Optimal design to test for heteroscedasticity in a regression model

Chiara Tommasi[1]

a joint work with

Lanteri, A.[1], **Leorato, S.**[1] and **López-Fidalgo, J.**[2]

[1]University of Milan (Italy)
Department of Economics, Management and Quantitatve Methods
[2]University of Navarre (Spain)
Institute of Data Science and Artificial Intelligence

22^{nd} Annual ENBIS Conference
Trondheim, June 26-30, 2022

Outline

1. Goal and notation.
2. Optimal design for model discrimination:
   D_s- and KL-optimality criteria.
3. D_s-criterion and noncentrality parameter of a likelihood-based test to detect heteroscedasticity.
4. Noncentrality parameter as a first order approximation of KL-criterion.
5. Simulation study: finite and asymptotic powers of the log-likelihood ratio test, using different designs.

Goal and notation

Non-linear Gaussian regression model

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

$x \in \mathcal{X} \subseteq \mathbb{R}^p$: **experimental condition**.
y_i: response variable with **mean function** $\eta(x_i;\beta)$ and **error variance** $\sigma^2 h(x_i;\gamma)$;
$h : \mathbb{R}^p \times \mathbb{R}^s \mapsto \mathbb{R}_+$: **specified** positive function.
$\beta \in \mathbb{R}^m$, σ^2 and $\gamma \in \mathbb{R}^s$: **unknown parameters**;
γ_0 leads to the **homoscedastic model**, i.e. $h(x_i;\gamma_0) = 1$.

Inferential goal: to detect model heteroscedasticity

To test **local alternatives**:

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

Likelihood-based tests

The above inferential issue is tackled by applying a **likelihood-based test** (log-likelihood ratio, score or Wald test), whose asymptotic distribution is a chi-squared r.v. with s df (it is a non-central chi-squared r.v, under H_1).

Aim of the work

To design an experiment: $\xi = \left\{ \begin{array}{ccc} x_1 & \dots & x_k \\ \xi(x_1) & \dots & \xi(x_k) \end{array} \right\}$,

where $\xi(x_j) \approx \dfrac{n_j}{n}$ (proportion of observations to be taken at x_j), which **maximizes** the **asymptotic power** of a likelihood-based test;

D_s-criterion: precise estimation for a subset of parameters

D_s-optimality for γ

The D_s-optimum design for γ minimizes the determinant of the asymptotic covariance matrix of the MLE for γ.

The Fisher information matrix can be partitioned as

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

$$\mathcal{I}(\xi;\beta,\sigma^2,\gamma)^{-1} = \begin{bmatrix} \mathcal{I}^{11} & \mathcal{I}^{12} \\ \mathcal{I}^{21} & \mathcal{I}^{22} \end{bmatrix}$$ asymptotic cov. matrix of MLE

Asymptotic covariance matrix of the MLE for γ is $\mathcal{I}^{22}$:

$\mathcal{I}^{22} = [\mathcal{I}_{22.1}(\xi;\gamma)]^{-1}$, where $\mathcal{I}_{22.1}(\xi;\gamma) = \mathcal{I}_{22} - \mathcal{I}_{12}^T\mathcal{I}_{11}^{-1}\mathcal{I}_{12}$.

Therefore, the D_s-optimum design for γ is

$$\xi_{D_s} = \arg\max_{\xi} |\mathcal{I}_{22.1}(\xi;\gamma)|.$$

Noncentrality parameter and D_s-optimality

Noncentrality parameter

Under local alternatives, $H_1 : \gamma = \gamma_0 + \lambda/\sqrt{n}$, a likelihood-based test-statistic is asymptotically distributed as a chi-squared r.v. with **noncentrality parameter:** $\zeta(\xi; \lambda; \gamma_0) = \lambda^T \mathcal{I}_{22.1}(\xi; \gamma_0)\lambda$.

