and the inverse information matrix for a given design ξ^* is simply a numerical matrix defining the coefficients of the polynomial d, the degree of this polynomial is equal to $2m - 2$. Since such a polynomial has at most m real extrema, the number of support points of the design cannot exceed m. At the same time, having fewer than m points leads to a singular information matrix, which is not allowed.

Hence, the D-optimal design must have exactly m support points, located strictly inside the interval and at its boundaries. Thus the candidate for the D-optimal design has the following form: It consists of m support points, x_1 through x_m, and their corresponding weights, w_1 through w_m. The points are ordered from -1 to 1, and all the weights are positive and sum up to one.

Proof: D-optimal design for polynomial regression (continued)

The information matrix of design ξ^* can be written as:

$$M(\xi^*) = \sum_{i=1}^{m} f(x_i) f^T(x_i) w_i$$

$$= \underbrace{\begin{pmatrix} 1 & \dots & 1 \\ x_1 & \dots & x_m \\ \vdots & & \vdots \\ x_1^{m-1} & \dots & x_m^{m-1} \end{pmatrix}}_{F} \underbrace{\begin{pmatrix} w_1 & 0 & \dots & 0 \\ 0 & w_2 & \dots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \dots & w_m \end{pmatrix}}_{W} \underbrace{\begin{pmatrix} 1 & x_1 & \dots & x_1^{m-1} \\ 1 & x_2 & \dots & x_2^{m-1} \\ \vdots & \vdots & & \vdots \\ 1 & x_m & \dots & x_m^{m-1} \end{pmatrix}}_{F^T}$$

From this factorization, we obtain:

$$\det(M(\xi^*)) = \det(FWF^T) = \left[\det(F^T)\right]^2 \cdot \prod_{i=1}^{m} w_i$$

Comments

By the definition, the information matrix $M(\xi^*)$ is equal to the sum over i from one to m of $f(x_i)f^T(x_i)w_i$.

We can write this as a matrix product. First, we define the matrix F, whose columns consist of the vectors $f(x_1)$ through $f(x_m)$. Since f consists of monomials one, x, x^2, and so on, F is the transposed Vandermonde matrix. Then we define a diagonal matrix W whose diagonal entries are the weights w_1 through w_m. Finally, F^T is the transpose of the first matrix. So the matrix $M(\xi^*)$ equals FWF^T.

By applying this identity, we compute the determinant of $M(\xi^*)$. The determinant evaluates to the square of the determinant of F^T, multiplied by the product of the weights. This result reveals an important property of D-optimal designs: the optimization problem naturally decomposes into two independent subproblems. Specifically, we can separately maximize the determinant with respect to (1) the support points and (2) their corresponding weights.

Proof: D-optimal design for polynomial regression (continued)

- The determinant $\det(F^T)$ is called the Vandermonde determinant and is calculated recursively. Let's denote

$$\triangle_{m-1} = \det(F^T).$$

- Then

$$\triangle_{m-1} = \begin{vmatrix} 1 & 0 & \dots & 0 \\ 1 & x_2 - x_1 & \dots & x_2^{m-1} - x_2^{m-2}x_1 \\ \vdots & \vdots & & \vdots \\ 1 & x_m - x_1 & \dots & x_m^{m-1} - x_m^{m-2}x_1 \end{vmatrix} = \prod_{i=2}^{m}(x_i - x_1) \cdot \triangle_{m-2}.$$

- Thus,

$$\triangle_{m-1} = \prod_{1 \le i < j \le m} (x_i - x_j).$$

- We find the values of x_i that maximize the expression $\triangle_{m-1}$. To do this, we differentiate with respect to the variable x_i, where $i = 2, \dots, m-1$, and set the result to zero:

$$\frac{\partial \triangle_{m-1}}{\partial x_i} = 0 \quad \Leftrightarrow \quad \frac{1}{x_i - x_1} + \dots + \frac{1}{x_i - x_{i-1}} + \frac{1}{x_i - x_{i+1}} + \dots + \frac{1}{x_i - x_m} = 0.$$

Comments

Let us now examine the structure of the determinant of F^T. This determinant is called the Vandermonde determinant and is denoted by $\triangle_{m-1}$. That is, $\triangle_{m-1}$ is equal to the determinant of F^T.

