Lecture #4 Notes Summary

Linear model, information matrices

The linear model

We assume that the experimenter wants to estimate an unknown parameter $\boldsymbol{\theta} \in \mathbb{R}^m$. He can make a trial at each point $\boldsymbol{x}$ of a compact region $\mathcal{X}$ (the *experimental region*), representing the different experimental conditions.

Let us assume that trials are done at some points $\boldsymbol{x}_1, \boldsymbol{x}_2, \ldots, \boldsymbol{x}_N \in \mathcal{X}$. In the standard linear model, it is assumed that:

- The observation y_i at the i^{th} point x_i satisfies

$$\mathbb{E}[y_i] = \boldsymbol{a}(\boldsymbol{x}_i)^T \boldsymbol{\theta},\ \mathrm{var}(y_i) = \sigma^2,$$

  where $a : \mathcal{X} \mapsto \mathbb{R}^m$ is continuous.

- The N observatoins are mutually independent. In particular, the responses are uncorrelated:

$$i \neq j \Longrightarrow \mathbb{E}\left[\left(y_i - \boldsymbol{a}(\boldsymbol{x}_i)^T\boldsymbol{\theta}\right)\left(y_j - \boldsymbol{a}(\boldsymbol{x}_j)^T\boldsymbol{\theta}\right)\right] = 0.$$

The set of points $\boldsymbol{x}_1, \boldsymbol{x}_2, \ldots, \boldsymbol{x}_N \in \mathcal{X}$ is called the *design*. NB: there can be some replicate, e.g. $\boldsymbol{x}_i = \boldsymbol{x}_j$. Obviously, the order of the trials does not play a role, so we can assume without loss of generality that the points $\boldsymbol{x}_1, \ldots, \boldsymbol{x}_s$ are distinct for some $s \leq N$. Then, the design can be represented by using the following notation:

> **Definition 1** (Exact design)**.** An *exact design* ξ corresponds to a collection of points $\boldsymbol{x}_1, \ldots, \boldsymbol{x}_s \in \mathcal{X}$ together with a vector of integers $\boldsymbol{n} \in \mathbb{N}^s$. We write
>
> $$\xi = \begin{pmatrix} \boldsymbol{x}_1 & \cdots & \boldsymbol{x}_s \\ n_1 & \cdots & n_s \end{pmatrix},$$
>
> or $\xi = \{\boldsymbol{x}, \boldsymbol{n}\}$ for short, which means that n_i observations are taken at $\boldsymbol{x}_i$.

Note that for an exact design $\xi = \{\boldsymbol{x}, \boldsymbol{n}\}$ with a total number of N trials, we have $\sum_i n_i = N$. Now, in vector notation, we have

$$\mathbb{E}[\boldsymbol{y}] = A(\xi)\boldsymbol{\theta},$$
$$\mathrm{Var}[\boldsymbol{y}] = \sigma^2 \boldsymbol{I}_N,$$

where the i^{th} row of the $N \times m-$ matrix $A(\xi)$ is $\boldsymbol{a}(\boldsymbol{x}_i)$.

A direct application of the Gauss-Markov theorem shows that if the estimability condition $\mathrm{im}\, K \subseteq \mathrm{im}\, A(\xi)^T$ is satisfied, then we have

$$\mathrm{Var}[\hat{\boldsymbol{\zeta}}] \succeq \mathrm{Var}[K^T\boldsymbol{\theta}^*] = \sigma^2 K^T \left(A(\xi)^T A(\xi)\right)^- K,$$

for all unbiased estimator $\hat{\boldsymbol{\zeta}}$ of $\boldsymbol{\zeta} := K^T\boldsymbol{\theta}$, and where $\boldsymbol{\theta}^* := A(\xi)^\dagger \boldsymbol{y}$ is the least square estimate of $\boldsymbol{\theta}$. Observe that $A(\xi)^T A(\xi) = \sum_i n_i \boldsymbol{a}(\boldsymbol{x}_i)\boldsymbol{a}(\boldsymbol{x}_i)^T$, where $\boldsymbol{n}$ is such that $\sum_{i=1}^s n_i = N$.

