Optimal design of experiments for model discrimination

Tommasi Chiara

Department of Economics, Management and Quantitative Methods
Univerisity of Milan, Italy

MASCOT-NUM
Giens Peninsula, Hyères (France)
April 2-5, 2024

Outline

1. Notation and background
2. Optimal design for model discrimination: D_s-, T- and KL-optimality criteria.
3. Equivalence theorem and the first order algorithm.
4. The double aim of model discrimination and parameter estimation: DKL-optimality criterion.
5. Removal of the parameter dependence:Bayesian KL-optimality criterion.
6. Discriminating among several models: Minimum KL-efficiency criterion.
7. Some references.

Notation

$x \in \mathcal{X}$: **experimental condition** chosen by the experimenter.
$y = y(x)$: observable experimental **response**.

Exact design: $\{x_1, \ldots, x_n\} \quad \longmapsto \quad \{y(x_1), \ldots, y(x_n)\}$

$$\xi_n = \left\{ \begin{array}{ccc} x_1 & \ldots & x_k \\ \xi_n(x_1) & \ldots & \xi_n(x_k) \end{array} \right\}, \quad \xi_n(x_j) = \frac{n_j}{n}, \; j = 1, \ldots k < n$$

Continuous design ξ: a probability measure on $\mathcal{X}$.

Responses and experimental conditions may be related through:

1. **Regression model:**
   $y = \eta(x, \theta) + \varepsilon, \quad \mathrm{E}(\varepsilon) = 0, \quad \mathrm{Var}(\varepsilon) = \sigma^2.$
2. **Statistical model:** parametric family of pdf's, $f(y, x, \theta)$.

Optimal design of experiments

An **optimality criterion function** is a concave function

$$\Phi : \Xi \longrightarrow \mathbb{R}, \quad \xi \in \Xi,$$

which summarizes the goal of the inferential study: estimation, prediction or discrimination.

Equivalence theorem

The design ξ^* is called Φ–optimal if and only if

$$\partial\Phi(\xi^*, \bar{\xi}) \leq 0, \qquad \text{for any design } \bar{\xi},$$

where $\partial\Phi(\xi, \bar{\xi}) = \lim_{\lambda \to 0^+} \frac{\Phi[(1-\lambda)\xi + \lambda\bar{\xi}] - \Phi(\xi)}{\lambda}$.

If differentiable, $\partial\Phi(\xi^*, \xi_x) \leq 0$, for any $x \in \mathcal{X}$, where $\xi_x = \left\{ \begin{array}{c} x \\ 1 \end{array} \right\}$.

The **efficiency** of a design ξ with respect to ξ^* is:

$$0 \leq \mathrm{Eff}_\Phi(\xi) = \frac{\Phi(\xi)}{\Phi(\xi^*)} \leq 1$$

Optimal design for model discrimination

1. $\eta_i(x, \theta_i)$: **regression model**, where $\theta_i \in \Omega_i \subset \mathbb{R}^{m_i}$ is an unknown parameter vector, $i = 1, 2$.
2. $f_i(y, x, \theta_i)$: **statistical model**, where $\theta_i \in \Omega_i \subset \mathbb{R}^{m_i}$ is an unknown parameter vector, $i = 1, 2$.

GOAL

To fix the experimental conditions with the aim of identifying:

1. which of two rival regression functions, $\eta_1(x, \theta_1)$ and $\eta_2(x, \theta_2)$, is the most adequate ($\mathrm{D_s}$- and T-optimality);
2. which of two rival statistical models, $f_1(y, x, \theta_1)$ and $f_2(y, x, \theta_2)$, is the most adequate ($\mathrm{D_s}$- and KL-optimality).

Nested models: D_s–criterion

$$\eta_1(x) \cong \theta_1^T f_1(x) + \theta_2^T f_2(x) \quad \text{and} \quad \eta_2(x) \cong \theta_2^T f_2(x)$$

$$\boxed{y \cong F_1\theta_1 + F_2\theta_2 + \varepsilon = F\theta + \varepsilon}, \quad F = [F_1, F_2], \theta = \begin{pmatrix} \theta_1 \\ \theta_2 \end{pmatrix}$$

$$M(\xi) \propto F^T F = \begin{bmatrix} F_1^T F_1 & F_1^T F_2 \\ F_2^T F_1 & F_2^T F_2 \end{bmatrix} \quad M^{-1}(\xi) = \begin{bmatrix} M^{11}(\xi) & M^{12}(\xi) \\ M^{21}(\xi) & M^{22}(\xi) \end{bmatrix}$$

$$\boxed{\mathrm{Var}(\hat{\theta}_1) \propto M^{11}(\xi) = \left\{ F_1^T [I - F_2 (F_2^T F_2)^{-1} F_2^T] F_1 \right\}^{-1}}$$

$\mathrm{D_s}$-optimality criterion

$$\Phi_{D_s}(\xi) = \log \left| M^{11}(\xi) \right|^{-1} = \log \frac{|M(\xi)|}{|M_{22}(\xi)|}$$

The $\mathrm{D_s}$-optimum design, $\xi^*_{D_s} = \arg\max_\xi \Phi_{D_s}(\xi)$, minimizes in some sense $\mathrm{Var}(\hat{\theta}_1)$.

D_s–optimality for model discrimination

- The **noncentrality parameter** of F test for the following system of hypothesis:

$$\begin{cases} H_0 : \ \eta(x) = \theta_2^T f_2(x) \\ H_1 : \ \eta(x) = \theta_1^T f_1(x) + \theta_2^T f_2(x) \end{cases} \qquad \Leftrightarrow \qquad \begin{cases} H_0 : \ \theta_1 = 0 \\ H_1 : \ \theta_1 \neq 0 \end{cases}$$

is

$$\boxed{\zeta(\xi;\theta_1) \propto \theta_1^T \ F_1^T [I - F_2(F_2^T F_2)^{-1} F_2^T] F_1 \, \theta_1}$$

Interpretation

The power of the F test is an increasing function of the noncentrality parameter. Since $\xi^*_{D_s} = \arg\max_\xi \left| F_1^T [I - F_2(F_2^T F_2)^{-1} F_2^T] F_1 \right|$, in some sense the D_s-optimum design **maximizes** the **power** of the F test **for any** value of θ_1.