Using elementary row operations, we can express this determinant recursively. Specifically, $\triangle_{m-1}$ is equal to the determinant of a matrix where the first row is one, zero, up to zero, and each subsequent row contains one, followed by $x_i - x_1$, then $x_i^{m-1} - x_i^{m-2}x_1$, and so on. This allows us to factor out the product over i from two to m of the differences $x_i - x_1$. What remains is a smaller determinant, namely $\triangle_{m-2}$.

By recursively applying this reduction, we eventually obtain the well-known closed formula for the Vandermonde determinant. That is, $\triangle_{m-1}$ is equal to the product over all pairs $i < j$ from one to m of $x_i - x_j$. This product is symmetric and strictly positive whenever the x-values are distinct.

To find the support points that maximize this expression, we differentiate with respect to each inner point x_i, for i ranging from two to $m - 1$. Setting the derivative equal to zero gives a system of equations. Each equation takes the form: the sum over all $j \neq i$ of one divided by $x_i - x_j$ equals zero. This condition characterizes the so-called electrostatic equilibrium of the support points, and the solution to this system will lead us to the optimal support for the D-optimal design.

Differential equation for support polynomial

- Define the function:

$$\varphi(x) = (x - x_2)\dots(x - x_{m-1})$$

- The equilibrium conditions can be rewritten as:

$$\frac{1}{x_i+1} + \frac{1}{x_i-1} + \frac{\varphi''(x_i)}{2\varphi'(x_i)} = 0, \quad i = 2, \dots, m-1$$

- Equivalent form:

$$(x_i^2 - 1)\varphi''(x_i) + 4x_i\varphi'(x_i) = 0, \quad i = 2, \dots, m-1$$

- This expression has the same degree and roots as $\varphi(x)$, so:

$$(x^2 - 1)\varphi''(x) + 4x\varphi'(x) = \text{const} \cdot \varphi(x)$$

- Matching the coefficient at x^{m-2} yields:

$$\text{const} = m(m-1) - 2$$

- The differential equation:

$$(x^2 - 1)\varphi''(x) + 4x\varphi'(x) - (m(m-1) - 2)\varphi(x) = 0$$

Comments

In order to reduce this system to a differential equation we introduce a new function: $\varphi(x)$ equals the product over j from 2 to $m-1$ of $(x - x_j)$. This function captures the internal support points of the D-optimal design, excluding the boundary points -1 and 1.

Using this notation, we can rewrite the equilibrium conditions in a more compact form. For each i from 2 to $m-1$, the equation becomes: $\frac{1}{x_i+1} + \frac{1}{x_i-1} + \frac{1}{2}\frac{\varphi''(x_i)}{\varphi'(x_i)} = 0$.

This expression can be simplified algebraically to: $(x_i^2 - 1)\varphi''(x_i) + 4x_i\varphi'(x_i) = 0$.

This equation has the same degree and the same roots as $\varphi(x)$ itself. Therefore, the left-hand side must be equal to a constant times $\varphi(x)$. To determine this constant, we match the coefficients of x^{m-2} on both sides. This gives us that the constant is equal to $m(m-1) - 2$.

Thus, we obtain the desired differential equation for $\varphi(x)$: $(x^2-1)\varphi''(x) + 4x\varphi'(x) - (m(m-1) - 2)\varphi(x) = 0$.

Completion of the proof

- Consider the differential equation:

$$(x^2-1)P''(x)+2xP'(x)-m(m-1)P(x)=0,$$

where $P(x)$ is a polynomial of degree $m-1$.
- The unique solution to this is the Legendre polynomial of order $m-1$:

$$P_{m-1}(x)=\frac{1}{2^{m-1}(m-1)!}\frac{d^{m-1}}{dx^{m-1}}(x^2-1)^{m-1}$$

- Note that $P'_{m-1}(x)$ satisfies the derivative of the previous equation:

$$\left[(x^2-1)P''(x)+2xP'(x)-m(m-1)P(x)\right]'=0$$

$$\Leftrightarrow (x^2-1)P'''(x)+4xP''(x)-(m(m-1)-2)P'(x)=0$$

- Hence, $\overline{\varphi}(x)=P'_{m-1}(x)$ is a solution of the desired differential equation.
- All we have to do is find the weights. So we solve the system:

$$\begin{cases}\frac{\partial\prod\omega_i}{\partial\omega_i}=0, & i=1,\dots,m-1\\ \sum_{i=1}^{m}\omega_i=1\end{cases}\implies \omega_1=\dots=\omega_m=\frac{1}{m}$$

Theorem is proved. □

Comments

We now consider the differential equation: $(x^2-1)P''(x)+2xP'(x)-m(m-1)P(x)=0$. Here, $P(x)$ is a polynomial of degree $m-1$.

