Lecture #6 Notes Summary

Information Criteria and confidence ellipsoids

Information criteria

Definition 1 (Generalized Information matrix for $K^T\boldsymbol{\theta}$)**.** Let K be a matrix such that $K^T\boldsymbol{\theta}$ is estimable. The *generalized information matrix* of the design $\xi = \{\boldsymbol{x}, \boldsymbol{w}\} \in \Xi(K)$ for the estimation of the parameter subsystem $\zeta = K^T\boldsymbol{\theta}$ is

$$M_K(\xi) = \left(K^T M(\xi)^- K\right)^{-1}.$$

Proposition 1. *For all $\xi \in \Xi(K)$, the matrix $K^T M(\xi)^- K$ is invertible. As a consequence, $M_K(\xi)$ is well defined for (and only for) the interesting designs.*

Proof. Let $\xi \in \Xi(K)$. Let the columns of $U \in \mathbb{R}^{m\times r}$ form a base of $\operatorname{im} K$, so there exists an invertible $r \times r-$matrix H such that $K = UH$. Since the estimability condition $\operatorname{im} K \subseteq \operatorname{im} M(\xi)$ is satisfied, we can complete the base U to form a base of $M(\xi)$. That is, if $\operatorname{rank} M(\xi) = q$, there exists a matrix $V \in \mathbb{R}^{m\times(q-r)}$ such that the columns of $W := [U, V]$ form a base of $M(\xi)$ (so we have $U^T U = I_r, V^T V = I_{q-r}$, and $U^T V = 0$). This implies decomposition of $M(\xi)$ of the form $M(\xi) = W\Sigma W^T$, for an invertible matrix $\Sigma \succ 0$ of size $q \times q$. So we have

$$K^T M(\xi)^- K = K^T W\Sigma^{-1}W^T K = H^T U^T [U, V]\Sigma^{-1}[U, V]^T UH = H^T \underbrace{[I_r, 0]\Sigma^{-1}[I_r, 0]^T}_{r\times r \text{ principal subblock of } \Sigma^{-1}} H \succ 0.$$

□

Recall that we want to choose the design ξ so as to minimize the matrix $K^T M(\xi)^- K$. Equivalently, we want to maximize the generalized $K-$information matrix $M_K(\xi) = (K^T M(\xi)^- K)^{-1}$, because

$$0 \prec P \preceq Q \Longleftrightarrow P^{-1} \succeq Q^1 \succ 0,$$

see Exercise 1. In particular, when $K = I_m$ is the identity matrix of size $m\times m$, (i.e. we want to estimate the whole parameter $\boldsymbol{\theta}$), the problem is to maximize the information matrix $M(\xi)$. We have seen in an exercise that maximizing $M_K(\xi)$ with respect to $\preceq$ was the same has minimizing the ellipoid

$$\mathcal{E}_K(\xi) = \{\boldsymbol{x} : \boldsymbol{x}^T M_K(\xi)\boldsymbol{x} \leq 1\}$$

for the inclusion relation. In fact, $\mathcal{E}_K(\xi)$ can be seen as a confidence ellipsoid of the estimator $\hat{\boldsymbol{\theta}}$ (resp. of $\hat{\boldsymbol{\zeta}} = K^T\hat{\boldsymbol{\theta}}$ if $K \neq I_m$). Indeed, if the error on the measurements follows a normal distribution, it can be seen that $\hat{\boldsymbol{\zeta}} \sim \mathcal{N}\left(K^T\boldsymbol{\theta}, K^T M(\xi)^- K\right)$. Then, it is a standard property of the (multivariate) normal distribution that for all $\alpha \in (0, 1)$,

$$\mathbb{P}\left(\frac{\boldsymbol{\zeta} - \hat{\boldsymbol{\zeta}}}{\kappa_\alpha} \in \mathcal{E}_K(\xi)\right) = 1 - \alpha,$$

where κ_α is a constant depending only on the confidence level α (in fact, κ_α is the square root of the $\alpha-$quantile of a χ^2 distribution with r degrees of freedom, where r is the number of columns of K).