Equivalence theorem

1. If $\xi^*_{D_s}$ is a **regular design**, i.e. if $M(\xi^*_{D_s})$ is a non singular matrix, then $\xi^*_{D_s}$ is $\mathrm{D_s}$-optimum if and only if:

$$\boxed{f(x)'M^{-1}(\xi^*_{D_s})f(x) - f_2(x)'M_{22}^{-1}(\xi^*_{D_s})f_2(x) - s \leq 0, \quad x \in \mathcal{X}}$$

2. In case of nested statistical models, $M(\xi;\theta) = \int J(x;\theta)d\xi(x)$ is the **Fisher information matrix** and $\xi^*_{D_s}$ is $\mathrm{D_s}$-optimum if and only if:

$$\boxed{\mathrm{tr}[M^{-1}(\xi^*_{D_s};\theta)J(x;\theta) - M_{22}^{-1}(\xi^*_{D_s};\theta_2)J_{22}(x;\theta)] - s \leq 0, \quad x \in \mathcal{X}}$$

Gaussian regression models (not necessarily nested): T–optimality (Atkinson and Fedorov, 1975)

$$\boxed{y = \eta_i(x, \theta_i) + \varepsilon, \quad \varepsilon \sim N(0, \sigma^2), \quad \sigma^2 \ \text{known}, \quad i = 1, 2}$$

Let $\eta(x) = \eta_1(x, \theta_1)$ be the **true** (known) model.

T-optimality criterion

The T-optimality criterion function is the minimum sum of squares for the lack of fit of the rival model:

$$T_{21}(\xi) = \inf_{\theta_2} \sum_{x \in \mathcal{X}} [\eta(x) - \eta_2(x, \theta_2)]^2 \, \xi(x)$$

The experiment should be designed to get the largest value of $T_{21}(\xi)$: $\xi^*_T = \arg\max_\xi T_{21}(\xi)$

Interpretation of T–optimality

- In the case of linear models: $\eta_j(x,\theta_j) = \theta_j^T f_j(x), \quad j = 1;2$

$$\begin{aligned} T_{21}(\xi) &= \min_{\theta_2} ||F_1\theta_1 - F_2\theta_2||^2, \qquad \hat{\theta}_2 = (F_2^T F_2)^{-1} F_2^T F_1 \theta_1 \\ &= \boxed{\theta_1^T \, F_1^T [I - F_2(F_2^T F_2)^{-1} F_2^T] F_1 \, \theta_1} \end{aligned}$$

- $T_{21}(\xi)$ is proportional to the **noncentrality parameter** of F test $\zeta(\xi;\theta_1)$ for testing:

$$\begin{cases} H_0 : \; \eta(x) = \theta_2^T f_2(x) \\ H_1 : \; \eta(x) = \theta_1^T f_1(x) + \theta_2^T f_2(x) \end{cases}$$

Interpretation

T-optimum design **maximizes** the **power** of the F test.

Equivalence theorem

If ξ_T^* is a regular design, i.e. if

$$\Omega_2(\xi_T^*) = \left\{ \hat{\theta}_2 : \; \hat{\theta}_2(\xi) = \arg\min_{\theta_2 \in \Omega_2} \sum_{x \in \mathcal{X}} [\eta(x) - \eta_2(x, \theta_2)]^2 \, \xi_T^*(x) \right\},$$

is singleton, then a necessary and sufficient condition for a design ξ_T^* to be T-optimum is

$$\boxed{[\eta(x) - \eta_2(x, \hat{\theta}_2)]^2 - \sum_{x \in \mathcal{X}} [\eta(x) - \eta_2(x, \hat{\theta}_2)]^2 \xi_T^*(x) \leq 0, \quad x \in \mathcal{X}}$$

Comparison between T- and D_s-criteria

- The D_s-optimum design $\xi^*_{D_s}$ maximizes $\left|M^{11}(\xi)\right|^{-1}$, therefore in some sense it also maximizes the power of the F test.
- When the rival models differ by one parameter, then $M^{11}(\xi)^{-1}$ is a scalar and thus, T- and D_s-criteria are equivalent.
- A disadvantage of D_s-optimality wrt the T-criterion is that it can be used only for nested models. However, the D_s-criterion is a more general tool as it can be applied to the Fisher information matrix, and therefore it enables to discriminate nested statistical models (not only regression functions).

Rival statistical models (nested or not; Gaussian or not): KL–optimality

Rival statistical models: $f_1(y,x,\theta_1)$ and $f_2(y,x,\theta_2)$.

Let $f(x) = f_1(y,x,\theta_1)$ be the **true** (known) model, which may include or not the rival model $f_2(y,x,\theta_2)$ as special case.

Kullback–Leibler divergence between f_1 and f_2:

$$\mathcal{I}\left[f_1, f_2\right] = \int f_1(y,x,\theta_1) \log \left[\frac{f_1(y,x,\theta_1)}{f_2(y,x,\theta_2)}\right] dy.$$

KL–optimality (López-Fidago, Tommasi and Trandafir '07)

$$I_{2,1}(\xi) = \min_{\theta_2} \int_{\mathcal{X}} \mathcal{I}\left[f_1(y,x,\theta_1), f_2(y,x,\theta_2)\right] \xi(dx)$$

KL-efficiency of ξ with respect to the KL-optimum design

$$\xi^*_{KL} = \arg\max_{\xi} I_{2,1}(\xi), \quad 0 \leq \mathrm{Eff}_{2,1}(\xi) = \frac{I_{2,1}(\xi)}{I_{2,1}(\xi^*_{KL})} \leq 1$$