D_s-optimality: a criterion for testing hypothesis

At $\gamma = \gamma_0$, $\xi_{D_s} = \arg\max_\xi |\mathcal{I}_{22.1}(\xi; \gamma_0)|$, therefore ξ_{D_s} maximizes "in some sense" $\zeta(\xi; \lambda; \gamma_0)$ (for any value of λ).

Scalar case ($s = 1$)

The D_1-optimal design:

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

maximizes the noncentrality parameter.

KL-optimality: a criterion for model discrimination

Let $f_1(y, x, \theta_1)$ and $f_2(y, x, \theta_2)$ be two **rival statistical models**.

If $f_1(y, x, \theta_1)$ is the **true** (known) model (which in general may include or not $f_2(y, x, \theta_2)$ as a special case) then the **Kullback–Leibler divergence** between f_1 and f_2 is:

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

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

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

$$\xi_{\theta_1}^{KL} = \arg\max_{\xi} I_{2,1}(\xi; \theta_1)$$

Discriminate heteroscedastic and homoscedastic models

Given the regression model $y_i = \eta(x_i; \beta) + \varepsilon_i$, $(i = 1, \dots, n)$ the KL-criterion can be applied to discriminate between

$$\varepsilon_i \sim N(0; \sigma^2 h(x_i; \gamma_1)) \quad \text{and} \quad \varepsilon_i \sim N(0; \sigma^2)$$

Let $(\beta_1^T, \sigma_1, \gamma_1^T)$ be the assumed known parameter values, we have that:

KL-criterion

$$I_{12}(\xi; \gamma_1) = 1 + \log A_h - \log G_h$$

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.

Analytical expression for the KL-optimum design

Usually, the KL-optimum design must be computed **numerically** and the computation is cumbersome.
In this context an analytical expression is available.

KL-optimum design

A KL-optimal design is

$$\xi_{\gamma_1}^{KL} = \left\{ \begin{array}{cc} \arg\inf_x h(x;\gamma_1) & \arg\sup_x h(x;\gamma_1) \\ \omega & 1-\omega \end{array} \right\},$$

$$\omega = \left(\frac{\overline{h}}{\overline{h}-\underline{h}} - \frac{1}{\log \overline{h} - \log \underline{h}} \right)$$

where $\underline{h} = \inf_x h(x;\gamma_1) > 0$ and $\overline{h} = \sup_x h(x;\gamma_1) < \infty$.
If $n\,\omega$ is not an integer number, then the best approximation is its integer part.

The "limiting" KL-optimum design

Under mild assumptions, the following theorem proves that, when n goes to infinity (and thus $\gamma_1 \to \gamma_0$ since $\gamma_1 = \gamma_0 + \lambda/\sqrt{n}$), the KL-optimum design tends to become equally supported at $\arg\inf_x h(x;\gamma_1)$ and $\arg\sup_x h(x;\gamma_1)$.

Limiting KL-optimum design

If

1. $\underline{x} = \arg\inf_x h(x;\gamma_1)$ and $\overline{x} = \arg\sup_x h(x;\gamma_1)$ do not depend on γ_1
2. $h(x;\gamma_1)$ is such that $\overline{h}/\underline{h} \to 1$ as $\gamma_1 \to \gamma_0$

then, the limiting KL-optimal design is

$$\xi_{\gamma_0}^{KL} = \left\{ \begin{array}{cc} \underline{x} & \overline{x} \\ 0.5 & 0.5 \end{array} \right\}.$$

KL-criterion and noncentrality parameter

From a Taylor expansion of $I_{12}(\xi;\gamma_1) = 1 + \log A_h - \log G_h$ at γ_0:

Connection between KL-criterion and noncentrality parameter

$$I_{12}(\xi;\gamma_1) = I_{12}\left(\xi;\frac{\lambda}{\sqrt{n}}\right) = 1 + \frac{1}{n}\zeta(\xi;\lambda;\gamma_0) + O\left(\frac{||\lambda||^3}{n^{\frac{3}{2}}}\right).$$

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

$$\xi^{KL}_{\gamma_1} = \arg\sup_{\xi} I_{12}\left(\xi;\frac{\lambda}{\sqrt{n}}\right) \to \arg\sup_{\xi} \zeta(\xi;\lambda;\gamma_0)$$

$\xi^{KL}_{\gamma_0}$ 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 $\zeta(\xi,\lambda,\gamma_0)$.