But when the matrix K has more than $r = 1$ columns, the problem of maximizing $M_K(\xi)$ is not well posed, because the Löwner ordering $\preceq$ is only a *partial order*. So we need to scalarize the problem, by considering a criterion $\Phi : \mathbb{S}_+^r \to \mathbb{R}$:

$$\max_{\xi\in\Xi(K)} \Phi\left(M_K(\xi)\right).$$

Definition 2 (Information criterion)**.** A continuous function $\Phi : \mathbb{S}_+^m \to \mathbb{R}$ is called *information criterion* if it satisfies the following properties:

1 *Positive homogeneity:*

$$\forall \alpha > 0,\ \forall M \in \mathbb{S}_+^m, \quad \Phi(\alpha M) = \alpha \Phi(M)$$

2 *Concavity:*

$$\forall \alpha \in [0, 1],\ \forall M, P \in \mathbb{S}_+^m, \quad \Phi(\alpha M + (1 - \alpha)P)) \geq \alpha\Phi(M) + (1 - \alpha)\Phi(P)$$

3 *Monotonicity with respect to the Löwner ordering:*

$$A \leq B \Longrightarrow \Phi(A) \leq \Phi(B)$$

We next introduce the family of information criterions introduced by Kiefer, and that contains the important cases of $D-$, $E-$, and $A-$optimality:

Definition 3 (Kiefer's $p-$optimality criterion)**.** Let $M \in \mathbb{S}_{++}^m$ with eigenvalues $0 < \lambda_1 \leq \ldots \leq \lambda_m$ (enumerated with multiplicities). For all $p \in [-\infty, 1]$, the Kiefer's $p-$criterion is:

$$\Phi_p(M) = \begin{cases} \lambda_{\min}(M) & = \lambda_1 & \text{if } p = -\infty \\ \det(M)^{\frac{1}{m}} & = \prod \lambda_i^{\frac{1}{m}} & \text{if } p = 0 \\ (\frac{1}{m} \operatorname{trace} M^p)^{\frac{1}{p}} & = (\frac{1}{m} \sum_i \lambda_i^p)^{\frac{1}{p}} & \text{otherwise.} \end{cases}$$

The definition of Φ_p is extended by continuity to singular matrices $M \in \mathbb{S}_+^m$, so that $\Phi_p(M) = 0$ if M is singular and $p \leq 0$.

We admit that Φ_p is an information criterion for all $p \leq 1$.

Definition 4 (optimal design)**.** A design ξ is called Φ_K*-optimal* if it maximizes $\Phi\big(M_K(\xi)\big)$ over $\xi(K)$. A design ξ is said to be an *exact* Φ_K*-optimal design of size* N if it maximizes $\Phi\big(M_K(\xi)\big)$ over

$$\xi_N(K) := \{\xi = \{\boldsymbol{x}, \boldsymbol{w}\} \in \Xi(K) : \forall i \in \{1, \ldots, s\},\ N w_i \in \mathbb{Z}_+\}.$$

It is always assumed that $K = I_m$ whenever the subscript K is ommited.

Definition 5 (D-criterion)**.** $\Phi_D := \Phi_0$ is called the *D-criterion.* A D_K-optimal design minimizes the volume of the confidence ellipsoid $\mathcal{E}_K(\xi)$, see Figure (a) below.

Proposition 2. *Let X be an invertible square matrix, and define $\boldsymbol{\theta'} = X\boldsymbol{\theta}$. Then, a design is $D-$optimal for $\boldsymbol{\theta'}$ iff it is $D-$optimal for $\boldsymbol{\theta}$. In other words, the $D-$optimal design is invariant to reparametrization.*