EQUIVALENCE THEOREM

Given a design ξ^*_{KL} , if

$$\Omega_2(\xi^*_{KL}) = \left\{ \hat{\theta}_2 : \ \hat{\theta}_2(\xi^*_{KL}) = \arg \min_{\theta_2 \in \Omega_2} \sum_{x \in \mathcal{X}} \mathcal{I}(f_1, f_2, x, \theta_2)\, \xi^*_{KL}(x) \right\},$$

is singleton, then ξ^*_{KL} is KL-optimum if and only if

$$\boxed{\underbrace{\mathcal{I}(f_1, f_2, x, \hat{\theta}_2) - \sum_{x \in \mathcal{X}} \mathcal{I}(f_1, f_2, x, \hat{\theta}_2)\, \xi^*_{KL}(x)}_{\psi(x, \xi^*_{KL}) = \partial I_{21}(\xi^*_{KL}, \xi_x)} \leq 0, \quad x \in \mathcal{X}}$$

Properties of the KL-criterion

- **Concavity**: $I_{2,1}(\xi;\theta_1)$ is concave (Tommasi, 2007).
- **Upper semi-continuity**: if the Kullback-Leibler divergence $\mathcal{I}(x,\theta_1,\theta_2)$ is continuous with respect to x then $I_{2,1}(\xi;\theta_1)$ is upper semi-continuous (May and Tommasi, 2014).
- **Continuity**: Under some mild conditions on $\mathcal{I}(x,\theta_1,\theta_2)$ and $I_{2,1}(\xi;\theta_1)$, the KL-criterion is continuous (Aletti, May & Tommasi, 2014).
- **Invariance**: If $\mathcal{Z} = \{z = \alpha + qx | x \in \mathcal{X}\}$ then the KL-optimum design on $\mathcal{Z}$ is $\eta^*_{KL}(dz) = \xi^*_{KL}(dx)$ where $z = \alpha + qx$ and $x \in \mathcal{X}$ (Aletti, May & Tommasi, 2014).

A first order algorithm for computing ξ^*_{KL}

1. given ξ_s, find

$$\begin{aligned}
\theta_{2,s} &= \arg\min_{\theta_2 \in \Omega_2} \int_{\mathcal{X}} \int_{\mathcal{Y}} \log \frac{f_1(y|x;\theta_1)}{f_2(y|x;\theta_2)}\, f_1(y|x;\theta_1)\, dy\, \xi_s(dx) \\
x_s &= \arg\max_{x \in \mathcal{X}} \int_{\mathcal{Y}} \log \frac{f_1(y|x;\theta_1)}{f_2(y|x;\theta_{2,s})}\, f_1(y|x;\theta_1)\, dy
\end{aligned}$$

2. Choose $\alpha_s = \arg\max_{\beta \in [0,1]} I_{2,1}((1-\beta)\xi_s + \beta \xi_{x_s})$ and construct

$$\xi_{s+1} = (1-\alpha_s)\xi_s + \alpha_s \xi_{x_s},$$

where ξ_{x_s} is a design with measure concentrated at the single point x_s.

A stopping rule for the algorithm

The directional derivative of $I_{2,1}(\xi;\theta_1)$ at ξ in the direction $\xi_x - \xi$ is

$$\psi(x,\xi) = \mathcal{I}(f_1, f_2, x, \hat{\theta}_2) - \sum_{x \in \mathcal{X}} \mathcal{I}(f_1, f_2, x, \hat{\theta}_2)\, \xi(x)$$

Since

$$\left[1 + \frac{\max_{x \in \chi} \psi(x;\xi)}{I_{2,1}(\xi)}\right]^{-1} \leq \frac{I_{2,1}(\xi)}{I_{2,1}(\xi^*_{KL})} \leq 1,$$

the iterative procedure **stops** at the step s if ξ_s is such that

$$\left[1 + \frac{\max_{x \in \chi} \psi(x;\xi_s)}{I_{2,1}(\xi_s)}\right]^{-1} > \delta$$

where $0 < \delta < 1$ is a suitably choosen value, e.g. $\delta = .99$.

Regular designs

- In the context of parameter estimation a design is regular if its information matrix is non singular.
- In the setting of discrimination between models a design is regular if the following set is a singleton:

$$\Omega_2(\xi) = \left\{ \tilde{\theta}_2 : \; \tilde{\theta}_2(\xi) = \arg \min_{\theta_2 \in \Theta_2} \int_{\mathcal{X}} \mathcal{I}(x, \theta_2)\, \xi(dx) \right\}$$

- These definitions of regular designs are equivalent if $f_2(y|x;\theta_2)$ is a **generalized linear model** or a **nonlinear Gaussian model**. For these models, the first order algorithm (used to compute a KL-optimum design) converges whenever the initial design ξ_0 has a non singular information matrix (Aletti et al., 2014).

KL-optimality versus T- and D_s-criteria

1. For **regression models** with Normal error distribution, the KL- optimality criterion coincides with
   - the T-optimality criterion, in the homoschedastic case;
   - the generalization of the T-optimality criterion provided by Uciński and Bogacka (2004), in the heteroschedastic case.
2. For **generalized linear models**, the KL-criterion coincides with the generalization of the T-optimality criterion provided by Ponce de Leon and Atkinson (1992).
3. If f_2 is nested in f_1, then the D_s-criterion is a competitor of the KL-criterion. Whereas, if the models are separate, the KL-criterion is the only possibility.

EXAMPLE 1: discrimination between copula models in clinical trials (Deldossi, Osmetti, Tommasi, 2019)

When efficacy and toxicity are jointly studied, it is necessary to find the right dependence structure between the responses probabilities.

Let $x \in \mathcal{X}$ denote the **dose** of a drug and (Y_1, Y_2) be a **binary** efficacy-toxicity response variable.
Both Y_1 and Y_2 take values in $\{0, 1\}$.