Comparing D_s- and KL-criteria

1. When γ is scalar, the D_1-optimal design is superior to the KL-optimum design, because the D_1-criterion is proportional to the noncentrality parameter.
   However, for n large enough, the KL-optimum design should almost coincide with ξ_{D_1}.
2. When γ is not scalar, with a large sample size n, the KL-optimum design should be superior to ξ_{D_s} because it approximatively maximizes the noncentrality parameter.
   Asymptotically, the KL-optimum design is

$$\xi_{\gamma_0}^{KL} = \arg\sup_{\xi} \zeta(\xi, \lambda, \gamma_0).$$

Simulation study: the model

We consider a linear regression model

$$y_i = \beta_0 + \beta_1 x + \varepsilon_i,\ \varepsilon_i \sim N(0; \sigma^2 h(x_i; \gamma)),$$

with $\beta_0 = \beta_1 = 1$, $\sigma^2 = 1$, $x \in [0, 1]$ and 3 different variance functions:

Case 1: $h(x, \gamma) = e^{\gamma x}$ (h monotone in x; γ scalar);
Case 2: $h(x, \gamma) = 1 + \gamma x + \sin(2\pi\gamma x)$ (h non-monotone in x);
Case 3: $h(x, \gamma) = 1 + \gamma_1 x + \gamma_2 x^2$. (γ not scalar)

We compute the powers of the LR-statistic to test

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

using several KL-optimal designs (for different choices of λ), the D_s-optimal design at $\gamma = 0$ and the uniform design:

$$\xi_{U_5} = \left\{ \begin{array}{ccccc} 0 & 0.25 & 0.5 & 0.75 & 1 \\ 0.2 & 0.2 & 0.2 & 0.2 & 0.2 \end{array} \right\}.$$

Case 1: $h(x,\gamma) = e^{\gamma x}$ (h monotone in x; γ scalar)

λ	n	$\gamma_1 = \frac{\lambda}{\sqrt{n}}$	$\xi_{\gamma_1}^{KL}$ $\hat{\alpha}$	$\xi_{\gamma_1}^{KL}$ $1-\hat{\beta}$	ξ_{D_1} $\hat{\alpha}$	ξ_{D_1} $1-\hat{\beta}$	ξ_{U_5} $\hat{\alpha}$	ξ_{U_5} $1-\hat{\beta}$
5	25	1	0.0680	0.3777	0.0672	0.4090	0.0686	0.2362
	100	0.5	0.0555	0.4185	0.0509	0.4280	0.0548	0.2342
	400	0.25	0.0463	0.4334	0.0500	0.4203	0.0488	0.2434
	∞	0	0.0500	0.7054	0.0500	0.7054	0.0500	0.2532
10	25	2	0.0674	0.8813	0.0672	0.9196	0.0672	0.6464
	100	1	0.0517	0.9304	0.0509	0.9371	0.0548	0.6937
	400	0.5	0.0517	0.9353	0.0500	0.9446	0.0488	0.7035
	∞	0	0.0500	0.9988	0.0500	0.9988	0.0500	0.9424
20	25	4	0.0863	0.9962	0.0672	1.0000	0.0672	0.9886
	100	2	0.0591	1.0000	0.0509	1.0000	0.0548	0.9976
	400	1	0.0557	1.0000	0.0500	1.0000	0.0488	0.9990
	∞	0	0.0500	1.0000	0.0500	1.0000	0.0500	1.0000

