Lecture #12 Notes Summary

Optimality of Block Designs

Optimality of BIBDs

In this lecture, we say that a binary design is a $(t, b, k)-$design if it is a block design on t treatments, with b blocks of size k. The set of all binary $(t, b, k)-$designs is denoted by $\mathcal{B}(t, b, k)$.

We are going to show that BIBDs are A-, D- and E-optimal in the class of binary designs with the same characteristic numbers (t, b, k). (In fact, they are also optimal in the class of all block designs (not necessarily binary), but the binary assumption is rather intuitive and simplifies the proofs.)

For our proofs, we next introduce the majorization order, which is a partial order on $\mathbb{R}^n$:

> **Definition 1** (Majorization)**.** Let $\boldsymbol{a}$ and $\boldsymbol{b}$ be two vectors of $\mathbb{R}^n$. We say that $\boldsymbol{a}$ *majorizes* $\boldsymbol{b}$, and we write $\boldsymbol{a} \succeq \boldsymbol{b}$, if
>
> - $\sum_i a_i = \sum_i b_i$;
> - $\forall k \in \{1, \ldots, n-1\}, \qquad \sum_{i=1}^k a_i^{\downarrow} \geq \sum_{i=1}^k b_i^{\downarrow}$,
>
> where $a_i^{\downarrow}$ denotes the i^th largest element of $\boldsymbol{a}$.

In $\mathbb{R}^3$, we have:

$$(6,0,0) \succeq (5,1,0) \succeq (3,3,0) \succeq (3,2,1) \succeq (2,2,2).$$

#1

We will not use the next theorem, but we give it because it is a nice and important result:

> **Theorem 1** (Hardy, Littlewood & Polya)**.** *Let $\boldsymbol{x}, \boldsymbol{y} \in \mathbb{R}^n$. We have $\boldsymbol{x} \preceq \boldsymbol{y}$ iff $\boldsymbol{x} = P\boldsymbol{y}$ for some* doubly stochastic matrix P *of size $n \times n$, that is, a matrix with row sums and column sums all equal to* 1.

We also point out that the set of doubly stochastic matrices is the convex hull of all permutation matrices, and that every doubly stochastic matrix of size $n \times n$ can be written as a barycenter of at most $n^2 - 2n + 2$ permutation matrices (this is Birkhoff's theorem).

> **Definition 2** (Schur-convex function)**.** A function $f : \mathcal{A} \subseteq \mathbb{R}^n \to \mathbb{R}$ is said to be *Schur-convex* iff for all $\boldsymbol{x}, \boldsymbol{y} \in \mathcal{A}$,
>
> $$\boldsymbol{x} \succeq \boldsymbol{y} \Longrightarrow f(\boldsymbol{x}) \geq f(\boldsymbol{y}).$$
>
> A function $f : \mathcal{A} \to \mathbb{R}$ is said to be *Schur-concave* iff $-f$ is Schur-convex.

The function $\boldsymbol{x} \to \max(\boldsymbol{x})$ is Schur-convex and $\boldsymbol{x} \to \min(\boldsymbol{x})$ is Schur-concave.

#2

Note that $\boldsymbol{x} \preceq P\boldsymbol{x} \preceq \boldsymbol{x}$ for all permutation matrices P. Hence it is clear if a function f is Schur-convex over a set $\mathcal{A}$ satisfying $(\boldsymbol{x} \in \mathcal{A} \Leftrightarrow P\boldsymbol{x} \in \mathcal{A})$, then f must be *symmetric*, that is $f(P\boldsymbol{x}) = f(\boldsymbol{x})$.

The next result makes it easy to check whether a function is Schur-convex:

Theorem 2 (Schur condition). *Let $\mathcal{A} \subseteq \mathbb{R}^n$ and let $f : \mathcal{A} \to \mathbb{R}$ be continuously differentiable. The function f is Schur-convex (resp. Schur-concave) on $\mathcal{A}$ if and only if f is* symmetric *(invariance to permutation), and for all $\boldsymbol{x} \in \mathcal{A}$,*

$$(x_1 - x_2)\left(\frac{\partial f}{\partial x_1} - \frac{\partial f}{\partial x_2}\right) \geq 0 \quad (\textit{resp.} \ \leq 0).$$

Proposition 3. *If $\mathcal{I} \subseteq \mathbb{R}$ is an interval and $g : \mathcal{I} \to \mathbb{R}$ is convex (resp. concave) and twice differentiable, then*

$$f : \boldsymbol{x} \to \sum_{i=1}^{n} g(x_i)$$

is Schur-convex (resp. Schur-concave) on $\mathcal{I}^n$.