▷ **Success probability of efficacy:**

$$\pi_1(x; \alpha) = P(Y_1 = 1|x; \alpha) = \frac{e^{\alpha_0+\alpha_1 x+\alpha_2 x^2}}{1 + e^{\alpha_0+\alpha_1 x+\alpha_2 x^2}}$$

▷ **"Success" probability of toxicity:**

$$\pi_2(x; \beta) = P(Y_2 = 1|x; \beta) = \frac{e^{\beta_0+\beta_1 x}}{1 + e^{\beta_0+\beta_1 x}}$$

- $\alpha = (\alpha_0, \alpha_1, \alpha_2)$ and $\beta = (\beta_0, \beta_1)$ are unknown coefficients

A copula model for efficacy and toxicity

A copula $C(\cdot,\cdot;\theta_C)$ represents the **dependence structure** between Y_1 and Y_2, as the joint probability can be expressed in terms of the copula:

	Toxicity		
Efficacy	1	0	
1	p^C_{11}	p^C_{10}	$\pi_1(x;\alpha)$
0	p^C_{01}	p^C_{00}	$1-\pi_1(x;\alpha)$
	$\pi_2(x;\beta)$	$1-\pi_2(x;\beta)$	1

where $p^C_{11}=P(Y_1=1,Y_2=1|x;\delta,\theta_C)=C\left[\pi_1(x;\alpha),\pi_2(x;\beta);\theta_C\right]$ and

$$\begin{aligned} p^C_{10} &= \pi_1(x;\alpha)-p^C_{11}(x;\delta,\theta_C), \\ p^C_{01} &= \pi_2(x;\beta)-p^C_{11}(x;\delta,\theta_C), \\ p^C_{00} &= 1-\pi_1(x;\alpha)-\pi_2(x;\beta)+p^C_{11}. \end{aligned}$$

The P-optimal dose depends on the copula function

Clinicians are interested in the optimal safe dose:

$$d_p^* = \arg\max_{d \in \mathcal{D}} p_{10}^{C}(d; \delta, \theta_C),$$

where the design region $\mathcal{X}$ has been transformed into $\mathcal{D} = [-1, 1]$ through

$$d = \frac{x - (x_{min} + x_{max})/2}{(x_{max} - x_{min})/2} \in \mathcal{D} = [-1, 1].$$

Motivation of the discriminating between rival copulas

The P-optimal dose d_P^* may change considerably under different dependence structures, therefore it is necessary to discriminate between rival copulas.

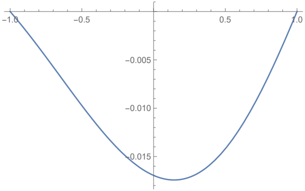
Discriminating Clayton versus Gumbel copulas

Copula	$C(u_1, u_2; \theta_C)$	$\theta_C \in \Theta_C$
Clayton	$(u_1^{-\theta_{Cl}} + u_2^{-\theta_{Cl}} - 1)^{-1/\theta_{Cl}}$	$\theta_{Cl} \in [0, \infty)$
Gumbel	$\exp\left(-\left[\{-\ln(u_1)\}^{\theta_G} + \{-\ln(u_2)\}^{\theta_G}\right]^{1/\theta_G}\right)$	$\theta_G \in [1, \infty)$

Assuming as true model the Clayton Copula (with $\theta_{Cl} = 2$), and setting $(\alpha_0, \alpha_1, \alpha_2) = (-.5, 1, 0)$ and $(\beta_0, \beta_1) = (0, 1)$, from the application of the first order algorithm, we obtain

$$\xi^*_{KL} = \left\{ \begin{array}{cc} \text{-1} & 1 \\ 0.58 & 0.42 \end{array} \right\}$$

which is optimal as shown by the figure of directional derivative:



Designing to detect model heteroscedasticity (Lanteri, Leorato, Lopez-Fidalgo, Tommasi, 2023)

Non-linear Gaussian regression model

$$y_i = \eta(x_i; \beta) + \varepsilon_i, \ \varepsilon_i \sim N(0; \sigma^2 h(x_i; \gamma)), \quad i = 1, \ldots, n$$

- The **error variance** depends on the experimental conditions through a **specific** positive function: $h : \mathbb{R}^p \times \mathbb{R}^s \mapsto \mathbb{R}_+$.
- $\beta \in \mathbb{R}^m$, σ^2 and $\gamma \in \mathbb{R}^s$ are **unknown parameters**, such that γ_0 leads to the **homoscedastic model**, i.e. $h(x_i; \gamma_0) = 1$.

Inferential goal: to test local alternatives

$$\begin{cases} H_0 : & \gamma = \gamma_0 \\ H_1 : & \gamma = \gamma_0 + \frac{\lambda}{\sqrt{n}}, \quad \lambda \neq 0 \end{cases}$$

by applying a **likelihood-based test** (log-likelihood ratio, score or Wald test).

Noncentrality parameter of a likelihood-based test

Noncentrality parameter

Under local alternatives, $H_1 : \gamma = \gamma_0 + \lambda/\sqrt{n}$, a likelihood-based statistic is asymptotically distributed as a chi-squared r.v. with s df and **noncentrality parameter:**

$$\zeta(\xi; \lambda; \gamma_0) = \lambda^T M_{22.1}(\xi; \gamma_0)\, \lambda$$

where $M_{22.1}(\xi; \gamma) = M_{22} - M_{12}^T M_{11}^{-1} M_{12}$ is the Shur complement matrix of M_{22} in the following partition of the Fisher information matrix

$$M(\xi; \beta, \sigma^2, \gamma) = \begin{bmatrix} M_{11} & M_{12} \\ M_{12}^T & M_{22} \end{bmatrix}$$

Noncentrality parameter and D_s-optimality

Let us recall that the asymptotic covariance matrix of MLE $(\hat{\beta}, \hat{\sigma}^2, \hat{\gamma})$ is

$$M(\xi; \beta, \sigma^2, \gamma)^{-1} = \begin{bmatrix} M^{11} & M^{12} \\ M^{21} & M^{22} \end{bmatrix}$$

Therefore, asymptotic covariance matrix of $\hat{\gamma}$ is $M^{22} = [M_{22.1}(\xi; \gamma)]^{-1}$ and the D_s-optimum design for γ is

$$\xi_{D_s} = \arg\min_{\xi} |M^{22}| = \arg\max_{\xi} |M_{22.1}(\xi; \gamma)|.$$

D_s-optimality (at γ_0): a criterion for testing hypothesis

$\xi_{D_s} = \arg\max_{\xi} |M_{22.1}(\xi; \gamma_0)|$ maximizes **in some sense** $\zeta(\xi; \lambda; \gamma_0)$ (for any value of λ).