Simulation study: the Monte Carlo powers $1-\hat{\beta}$

Case 2: $h(x,\gamma) = 1 + \gamma x + \sin(2\pi\gamma x)$ (h non-monotone; γ scalar)

λ	n	$\gamma_1 = \frac{\lambda}{\sqrt{n}}$	$\xi^{KL}_{\gamma_1}$		ξ_{D_1}		ξ_{U_5}	
			$\hat{\alpha}$	$1-\hat{\beta}$	$\hat{\alpha}$	$1-\hat{\beta}$	$\hat{\alpha}$	$1-\hat{\beta}$
10	100	1	0.0000	0.0003	0.0000	0.0021	0.0000	0.0005
	400	0.5	0.0081	0.0305	0.0085	0.0108	0.0036	0.0120
	1600	0.25	0.0480	0.3743	0.0482	0.3709	0.0596	0.2222
	6400	0.125	0.0506	0.6096	0.0492	0.6159	0.0572	0.3921
	25600	0.0625	0.0489	0.6916	0.0514	0.6943	0.0512	0.4049
	∞	0	0.0500	0.9537	0.0500	0.9537	0.0500	0.7307
20	100	2	0.0000	0.0217	0.0000	0.0389	0.0000	0.0090
	400	1	0.0000	0.0899	0.0085	0.0814	0.0036	0.0739
	1600	0.5	0.0493	0.3819	0.0482	0.0963	0.0596	0.2709
	6400	0.25	0.0486	0.9133	0.0501	0.9125	0.0596	0.7274
	25600	0.125	0.0518	0.9939	0.0514	0.9937	0.0541	0.8952
	∞	0	0.0500	1.0000	0.0500	1.0000	0.0500	0.9993
40	100	4	0.0000	0.2236	0.0000	0.1945	0.0000	0.0635
	400	2	0.0000	0.3543	0.0085	0.2547	0.0036	0.1579
	1600	1	0.0109	0.5399	0.0482	0.2660	0.0596	0.4365
	6400	0.5	0.0521	0.9206	0.0501	0.2932	0.0596	0.7294
	25600	0.25	0.0492	1.0000	0.0514	1.0000	0.0541	0.9955
	∞	0	0.0500	1.0000	0.0500	1.0000	0.0500	1.0000

Simulation study: the Monte Carlo powers $1-\hat{\beta}$

Case 3: $h(x,\gamma) = 1 + \gamma_1 x + \gamma_2 x^2$ (γ not scalar)