Thus, we can give a first informal version of the optimal design problem: *Find a design $\xi = \{\boldsymbol{x}, \boldsymbol{n}\}$ which satisfies the estimability condition, and which minimizes (in a certain sense) the matrix*

$$K^T \left(\sum_{i=1}^{s} n_i \boldsymbol{a}(\boldsymbol{x}_i) \boldsymbol{a}(\boldsymbol{x}_i)^T \right)^{-} K$$

To get rid of the combinatorics of the problem, we can write $n_i = N w_i$ for some vector $\boldsymbol{w}$ in the probability simplex $\Delta_s := \{\boldsymbol{w} \in \mathbb{R}_+^s : \sum_i w_i = 1\}$, and we (temporarily) ignore the integer constraint $N w_i \in \mathbb{Z}_+$.

Definition 2 (Approximate design)**.** An *approximate design* ξ corresponds to a collection of points $\boldsymbol{x}_1, \ldots, \boldsymbol{x}_s \in \mathcal{X}$ together with a vector of weights $\boldsymbol{w} \in \Delta_s$. We write

$$\xi = \begin{pmatrix} \boldsymbol{x}_1 & \cdots & \boldsymbol{x}_s \\ w_1 & \cdots & w_s \end{pmatrix},$$

and w_i can be interpreted as the *percentage of experimental effort* spent at the design point $\boldsymbol{x}_i$.

Definition 3 (Information matrix)**.** The *information matrix* of the design $\xi = \{\boldsymbol{x}, \boldsymbol{w}\}$ is

$$M(\xi) = \sum_{i=1}^{s} w_i \boldsymbol{a}(\boldsymbol{x}_i) \boldsymbol{a}(\boldsymbol{x}_i)^T.$$

An anoying point of these definitions is that we do not know what should be the value of s for an approximate design. But we can use the following geometrical result:

Theorem 1 (Carathéodory)**.** *Let $P \subset \mathbb{R}^n$ and $\boldsymbol{x} \in$ convex-hull(P). Then, $\boldsymbol{x}$ can be written as the barycenter of $n+1$ points $\boldsymbol{a}_1, \ldots, \boldsymbol{a}_{n+1} \in P$. That is,*

$$\exists \boldsymbol{a}_1, \ldots, \boldsymbol{a}_{n+1} \in P, \quad \exists \boldsymbol{\lambda} \in \Delta_{n+1} : \quad \boldsymbol{x} = \sum_i \lambda_i \boldsymbol{a}_i.$$

Proposition 2. *Let $\xi = \{\boldsymbol{x}, \boldsymbol{w}\}$ be a design (exact or approximate) for a model with m parameters, supported by N points (i.e., $\boldsymbol{x}_1, \ldots, \boldsymbol{x}_N \in \mathbb{R}^m$). Then, there exists an approximate design $\xi' = \{\boldsymbol{x}, \boldsymbol{w}\}$ supported by at most $s = \frac{m(m+1)}{2} + 1$ points such that $M(\xi) = M(\xi')$.*

Remark. An alternative way to define designs uses the concept of probability measures: A design ξ can be seen as a probability measure over $\mathcal{X}$, in which case we define

$$M(\xi) = \int_{\boldsymbol{x} \in \mathcal{X}} \boldsymbol{a}(\boldsymbol{x}) \boldsymbol{a}(\boldsymbol{x})^T d\xi(\boldsymbol{x}).$$

In this case, Carathéodory's theorem implies that every information matrix of the above form can be obtained from a discrete measure with $s = \frac{m(m+1)}{2} + 1$ support points.

Remark. When the design region $\mathcal{X}$ is finite, or when the design points $\boldsymbol{x}_1, \ldots, \boldsymbol{x}_s$ are given, we simply represent the design by $\boldsymbol{w} \in \Delta_s$ and we write $\boldsymbol{a}_i$ instead of $\boldsymbol{a}(\boldsymbol{x}_i)$:

$$M(\boldsymbol{w}) = \sum_{i=1}^{s} w_i \boldsymbol{a}_i \boldsymbol{a}_i^T.$$

Exercises

1. Proof of Carathéodory's theorem.

   The proof is by induction. Assume that $\boldsymbol{x}$ is a barycenter of $N > n+1$ points $\boldsymbol{a}_1, \ldots, \boldsymbol{a}_N$.

   - Show that $\boldsymbol{a}_2 - \boldsymbol{a}_1, \ldots, \boldsymbol{a}_N - \boldsymbol{a}_1$ are linearly dependents, and deduce the existence of a vector $\boldsymbol{\mu} \neq \mathbf{0}$ such that $\sum_i \mu_i = 0$, $\sum_i \mu_i \boldsymbol{a}_i = \mathbf{0}$.
   - Use the above vector $\boldsymbol{\mu}$ to show that $\boldsymbol{x}$ is a barycenter of $N-1$ points.