In the **scalar case** ($s = 1$), the ξ_{D_1} maximizes **exactly** the noncentrality parameter.

Analytical expression of the D_1-optimum design

D_s-optimality for γ is equivalent to D-optimality

The D_s-optimum design for γ coincides with the D-optimum design for estimating (α_0, α^T), with $\alpha^T = (\alpha_1, \ldots, \alpha_s)$, in the following linear regression model:

$$y_i = \alpha_0 + \alpha^T \nabla \log h(x_i; \gamma) + \varepsilon_i, \; \varepsilon_i \sim N(0; \sigma^2), \quad i = 1, \ldots, n,$$

where $\nabla \log h(x; \gamma) = \left(\frac{\partial \log h(x;\gamma)}{\partial \gamma_1}, \ldots, \frac{\partial \log h(x;\gamma)}{\partial \gamma_s} \right)^T$.

Scalar case: D_1-optimal design

$$\xi_{\mathrm{D}_1} = \left\{ \begin{array}{cc} \text{argmin}_x \left. \frac{\partial \log h(x_i;\gamma)}{\partial \gamma} \right|_{\gamma=\gamma_0} & \text{argmax}_x \left. \frac{\partial \log h(x_i;\gamma)}{\partial \gamma} \right|_{\gamma=\gamma_0} \\ 0.5 & 0.5 \end{array} \right\}$$

Analytical expression for the KL-optimum design

Usually, the KL-optimum design must be computed numerically and the computation may be **cumbersome**.

KL-optimum design to discriminate
$\varepsilon_i \sim N(0;\sigma^2 h(x_i;\gamma_1)) \quad \text{vs} \quad \varepsilon_i \sim N(0;\sigma^2)$, where $\gamma_1 = \gamma_0 + \frac{\lambda}{\sqrt{n}}$

$$I_{2,1}(\xi;\gamma_1) = 1 + \log A_h - \log G_h$$

where $A_h = \sum_{i=1}^k h(x_i;\gamma_1)\,\xi(x_i)$ and $G_h = \prod_{i=1}^k [h(x_i;\gamma_1)]^{\xi(x_i)}$ are the arithmetic and the geometric means $h(x_i;\gamma_1)$, $i = 1,\dots,k$, respectively.

$$\begin{aligned} \xi_{\gamma_1}^{KL} &= \arg\max_{\xi} I_{2,1}(\xi;\gamma_1) = \left\{ \begin{array}{cc} \arg\inf_x h(x;\gamma_1) & \arg\sup_x h(x;\gamma_1) \\ \omega & 1-\omega \end{array} \right\}, \\ \omega &= \left(\frac{\overline{h}}{\overline{h} - \underline{h}} - \frac{1}{\log \overline{h} - \log \underline{h}} \right) \end{aligned}$$

KL-criterion and noncentrality parameter

From a Taylor expansion of $I_{2,1}(\xi;\gamma_1) = 1 + \log A_h - \log G_h$ at γ_0:

Connection between KL-criterion and noncentrality parameter

$$I_{2,1}(\xi;\gamma_1) = I_{2,1}\left(\xi;\gamma_0 + \frac{\lambda}{\sqrt{n}}\right) = 1 + \frac{1}{n}\zeta(\xi;\lambda;\gamma_0) + O\left(\frac{||\lambda||^3}{n^{\frac{3}{2}}}\right).$$

This expansion holds **uniformly** in ξ, therefore as $n \to \infty$,

$$\xi^{KL}_{\gamma_1} = \arg\sup_{\xi} I_{2,1}(\xi;\gamma_1) \to \arg\sup_{\xi} \zeta(\xi;\lambda;\gamma_0)$$

$\xi^{KL}_{\gamma_0}$ maximizes the noncentrality parameter

$$\xi^{KL}_{\gamma_1} \to \xi^{KL}_{\gamma_0} = \left\{ \begin{array}{cc} \underline{x} & \overline{x} \\ 0.5 & 0.5 \end{array} \right\}$$

therefore, $\xi^{KL}_{\gamma_0}$ maximizes the noncentrality parameter $\zeta(\xi,\lambda,\gamma_0)$.

EXAMPLE 2: The Hill model

- The Hill model

$$y = \eta(x,\beta) + \varepsilon = \frac{(E_{con} - b) \cdot (x/IC_{50})^s}{(1 + (x/IC_{50})^s)} + b + \varepsilon,$$

  is widely applied in dose-response contexts, biology and enzymatic kinetics.

- Physical interpretation of the parameters:E_{con} the effect on the control for dose $x = 0$, b the asymptotic value of the response when $x \to \infty$, IC_{50} corresponds to the value for which the response would be the middle of the range $E_{con} - b$ and finally s is a shape parameter $s > 0$ makes the response strictly increasing and $s < 0$ strictly decreasing.
- The random error term may be: $\varepsilon \sim \mathcal{N}(0, \sigma^2)$ (constant absolute error), or $\varepsilon \sim \mathcal{N}(0, \sigma^2\, \eta(x,\beta)^2)$ (constant relative error).

Constant absolute error vs constant relative error

- To decide in favour of one of the two error-variance structures, we compute the KL-optimum design (setting $\beta = (E_{con}, b, IC_{50}, s)^T = (1.70, 0.137, 111, -1.03)^T$ and $\mathcal{X} = [0.01, 1500]$):

$$\xi^{\star}_{KL} = \left\{ \begin{array}{cc} 0.01 & 1500 \\ 0.23 & 0.77 \end{array} \right\},$$

  that does not allow to estimate all the parameters of the model.
- The KL-optimum design has only 2 support points, therefore the parameters of the Hill model cannot be estimated.
- **The aim of a design should be dual:** to discriminate between the rival models and to estimate efficiently the parameters of the models.

MODEL DISCRIMINATION AND PARAMETER ESTIMATION

Background

- Nested regression models with **Gaussian** errors:
  DD$_1$-criterion: Dette (1993).
  DD$_s$-criterion: Tsai and Zen (2004) and Zen and Tsai (2004).
  DT-criterion: Atkinson (2008).