According to the theory of orthogonal polynomials, the unique solution up to a multiplicative constant is the Legendre polynomial of order $m-1$. This polynomial is given by $\frac{1}{2^{m-1}(m-1)!}\frac{d^{m-1}}{dx^{m-1}}(x^2-1)^{m-1}$.

Differentiating the original differential equation, we get a new equation involving the third derivative of P. This leads to: $(x^2-1)P'''(x)+4xP''(x)-(m(m-1)-2)P'(x)=0$.

This implies that the derivative of the Legendre polynomial, P'_{m-1}, satisfies the differential equation we previously obtained for φ.

Finally, solving the system of equations for the weights shows that all ω_i must be equal, since the only solution satisfying the product-maximization condition and the sum-to-one constraint is $\omega_i=\frac{1}{m}$ for all i.

The theorem is proved.

Let's consider D-optimal designs for the trigonometric regression model.

- The trigonometric regression model has the function:

$$\eta(x,\theta) = \theta_0 + \sum_{j=1}^{m} \theta_{2j-1}\sin(jx) + \sum_{j=1}^{m} \theta_{2j}\cos(jx),$$

where $\theta = (\theta_0, \theta_1, ...\theta_{2m})^T$ is the vector of unknown parameters.
- The regression function vector is $f(t) = (1, \sin t, \cos t, \ldots, \sin(mt), \cos(mt))^T$.
- The design interval is $\chi = [-\pi, \pi]$.

Theorem 10

Let $\chi = [-\pi, \pi]$. An approximate D-optimal design for the trigonometric regression model is any design:

$$\xi_N^* = \begin{pmatrix} t_1^* & \ldots & t_N^* \\ 1/N & \ldots & 1/N \end{pmatrix},$$

where $t_i^* = \frac{i-1}{N}2\pi - \pi$ for $i = 1, \ldots, N$, and $N \geq 2m+1$ (m is the order of the regression model).

The uniform design is also D-optimal: $\xi^* = \dfrac{1}{2\pi}dx$.

Comments

Let us now examine the D-optimal design for a different class of models — trigonometric regression.

In this case, the regression function is a trigonometric polynomial: it consists of the constant term θ_0, followed by alternating sine and cosine terms of increasing frequency. Specifically, we have $\sin(x)$, $\cos(x)$, $\sin(2x)$, $\cos(2x)$, and so on, up to $\sin(mx)$ and $\cos(mx)$. Altogether, there are $2m+1$ parameters to estimate, so the design must be able to support at least that many.

The design domain is the interval from $-\pi$ to π, which reflects the periodic nature of trigonometric functions.

The theorem on the slide gives a simple and elegant solution: if we take N equally spaced points on the interval from $-\pi$ to π, where N is at least $2m+1$, and assign equal weights to each point, then the resulting design is D-optimal. These support points are computed using the formula: $t_i^* = \frac{i-1}{N}2\pi - \pi$. So the points are uniformly spaced and symmetric.

This design distributes the observations uniformly over the interval, and thanks to the orthogonality of trigonometric functions on this domain, the information matrix turns out to be diagonal.

Moreover, the continuous uniform distribution on the interval from $-\pi$ to π is also D-optimal. That is, the design with density $\frac{1}{2\pi}$ also leads to the maximum determinant of the information matrix.

Let's prove this theorem.