Proof. f is clearly symmetric. Then, we check the Schur condition: Observe that

$$\frac{\partial f}{\partial x_1} - \frac{\partial f}{\partial x_2} = g'(x_1) - g'(x_2),$$

and since g is convex, g' is nondecreasing and $g'(x_1) - g'(x_2)$ has the same sign as $x_1 - x_2$. □

Corollary 4. *The following functions are Schur-concave over $\mathbb{R}^n_{++}$:*

- $\boldsymbol{x} \to \sum_i \log x_i$;
- *For all* $0 < p \leq 1, p \neq 0$, $\boldsymbol{x} \to \sum_i x_i^p$;

And for all $p < 0$, $\boldsymbol{x} \to \sum_i x_i^p$ is Schur-convex over $\mathbb{R}^n_{++}$.

Proposition 5. *Let C be the Information matrix of a design $\xi \in \mathcal{B}(t, b, k)$. Then,*

$$\text{trace } C = b(k-1).$$

Proof. For a binary design, the diagonal elements of Λ are $\lambda_{ii} = r_i$, and so the diagonal elements of $L = kR - \Lambda$ are equal to $(k-1)\times$ the replications r_i. We have $\sum_{i=1}^{t} r_i = bk$ (total number of plots). So

$$\operatorname{trace} C = \frac{1}{k} \operatorname{trace} L = \frac{1}{k}(k-1)\sum_{i=1}^{t} r_i = b(k-1).$$

□

Theorem 6. *Let the integers (t, b, k) be such that a BIBD ξ exists in $\mathcal{B}(t, b, k)$. Then, ξ is $A-$, $E-$, and $D-$ optimal in the class of all designs in $\mathcal{B}(t, b, k)$.*

Proof. The information matrix of a BIBD with replication r and concurrence λ is the completely symmetric matrix of size $t \times t$, with $\frac{k-1}{k} r$ on the diagonal and $-\frac{1}{k}\lambda$ on all off-diagonal coefficients. Its eigenvalues are

$\frac{k-1}{k}r - (t-1)\frac{1}{k}\lambda = 0$, and $\frac{k-1}{k}r + \frac{1}{k}\lambda = (r(k-1)+\lambda)\frac{1}{k}$ with multiplicity $(t-1)$. Using the relations $bk = tr$ and $r(k-1) = \lambda(t-1)$, the nonzero eigenvalues simplify to

$$\frac{1}{k}\left(bk\frac{(k-1)}{t} + bk\frac{(k-1)}{t(t-1)}\right) = b\frac{k-1}{t-1}.$$

Note that the nonzero eigenvalues sum to $b(k-1)$ in accordance with Proposition 5.

Now, let ξ' be another binary $(t,b,k)-$design, and denote by $\boldsymbol{\lambda}(\xi)$ the vector of nonzero eigenvalues of the information matrix of ξ. By proposition 5, $\boldsymbol{\lambda}(\xi)$ and $\boldsymbol{\lambda}(\xi')$ have the same sum, and $\boldsymbol{\lambda}(\xi) \preceq \boldsymbol{\lambda}(\xi')$ because all values of $\boldsymbol{\lambda}(\xi)$ are equal. So ξ maximizes $\Phi(\boldsymbol{\lambda}(\xi))$ over the class $\mathcal{B}(t,b,k)$ for all Schur-concave function Φ. For $E-$optimality, we take $\Phi = \max(\cdot)$; for $D-$optimality, we note that a design maximizing the geometric mean of $\boldsymbol{\lambda}$ also maximizes $\sum_i \log \lambda_i$; and for $A-$optimality we observe that a design maximizing the harmonic average of $\boldsymbol{\lambda}$ must minimize $\sum_i \frac{1}{\lambda_i}$. □

Optimality of sparse designs

When Levi Graph is a tree The Levi Graph has bk edges and $b+t$ vertives. So a design can be connected only if $bk \geq b+t-1$. The extreme case is $bk = b+t-1$, that is, $b(k-1) = t-1$, and the Levi Graph is a tree. So we cannot distinghish the designs using the D-criterion (there is always a single spanning tree). But the next proposition gives the form of the A- and E-optimal designs.