- **DKL-optimality criterion** (Tommasi, 2009):

$$\Phi_{DKL}(\xi)=\left(\frac{I_{21}(\xi)}{I_{21}(\xi_{21}^*)}\right)^{\alpha_1}\left(\frac{I_{12}(\xi)}{I_{12}(\xi_{12}^*)}\right)^{\alpha_2}\left(\frac{|M_1(\xi)|}{|M_1(\xi_{D_1}^*)|}\right)^{\frac{\alpha_3}{m_1}}\left(\frac{|M_2(\xi)|}{|M_2(\xi_{D_2}^*)|}\right)^{\frac{\alpha_4}{m_2}},\ \sum_{i=1}^{4}\alpha_i=1$$

  DKL-optimum design: $\boxed{\xi_{DKL}^* = \arg\max_\xi \log \Phi_{DKL}(\xi)}$

- May and Tommasi (2013) generalize the DKL-optimality to the case of **several nested non-linear models**. They also provide a sequential version of their criterion which asymptotically selects the true model and converges to the *D*-optimum design for the true model.

BAYESIAN KL-CRITERION

Drawback of KL-criterion

The KL-criterion depends on the unknown parameters θ_1 of the assumed known model $f_1(y; x, \theta_1)$: ξ^*_{KL} is only **locally** optimum.

- **Standardized Bayesian KL-criterion**:

$$\boxed{I^{SB}(\xi) = \pi_1\, E_{\theta_1}\left[\frac{I_{2,1}(\xi,\theta_1)}{I_{2,1}(\xi^*_{2,1},\theta_1)}\right] + (1-\pi_1)\, E_{\theta_2}\left[\frac{I_{1,2}(\xi,\theta_2)}{I_{1,2}(\xi^*_{1,2},\theta_2)}\right]}$$

  $I_{2,1}(\xi;\theta_1)$ and $I_{1,2}(\xi;\theta_2)$ may have different magnitudes for this reason they are standardized.

- **Standardized Bayesian KL-optimum design**:

$$\xi^*_{SB} = \arg\max_{\xi} I^{SB}(\xi),$$

  which **does not depend** any more on the nominal values for the parameters.

KL–optimality for several models

Rival models: $f_i(y, x, \theta_i)$ with $i = 1, \dots, k$.

Extended model: $f_{k+1}(y, x, \theta_{k+1})$, which includes each $f_i(y, x, \theta_i)$ as a special case.

Kullback–Leibler divergence between f_{k+1} and f_i:

$$\mathcal{I}\left[f_{k+1}, f_i\right] = \int f_{k+1}(y, x, \theta_{k+1}) \log\left[\frac{f_{k+1}(y, x, \theta_{k+1})}{f_i(y, x, \theta_i)}\right] dy.$$

KL–optimality to discriminate between f_i and f_{k+1}

$$I_{i,k+1}(\xi) = \min_{\theta_i} \int_{\mathcal{X}} \mathcal{I}\left[f_{k+1}(y, x, \theta_{k+1}), f_i(y, x, \theta_i)\right] \xi(dx)$$

KL-efficiency of ξ with respect to the KL-optimum design

$$\mathrm{Eff}_{i,k+1}(\xi) = \frac{I_{i,k+1}(\xi)}{I_{i,k+1}(\xi_i^*)}, \qquad \xi_i^* = \arg\max_{\xi} I_{i,k+1}(\xi)$$

Generalized KL–criterion

Generalized KL–criterion (Tommasi '07)

$$I_\alpha(\xi)=\sum_{i=1}^{k}\alpha_i\,\mathrm{Eff}_{i,k+1}(\xi),\quad \alpha_i\geq 0,\ \sum_{i=1}^{k}\alpha_i=1$$

Equivalence Theorem

ξ_α is a Generalized KL-optimum design, i.e. $\xi_\alpha=\arg\max_\xi I_\alpha(\xi)$, **iff**

$$\sum_{i=1}^{k}\alpha_i\left[\mathcal{I}^s_{i,k+1}(x)-\int_{\mathcal{X}}\mathcal{I}^s_{i,k+1}(x)\,\xi_\alpha(dx)\right]\leq 0,\quad x\in\mathcal{X}$$

Standardized Kullback–Leibler divergence:

$$\mathcal{I}^s_{i,k+1}(x)=\frac{\mathcal{I}\left[f_{k+1}(y,x,\theta_{k+1}),f_i(y,x,\hat{\theta}_i)\right]}{I_{i,k+1}(\xi_i^*)},\quad \text{where}$$

$$\hat{\theta}_i=\arg\min_{\theta_i\in\Omega_i}\int_{\mathcal{X}}\mathcal{I}\left[f_{k+1}(y,x,\theta_{k+1}),f_i(y,x,\theta_i)\right]\xi_i^*(dx).$$

Minimum KL–efficiency criterion

Minimum KL–efficiency criterion (Tommasi, Martín-Martín, López-Fidalgo, 2015)

$$I_m(\xi) = \min_{i \in \{1,\dots,k\}} \mathrm{Eff}_{i,k+1}(\xi)$$

Motivation: $I_m(\xi)$ gives equal efficiencies to those models that are more difficult to be discriminated.

Equivalence Theorem

ξ_m^* is a max-min KL-efficiency design, i.e. $\xi_m^* = \arg\max_{\xi} I_m(\xi)$, **iff** there exists a set of weights α_i^* on the index set of the models with the same minimum efficiency, i.e.

$\mathcal{C}(\xi_m^*) = \left\{ \arg \min_{i \in \{1,\dots k\}} \mathrm{Eff}_{i,k+1}(\xi_m^*) \right\}$, such that

$$\sum_{i \in \mathcal{C}(\xi_m^*)} \alpha_i^* \left[\mathcal{I}_{i,k+1}^s(x) - \int_{\mathcal{X}} \mathcal{I}_{i,k+1}^s(x)\, \xi_m^*(dx) \right] \leq 0, \quad x \in \mathcal{X}$$

Properties

- The Equivalence Theorem cannot be used to check for the minimum KL–efficiency optimality of a design, because both the index set $\mathcal{C}(\xi_m^*)$ and the weights α_i^* on $\mathcal{C}(\xi_m^*)$ are **unknown**.