$\lambda^T = (\lambda_1, \lambda_2)$	n	$\gamma_1^T = \left(\frac{\lambda_1}{\sqrt{n}}, \frac{\lambda_2}{\sqrt{n}}\right)$	$\xi_{\gamma_1}^{KL}$ $\hat{\alpha}$	$\xi_{\gamma_1}^{KL}$ $1-\hat{\beta}$	ξ_{D_2} $\hat{\alpha}$	ξ_{D_2} $1-\hat{\beta}$	ξ_{U_5} $\hat{\alpha}$	ξ_{U_5} $1-\hat{\beta}$
(2.5, 2.5)	25	(0.5 , 0.5)	0.0202	0.1104	0.0782	0.1522	0.0973	0.1529
	100	(0.25 , 0.25)	0.0168	0.1532	0.0556	0.1679	0.0634	0.1457
	400	(0.125 , 0.125)	0.0129	0.1861	0.0535	0.1980	0.0532	0.1661
	∞	(0 , 0)	0.0500	0.6028	0.0500	0.4384	0.0500	0.3400
(2.5, 5)	25	(0.5 , 1)	0.0218	0.1837	0.0782	0.2160	0.0973	0.2095
	100	(0.25 , 0.5)	0.0194	0.3083	0.0556	0.2802	0.0634	0.2445
	400	(0.125 , 0.25)	0.0146	0.4165	0.0535	0.3598	0.0532	0.2906
	∞	(0 , 0)	0.0500	0.9291	0.0500	0.8038	0.0500	0.6753
(5, 5)	25	(1 , 1)	0.0206	0.2697	0.0782	0.2766	0.0973	0.2459
	100	(0.5 , 0.5)	0.0157	0.4855	0.0556	0.4211	0.0634	0.3264
	400	(0.25 , 0.25)	0.0133	0.6579	0.0535	0.5444	0.0532	0.4372
	∞	(0 , 0)	0.0500	0.9965	0.0500	0.9668	0.0500	0.9028
(2.5, 10)	25	(0.5 , 2)	0.0240	0.3744	0.0782	0.3418	0.0973	0.3143
	100	(0.25 , 1)	0.0158	0.6458	0.0556	0.5596	0.0634	0.4500
	400	(0.125 , 0.5)	0.0159	0.8361	0.0535	0.7241	0.0532	0.5969
	∞	(0 , 0)	0.0500	1.0000	0.0500	0.9983	0.0500	0.9876
(5, 10)	25	(1 , 2)	0.0221	0.4560	0.0782	0.4119	0.0973	0.3558
	100	(0.5 , 1)	0.0163	0.7638	0.0556	0.6509	0.0634	0.5378
	400	(0.25 , 0.5)	0.0149	0.9280	0.0535	0.8376	0.0532	0.7131
	∞	(0 , 0)	0.0500	1.0000	0.05	1.0000	0.05	0.9990
(10, 10)	25	(2 , 2)	0.0231	0.5664	0.0782	0.5232	0.0973	0.4285
	100	(1 , 1)	0.0163	0.9067	0.0556	0.8084	0.0634	0.6919
	400	(0.5 , 0.5)	0.0135	0.9908	0.0535	0.9558	0.0532	0.8835
	∞	(0 , 0)	0.0500	1.0000	0.0500	1.0000	0.0500	1.0000

Simulation study: the designs

Case 1

$$\xi_{D_1} = \begin{Bmatrix} 0 & 1 \\ 0.500 & 0.500 \end{Bmatrix} \quad \xi^{KL}_{0.25} = \begin{Bmatrix} 0 & 1 \\ 0.521 & 0.479 \end{Bmatrix} \quad \xi^{KL}_{0.5} = \begin{Bmatrix} 0 & 1 \\ 0.542 & 0.458 \end{Bmatrix}$$

$$\xi^{KL}_{1} = \begin{Bmatrix} 0 & 1 \\ 0.582 & 0.418 \end{Bmatrix} \quad \xi^{KL}_{2} = \begin{Bmatrix} 0 & 1 \\ 0.657 & 0.343 \end{Bmatrix} \quad \xi^{KL}_{4} = \begin{Bmatrix} 0 & 1 \\ 0.769 & 0.231 \end{Bmatrix}$$

Case 2

$$\xi_{D_1} = \begin{Bmatrix} 0 & 1 \\ 0.500 & 0.500 \end{Bmatrix} \quad \xi^{KL}_{0.0625} = \begin{Bmatrix} 0 & 1 \\ 0.504 & 0.496 \end{Bmatrix} \quad \xi^{KL}_{0.125} = \begin{Bmatrix} 0 & 1 \\ 0.507 & 0.493 \end{Bmatrix} \quad \xi^{KL}_{0.25} = \begin{Bmatrix} 0 & 1 \\ 0.510 & 0.490 \end{Bmatrix}$$

$$\xi^{KL}_{0.5} = \begin{Bmatrix} 0.00 & 0.55 \\ 0.510 & 0.490 \end{Bmatrix} \quad \xi^{KL}_{1} = \begin{Bmatrix} 0.72 & 0.28 \\ 0.512 & 0.488 \end{Bmatrix} \quad \xi^{KL}_{2} = \begin{Bmatrix} 0.36 & 0.64 \\ 0.520 & 0.480 \end{Bmatrix} \quad \xi^{KL}_{4} = \begin{Bmatrix} 0.18 & 0.89 \\ 0.532 & 0.468 \end{Bmatrix}$$