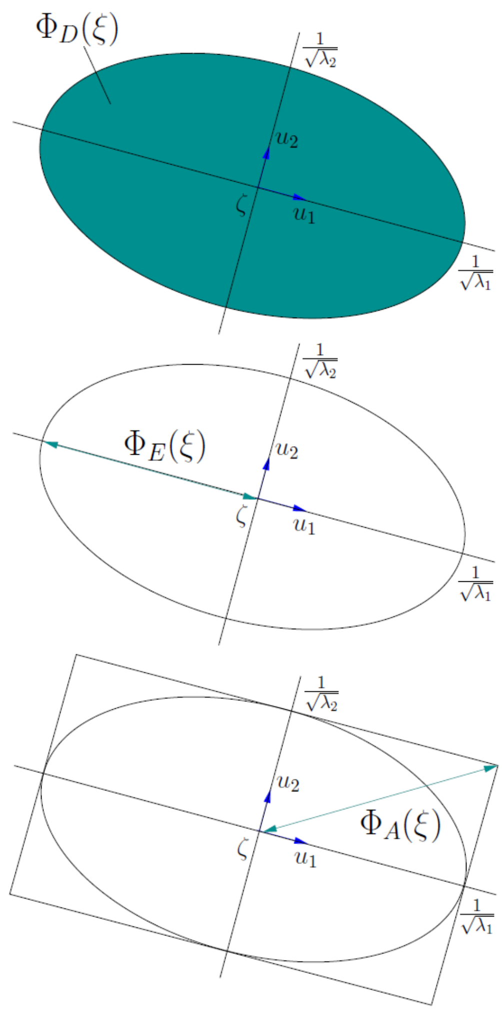
Proof. Since X is invertible, we have $\Xi(X) = \Xi(I_m)$, i.e. $M(\xi)$ is invertible for all feasible designs. Hence,

$$\Phi_D\big(M_X(\xi)\big) = \det\left(X^T M(\xi)^- X\right)^{-\frac{1}{m}} = \left((\det X)^2 (\det M(\xi))^{-1}\right)^{-\frac{1}{m}},$$

which is proportional to $\Phi_D\big(M(\xi)\big)$. □

Definition 6 (E-criterion)**.** $\Phi_E := \Phi_{-\infty}$ is called the *E-criterion.* An E_K-optimal design minimizes the largest radius of the confidence ellipsoid $\mathcal{E}_k(\xi)$, see Figure (b) below.

Definition 7 (A-criterion)**.** $\Phi_A := \Phi_{-1}$ is called the *A-criterion.* An A_K-optimal design minimizes the diagonal of the bounding box of the confidence ellipsoid $\mathcal{E}_K(\xi)$, see Figure (c) below.



(a)

(b)

(c)

Exercises

1. By using a Schur complement, show that $0 \prec P \preceq Q \iff P^{-1} \succeq Q^{-1} \succ 0$.

2. Let ξ be an $E-$optimal design. Show that ξ is *good* for the estimation of any linear combination of the form $\boldsymbol{c}^T\boldsymbol{\theta}$. More precisely, show that ξ minimizes the worst case of the variance $\boldsymbol{c}^T M(\xi)^- \boldsymbol{c}$, over all vectors $\boldsymbol{c} \in B_m := \{\boldsymbol{c} \in \mathbb{R}^m : \|\boldsymbol{c}\| = 1\}$. In other words, ξ minimizes $\Big(\max_{\boldsymbol{c}\in B_m} \boldsymbol{c}^T M(\xi)^- \boldsymbol{c}\Big)$.

3. The variance of the best estimator for $\boldsymbol{a}_i^T\boldsymbol{\theta}$ is $\lambda_i(\xi) := \sigma^2 \boldsymbol{a}_i^T M(\xi)^- \boldsymbol{a}_i$. Show that a design $\xi \in \cap_i \Xi(\boldsymbol{a}_i)$ minimizing a weighted sum of the variances $\sum_i \alpha_i \lambda_i(\xi)$ must be $A_K$optimal for some matrix K.

4. Show that $\mathcal{E}_K := \{\boldsymbol{y} : \boldsymbol{y}^T M_K(\xi)\boldsymbol{y} \leq 1\}$ is a projection of $\mathcal{E} := \{\boldsymbol{x} : \boldsymbol{x}^T M(\xi)\boldsymbol{x} \leq 1\}$ onto $\mathrm{im}(K^T)$, in the following sense:
$$\mathcal{E}_K = \{K^T\boldsymbol{x} : \ \boldsymbol{x} \in \mathcal{E}\}.$$