- A max-min KL-efficiency design is **always** a Generalized KL-optimum design.

- This suggests to search the max-min KL-efficency design $\xi_m^* = \arg\max_{\xi} \min_{i \in \{1,\dots,k\}} \mathrm{Eff}_{i,k+1}(\xi)$ in the class of the Generalized KL-optimum designs:

$$\xi_\alpha = \arg\max_{\xi} I_\alpha(\xi), \quad \alpha_i \geq 0, \; \sum_{i=1}^{k} \alpha_i = 1$$

.

Corollaries of the Equivalence Theorem

Corollary 1

A max-min KL–efficiency design: $\xi_m^* = \arg\max_\xi \min_{i\in\{1,\dots,k\}} \mathrm{Eff}_{i,k+1}(\xi)$ is also optimum for the optimality criterion $\min_{i\in\mathcal{C}(\xi_m^*)} \mathrm{Eff}_{i,k+1}(\xi)$.

Corollary 2

If for any subset $\mathcal{A} \subseteq \{1, 2, \dots, k\}$, such that $\#\mathcal{A} \leq l < k$, the max-min KL–efficiency design to discriminate among models $\{f_i : i \in \mathcal{A}\}$ is not optimum to discriminate among all the k models, then $\#\mathcal{C}(\xi_m^*)$ is at least equal to $l + 1$.

Corollary 3

A design $\xi_i^* = \arg\max_\xi \mathrm{Eff}_{i,k+1}(\xi)$, which is KL–optimum to discriminate between f_i and f_{k+1}, **cannot** be a max-min KL–efficiency design, for any $i = 1, \dots, k$. In other words $\#\mathcal{A} \neq 1$.

Iterative algorithm

1. Set $l = l + 1, \quad l = 1, \dots, k-1.$
2. For a subset of l models $\{f_i : i \in \mathcal{A}\}$ with $\mathcal{A} = \{i_1, \dots, i_l\}$, **find** the $\tilde{\alpha}$ for which the corresponding Generalized KL-optimum design
$$\xi_{\tilde{\alpha}} = \arg\max_{\xi} \sum_{j=1}^{l} \tilde{\alpha}_j \cdot \mathrm{Eff}_{i_j,k+1}(\xi), \quad \text{satisfies}$$
$$\mathrm{Eff}_{i_1,k+1}(\xi_{\tilde{\alpha}}) = \mathrm{Eff}_{i_2,k+1}(\xi_{\tilde{\alpha}}) = \cdots = \mathrm{Eff}_{i_l,k+1}(\xi_{\tilde{\alpha}}). \qquad (1)$$
3. **If**
$$\mathrm{Eff}_{r,k+1}(\xi_{\tilde{\alpha}}) > \mathrm{Eff}_{i_1,k+1}(\xi_{\tilde{\alpha}}) = \cdots = \mathrm{Eff}_{i_l,k+1}(\xi_{\tilde{\alpha}}),$$
$$r \neq i_j; \;\; j = 1, \dots, l \qquad (2)$$
then STOP, since $\tilde{\alpha} = \alpha^*$ and thus $\xi_{\tilde{\alpha}} = \xi_m^*$.
Else, try another subset of l models and go to Step 2.
4. **If** for any subset of l models the corresponding $\xi_{\tilde{\alpha}}$ under restriction (1) does not satisfy (2), **then** go to Step 1.

EXAMPLE 3: logistic regression models

- **Rival models** (they belong to the **same** class of models):

$$\boxed{P_i(Y=1; x, \theta_i) = \frac{e^{\eta_i(x,\theta_i)}}{1+e^{\eta_i(x,\theta_i)}}, \quad \mathcal{X} = [0,1]}$$

$$\begin{aligned}
\eta_1(x,\theta_1) &= \theta_{1,1}x \\
\eta_2(x,\theta_2) &= \theta_{2,0} + \theta_{2,1}x \\
\eta_3(x,\theta_3) &= \theta_{3,1}x + \theta_{3,2}x^2
\end{aligned}$$

- A reasonable **extended model** corresponds to the following predictor function:

$$\eta_4(x,\theta_4) = \theta_{4,0} + \theta_{4,1}x + \theta_{4,2}x^2.$$

Nominal values: $\theta_{4,0} = \theta_{4,1} = \theta_{4,2} = 1.$

Numerical computation of max-min KL-efficiency designs

STEP 1. For a pair of models (P_{i_1}, P_{i_2}), find $\tilde{\alpha}$ and compute

$$\xi_{\tilde{\alpha}} = \arg\max_{\xi} [\tilde{\alpha} \cdot \mathrm{Eff}_{i_1,m}(\xi) + (1-\tilde{\alpha}) \cdot \mathrm{Eff}_{i_2,m}(\xi)]$$

for which $\mathrm{Eff}_{i_1,m}(\xi_{\tilde{\alpha}}) = \mathrm{Eff}_{i_2,m}(\xi_{\tilde{\alpha}})$.

STEP 2. If

$$\mathrm{Eff}_{i_3,m}(\xi_{\tilde{\alpha}}) > \mathrm{Eff}_{i_1,m}(\xi_{\tilde{\alpha}}) = \mathrm{Eff}_{i_2,m}(\xi_{\tilde{\alpha}}) \qquad (3)$$

STOP since $\xi_{\tilde{\alpha}} = \xi^*_m$. Else try another couple of models and go back to Step 1.