Case 3

$$\xi_{D_2} = \begin{Bmatrix} 0 & 0.5 & 1 \\ 0.333 & 0.333 & 0.333 \end{Bmatrix} \quad \xi^{KL}_{(0,0)} = \begin{Bmatrix} 0.00 & 1.00 \\ 0.5000 & 0.5000 \end{Bmatrix} \quad \xi^{KL}_{(0.05,0.05)} = \begin{Bmatrix} 0.00 & 1.00 \\ 0.5008 & 0.4992 \end{Bmatrix}$$

$$\xi^{KL}_{(0.125,0.125)} = \begin{Bmatrix} 0.00 & 1.00 \\ 0.5186 & 0.4814 \end{Bmatrix} \quad \xi^{KL}_{(0.125,0.25)} = \begin{Bmatrix} 0.00 & 1.00 \\ 0.5265 & 0.4735 \end{Bmatrix} \quad \xi^{KL}_{(0.25,0.25)} = \begin{Bmatrix} 0.00 & 1.00 \\ 0.5337 & 0.4663 \end{Bmatrix}$$

$$\xi^{KL}_{(0.125,0.5)} = \begin{Bmatrix} 0.00 & 1.00 \\ 0.5403 & 0.4597 \end{Bmatrix} \quad \xi^{KL}_{(0.25,0.5)} = \begin{Bmatrix} 0.00 & 1.00 \\ 0.5464 & 0.4536 \end{Bmatrix} \quad \xi^{KL}_{(0.5,0.5)} = \begin{Bmatrix} 0.00 & 1.00 \\ 0.5573 & 0.4427 \end{Bmatrix}$$

$$\xi^{KL}_{(0.25,1)} = \begin{Bmatrix} 0.00 & 1.00 \\ 0.5668 & 0.4332 \end{Bmatrix} \quad \xi^{KL}_{(0.5,1)} = \begin{Bmatrix} 0.00 & 1.00 \\ 0.5753 & 0.4247 \end{Bmatrix} \quad \xi^{KL}_{(1,1)} = \begin{Bmatrix} 0.00 & 1.00 \\ 0.5898 & 0.4102 \end{Bmatrix}$$

$$\xi^{KL}_{(0.5,2)} = \begin{Bmatrix} 0.00 & 1.00 \\ 0.6018 & 0.3982 \end{Bmatrix} \quad \xi^{KL}_{(1,2)} = \begin{Bmatrix} 0.00 & 1.00 \\ 0.6120 & 0.3880 \end{Bmatrix} \quad \xi^{KL}_{(2,2)} = \begin{Bmatrix} 0.00 & 1.00 \\ 0.6287 & 0.3713 \end{Bmatrix}$$

Remark. In Cases 1 and 2, as $\gamma = \lambda/\sqrt{n}$ goes to $\gamma_0 = 0$, the ξ^{KL}_{γ} approaches ξ_{D_1}, consistently with the theoretical results. This is not true for the multi-dimensional Case 3.

References

Dette, H. and Titoff, S. (2008). Optimal discrimination designs. Annals of Statistics, 37(4), 2056-2082.

López-Fidalgo, J., Tommasi, C. and Trandafir, P.C. (2007). An optimal experimental design criterion for discriminating between non-normal models. JRSSB, 69(2), 231-242.

Tommasi, C. and López-Fidalgo, J. (2010). Bayesian optimum designs for discriminating between models with any distribution. CSDA, 54, 143-150.

Tung, S. H. (1975). On lower and upper bounds of the difference between the arithmetic and the geometric mean. Mathematics of Computation, 29(131), 834-836.

Thanks a lot for your attention.

Do you have any questions?