Proof: D-optimal design for trigonometric model

Proof:

Due to the orthogonality of trigonometric functions, for any integers $i, j > 0$:

$$\int_{-\pi}^{\pi} \sin ix \, dx = \int_{-\pi}^{\pi} \cos jx \, dx = \int_{-\pi}^{\pi} \sin ix \cos jx \, dx = 0$$

The orthogonality conditions also yield:

$$\frac{1}{2\pi}\int_{-\pi}^{\pi} \sin^2 ix \, dx = \frac{1}{2\pi}\int_{-\pi}^{\pi} \cos^2 jx \, dx = \frac{1}{2}$$

Using these properties, the information matrix of the uniform design $\xi^* = \frac{1}{2\pi}dx$ is diagonal:

$$M(\xi^*) = \frac{1}{2\pi}\int_{-\pi}^{\pi} f(x)f^T(x)dx = \begin{pmatrix} 1 & 0 & \dots & 0 \\ 0 & \frac{1}{2} & \dots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \dots & \frac{1}{2} \end{pmatrix}.$$

Comments

Let's now go through the proof of this result. The key idea here is to exploit the orthogonality of trigonometric functions on the interval from $-\pi$ to π.

We begin with three basic integral identities. First, for any positive integer i, the integral from $-\pi$ to π of sin(ix) is zero. Second, for any positive integer j, the integral from $-\pi$ to π of cos(jx) is also zero. Third, the integral from $-\pi$ to π of the product of sin(ix) and cos(jx) is again zero.

These three equalities follow from the fact that sine and cosine functions are orthogonal over the interval from $-\pi$ to π. In particular, sine and cosine are orthogonal not just to each other, but also to the constant function, which explains why these integrals vanish.

Next, the orthogonality properties also tell us how the squared terms behave. If we integrate $\sin^2(ix)$ over the interval from $-\pi$ to π and divide by 2π, we get $\frac{1}{2}$. The same is true for $\cos^2(jx)$: its average over this interval is also $\frac{1}{2}$. This again holds for any positive integers i and j.

From this we automatically get that the information matrices of the designs defined in the conditions of the theorem are diagonal. The first element on the diagonal corresponds to the constant function and equals one. All the remaining diagonal elements correspond to sine and cosine terms and equal $\frac{1}{2}$. So the matrix has a one in the top-left corner and $\frac{1}{2}$ along the rest of the diagonal.

Proof (continued)

- The Kiefer–Wolfowitz theorem requires showing that the extremal polynomial $d(x,\xi^*) = f^T(x)D(\xi^*)f(x)$ satisfies $\max d(x,\xi^*) = 1+2m$.
- The inverse information matrix is:

$$D(\xi^*) = \begin{pmatrix} 1 & 0 & \dots & 0 \\ 0 & 2 & \dots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \dots & 2 \end{pmatrix},$$

  which gives the extremal polynomial

$$d(x,\xi^*) = 1 + 2(\sin^2 x + \cos^2 x + \ldots + \sin^2(mx) + \cos^2(mx)) = 1+2m.$$

- The optimality of the discrete design ξ_N^* follows from the fact that for any $j = 1,\ldots,2m$, the sums of the trigonometric functions at the design points are zero:

$$\sum_{i=1}^{N} \sin jt_i^* = 0, \quad \sum_{i=1}^{N} \cos jt_i^* = 0.$$

The theorem is proved. □

K–W Theorem (proof)

Example: Checking D-optimality

Equivalence theorems

D-optimal designs

Comments

To complete the proof, we apply the Kiefer–Wolfowitz equivalence theorem. We must show that the extremal polynomial, denoted $d(x,\xi^*)$, defined as $f^T(x)D(\xi^*)f(x)$, reaches its maximum value equal to $1+2m$.

Recall that $D(\xi^*)$ is the inverse of the information matrix. Since the original matrix is diagonal with one as the first entry and $\frac{1}{2}$ on all others, its inverse is also diagonal with entries one and two. So, $D(\xi^*)$ has one in the top-left corner, and twos along the remaining diagonal positions.

Substituting this into the definition of the extremal polynomial, we compute $d(x,\xi^*)$ as: one plus two times $\sin^2(x)$ plus two times $\cos^2(x)$, and so on, up to two times $\sin^2(mx)$ and two times $\cos^2(mx)$.

Each pair $\sin^2(jx) + \cos^2(jx)$ equals one. There are m such pairs, so the total becomes $1+2m$, as required. This proves that the continuous uniform design satisfies the optimality condition.

Finally, we show that the approximate design ξ_N^* with support points t_i^* — spaced uniformly on the interval from $-\pi$ to π — is also D-optimal. This follows from the fact that for every integer j from one to 2m, the sum of $\sin(jt_i^*)$ over all i equals zero, and similarly for the cosine terms.

Thus, the information matrix for the discrete design coincides with the continuous one. The theorem is proved.