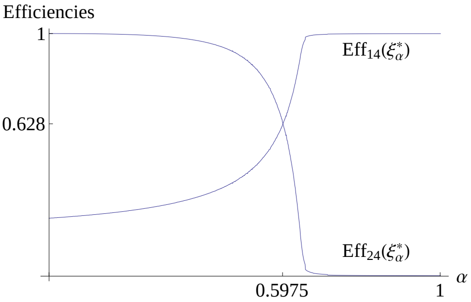
STEP 3. If for any pair of models condition (3) is never satisfied, then the max-min KL-efficiency design is

$$\xi^*_m = \arg\max_{\xi} \mathrm{Eff}_{i_1,m}(\xi)$$

under the common efficiency restriction:

$$\mathrm{Eff}_{i_1,m}(\xi) = \mathrm{Eff}_{i_2,m}(\xi) = \mathrm{Eff}_{i_3,m}(\xi)$$

Subset $\{P_1, P_2\}$



$$\tilde{\alpha} = .5975 \quad \text{and} \quad \xi^{*(1,2)}_{\tilde{\alpha}} = \left\{ \begin{array}{ccc} 0. & 0.344875 & 1. \\ 0.610709 & 0.245741 & 0.143551 \end{array} \right\}$$

is **not** the max-min KL-efficiency design since

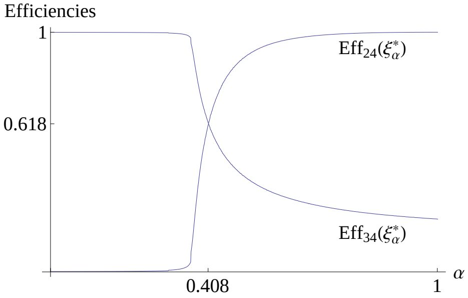
$$\mathrm{Eff}_{3,4}\left(\xi^{*(1,2)}_{\tilde{\alpha}}\right) = .611 < \mathrm{Eff}_{1,4}\left(\xi^{*(1,2)}_{\tilde{\alpha}}\right) = \mathrm{Eff}_{2,4}\left(\xi^{*(1,2)}_{\tilde{\alpha}}\right) = .628.$$

Subset $\{P_1, P_3\}$

The Generalized KL-optimum is concentrated at the point $x = 0$ for any value of α, that is $\xi_\alpha^{*(1,3)} = \xi_0$.

$\mathrm{Eff}_{2,4}(\xi_0) < \mathrm{Eff}_{1,4}(\xi_0) = \mathrm{Eff}_{3,4}(\xi_0) = 1$, thus ξ_0 **cannot** be the max-min KL-efficiency design.

Subset $\{P_2, P_3\}$



$$\tilde{\alpha} = .408 \quad \text{and} \quad \xi^{*(2,3)}_{\tilde{\alpha}} = \left\{ \begin{array}{ccc} 0. & 0.361548 & 1. \\ 0.618394 & 0.239079 & 0.142527 \end{array} \right\}$$

$\mathrm{Eff}_{1,4}(\xi^{*(2,3)}_{\tilde{\alpha}}) = .633 > \mathrm{Eff}_{2,4}(\xi^{*(2,3)}_{\tilde{\alpha}}) = \mathrm{Eff}_{3,4}(\xi^{*(2,3)}_{\tilde{\alpha}}) = .618$,
therefore $\xi^{*(2,3)}_{\tilde{\alpha}}$ **is** the max-min KL-efficiency design.

Ongoing research in discrimination problems

- This algorithm for computing max-min efficiency designs is very slow and may give numerical problems if the rival models are not smooth enough. Chen et al. (2020) have used the particle swarm optimization (PSO) to numerically search KL-optimum designs. In collaboration with Pozuelo-Campos, Casero-Alonso, López-Fidalgo and Wong, we are applying the PSO algorithm to discriminate among several **random effects linear models**.
- When $\theta_i \sim N(b_i; \sigma^2 T_i)$, with $i = 1, 2$, the KL-criterion has the same interpretation of the T-criterion:

$$I_{2,1}(\xi) = \min_{b_2} ||F_1 b_1 - F_2 b_2||^2_{\Sigma_1^{-1}},$$

where $||\cdot||_{\Sigma_1^{-1}}$ denotes the norm which corresponds to the inner product $< v, w >_{\Sigma_1^{-1}} = v' \Sigma_1^{-1} w$ and $\Sigma_1 = I + F_1\, T_1\, F_1'$.

References

Aletti, May and Tommasi. KL-optimum designs: theoretical properties and practical computation. Statistics and Computing, 2014.

Atkinson. Planning experiments to detect inadequate regression models. Biometrika, 1972.

Atkinson and Cox. Planning experiments for discriminating between models. JRSSB, 1974.

Atkinson and Fedorov . The design of experiments for discriminating between two rival models (several models). Biometrika, 1975 (a, b).

Biedermann, Dette and Pepelyshev. Optimal discrimination designs for exponential regression models. JSPI, 2007.

Braess and Dette. Optimal discriminating designs for several competing regression models. Annals, 2013.

Dette, Melas, Pepelyshev, and Strigul. Efficient design of experiments in the monod model. JRSSB, 2003.

Dette and Titoff. Optimal discrimination designs. Annals of Statistics, 2008.

Deldossi, Osmetti, Tommasi. Optimal design to discriminate between rival copula models for a bivariate binary response. TEST, 2019.

References

Lanteri, Leorato, López-Fidalgo and Tommasi. Designing to detect heteroscedasticity in a regression model. JRSSB, 2023.

López–Fidalgo, Tommasi, and Trandafir. An optimal experimental design criterion for discriminating between non-normal models. JRSSB, 2007.

May and Tommasi. Model selection and parameter estimation in non-linear nested models. Statistica Sinica, 2014.

Ponce de Leon and Atkinson. Advances in GLM and Statistical Modelling, chapter The design of experiments to discriminate between two rival generalized linear models. Lecture Notes in Statistics. Springer-Verlag, 1992.

Tommasi. Optimal designs for discriminating among several non-normal models. In mODa 8, Physica-Verlag, 2007.

Tommasi. Optimal designs for both model discrimination and parameter estimation. JSPI, 2009.

Tommasi and López-Fidalgo. Bayesian optimum designs for discriminating between models with any distribution. CSDA, 2010.

Tommasi C., Martín-Martín R. and López-Fidalgo J. Max-min optimal discriminating designs for several statistical models. Statistics and Computing, 2015.

Ucinski and Bogacka. T-optimum designs for multiresponse dynamic heteroscedastic models. In MODA7, Springer Verlag, 2004.

Thanks for your attention!