> **Proposition 7.** *If $b(k-1) = t-1$, then* queen-bee designs *are $A-$ and $E-$optimal. These are the designs in which one treatment is replicated b times (this treatment appear in all blocks), and each other treatment appears in a single block.*

Proof. Since the LG is a tree, the effective resistance between 2 nodes is equal to the distance between these nodes. Recall that $A-$optimal designs must minimize the sum of all pairwise resistances. In any tree-shaped LG, the resistance between two treatments that appear in the same block is 2, and otherwise it must be at least 4. These bounds are attained for the queen-bee design, because the distance between any two non-concurrent treatments is 4 in the LG.

The proof for $E-$optimality requires the *cut-set* lemma, which we will prove in exercises. □

When Levi Graph is unicyclic In the case $b(k-1) = t$, just slightly less extreme, the Levi Graph contains a single cycle.

> **Proposition 8.** *If $b(k-1) = t$, then the Levi-Graph of a $D-$optimal design must contain a big cycle of length $2b$.*

Proof. In a graph with a single cycle, the number of spanning trees is equal to the length of the cycle. In a $(b,t)-$bipartite graph, with $b \leq t$, it is clear that cycles must be of length $\leq 2b$. □

It is also possible to characterize the $A-$ and $E-$optimal designs in the unicyclic case. For example, we can show that if $9 \leq t = b(k-1) \leq 12$, we obtain

- an $E-$optimal design by forming a cycle of length 6 in the LG (that is, a cycle connecting three blocks), and attaching all remaining blocks to one treatment of the cycle.
- an $A-$optimal design by forming a cycle of length 8 in the LG (that is, a cycle connecting four blocks), and attaching all remaining blocks to one treatment of the cycle.

Exercises

1. Show that the function $f : \boldsymbol{x} \to \log(\frac{1}{x_1}-1)+\log(\frac{1}{x_2}-1)$ is Schur-convex on $\mathcal{A} := \{\boldsymbol{x} \in (0,1)^2 : x_1 + x_2 \leq 1\}$, but f is not convex on $\mathcal{A}$.

2. Find the A-, E- and D-optimal (t, b, k)-block designs with
   - $b = 3,\ k = 4,\ t = 10$;
   - $b = 7,\ k = 3,\ t = 7$;
   - $b = 6,\ k = 3,\ t = 12$.

3. Proof of the *Cut-set lemma*

   Lemma 9. *Let G be a connected graph on n vertices. Assume that there is a subset of m edges whose removal disconnects G and separates th vertices into two vertex sets of size n_1 and n_2, with $n_1 + n_2 = n$. Then, the smallest nonzero eigenvalue μ of the Laplacian of G satisfies*

   $$\mu \leq m\left(\frac{1}{n_1} + \frac{1}{n_2}\right).$$

   Let V_1 and V_2 be the vertex sets of the theorem, and let $\boldsymbol{v}$ be the vector with n_2 on the vertices of V_1, and $-n_1$ on the vertices of V_2.

   - Show that $\boldsymbol{v}$ is orthogonal to the trivial eigenspace of L.
   - Compute $\boldsymbol{v}^T\boldsymbol{v}$ and $\boldsymbol{v}^T L\boldsymbol{v}$.
   - Conclude by using the Rayleigh principle.

4. Use the cut-set lemma to prove $E-$optimality of queen-bee designs (in the class of $(t, b, k)-$designs where $b(k - 1) = t$). We admit that the smallest nontrivial eigenvalue of the Laplacian of a CG of a queen-bee design is $\mu = 1$.
   - Start to show the theorem for blocks of size $k = 2$.
   - For blocks of size k, show that a design in which a treatment appears in at least 3 blocks has $\mu \leq 1$.
   - Similarly, show that a design in which all treatments appear in at most 2 blocks, and the number of blocks is $b \geq k + 1$, cannot be $E-$optimal.
   - The missing cases (treatments appear in at most 2 blocks, $b \leq k$) require an alternative cut-set lemma.