Maximum Likelihood Inference and Adaptation for Non-Linear Models with Gaussian Errors

∼ GMDS CEN - IBS 2020 ∼

∼ Caterina May ∼
∼ joint work with Chiara Tommasi and Nancy Flournoy ∼

Università del Piemonte Orientale, University of Missouri and Università di Milano

2th December 2020

Outline

Background and notation

- Assume n independent observations follow the model

$$y_j = \eta(x_j, \theta) + \varepsilon_j, \quad \varepsilon_j \sim \mathcal{N}(0, \sigma^2), \quad j = 1, \ldots, n;$$

  y_j: response of the unit j, treated at the dose $x_j \in \mathscr{X}$;

$$\theta \in \Theta \subseteq \mathbb{R}^{p+1}.$$

- Experimental design

$$\xi = \begin{Bmatrix} x_1 & \cdots & x_M \\ \omega_1 & \cdots & \omega_M \end{Bmatrix}, \quad 0 \leq \omega_m \leq 1, \sum_{m=1}^{M} \omega_m = 1$$

  x_m: different experimental conditions to be used in the analysis, $m = 1, \ldots, M$;
  ω_m: proportions of units to be taken at each condition.

Inferential goal: Precise estimation of θ

Best choice of the experimental conditions

- **Optimal design:** $\xi^*(\theta) = \arg\max_\xi \Phi[M(\xi;\theta)]$.
- **Information matrix:**

$$M(\xi;\theta) = \int_{\mathscr{X}} \nabla\eta(x,\theta)\nabla\eta(x,\theta)^T d\xi(x)$$

- $\Phi(\cdot)$: functional which reflects the inferential aim, since the inverse of $M(\xi;\theta)$ is (approximately) proportional to the variance of the maximum likelihood estimator (MLE).
- **Locally optimal design**: $\xi^*(\theta)$ can be computed only if a guessed value θ_0 for θ is available.

A two-stage adaptive procedure: a way to overcome the problem of the choice of θ_0

A two-stage adaptive procedure

First stage

1. Fix a first stage design:
   $\xi_1 = \left\{ \begin{array}{ccc} x_{11} & \cdots & x_{1M_1} \\ \omega_{11} & \cdots & \omega_{1M_1} \end{array} \right\}$.
2. Take n_1 independent responses according to ξ_1:

   for $m = 1, \ldots, M_1$, take $n_{1m} \simeq n_1 \omega_{1m}$ independent observations at x_{1m} such that $\sum_{m=1}^{M_1} n_{1m} = n_1$;

   $\{y_{1mj}\}_{1,1}^{M_1, n_{1m}}$ denote the first stage observations.

3. Compute the MLE $\hat{\theta}_{n_1}$ using the first stage observations $\{y_{1mj}\}_{1,1}^{M_1, n_{1m}}$:

   $\hat{\theta}_{n_1} = \hat{\theta}_{n_1}(\bar{y}_1)$ where $\bar{y}_1 = (\bar{y}_{11}, \ldots, \bar{y}_{1M_1})^T$ is the complete sufficient statistic for θ.

A two-stage adaptive procedure

Second stage

1. Determine a second stage **locally optimal** design where $\hat{\theta}_{\boldsymbol{n}_1} = \hat{\theta}_{n_1}(\bar{y}_1)$ is used as guessed value for θ:
   $\xi_2^* = \xi^*(\hat{\theta}_{n_1}) = \left\{ \begin{array}{ccc} x_{21} & \cdots & x_{2M_2} \\ \omega_{21} & \cdots & \omega_{2M_2} \end{array} \right\},$

   $\Rightarrow$ ξ_2^* depends on the first stage observations through $\bar{y}_1$
2. Given $\bar{y}_1$, take n_2 (conditionally) independent responses accordingly to ξ_2^*:

   for $m = 1, \ldots, M_2$, take $n_{2m} \simeq n_2 \omega_{2m}$ independent observations at x_{2m} such that $\sum_{m=1}^{M_2} n_{2m} = n_2$;

   $\{y_{2mj}\}_{1,1}^{M_2, n_{2m}}$ denote the second stage observations.
3. Compute the MLE $\hat{\theta}_n$ using the **whole** set of $n = n_1 + n_2$ observations.

Model for the the first and the second stage responses

Natural assumption

Second stage observations $\{y_{2mj}\}_{1,1}^{M_2,n_{2m}}$ depend on the first stage information $\{y_{1mj}\}_{1,1}^{M_1,n_{1m}}$ only through ξ_2^*.

$$\Downarrow$$

Given ξ_2^*, $\{y_{2mj}\}_{1,1}^{M_2,n_{2m}}$ are conditionally independent on $\{y_{1mj}\}_{1,1}^{M_1,n_{1m}}$ and we have the following model for the whole set of observations $\{y_{imj}\}_{1,1,1}^{2,M_i,n_{im}}$:

$$y_{imj} = \eta(x_{im}, \theta) + \varepsilon_{imj}, \quad \varepsilon_{imj} \sim \mathcal{N}\left(0, \sigma^2\right).$$

Asymptotic properties of the MLE

Three scenarios are possible in the two-stage experiment:

1. Both stages sample size n_1 and n_2 increase to infinity:

$$n_1 \to +\infty, \quad n_2 \to +\infty.$$

   ⇒ classical approach which eliminates the dependence between stages (Dragalin et al. (2008), Dette et al. (2012), Pronzato and Pazman (2013), Kim and Flournoy (2015))

2. Only firs-stage sample size n_1 grows to infinity

   ⇒ no need of second stage observations

3. Only second-stage sample size n_2 goes to infinity, while n_1 is fixed:

$$n_1 < +\infty, \quad n_2 \to +\infty.$$

   ⇒ considered by Lane, Flournoy et al. (2012, 2014) for the uni-dimensional parameter

In this work: extension to multi-parameter models (non trivial)

Likelihood and score function

- In our two-stage procedure the total **likelihood** can be factorized as:

$$\mathscr{L}_n(\theta|\bar{y}_1,\bar{y}_2) \propto \mathscr{L}_{1n}(\theta|\bar{y}_1)\cdot\mathscr{L}_{2n}(\theta|\bar{y}_2), \quad \text{where}$$

$$\mathscr{L}_{in}(\theta|\bar{y}_i) = \exp\left\{-\frac{1}{2\sigma^2}\sum_{m=1}^{M_i} n_{im}\left[\bar{y}_{im}-\eta(x_{im},\theta)\right]^2\right\}, \quad i=1,2,$$

- The **total score** is

$$S = \nabla \ln \mathscr{L}_n(\theta|\bar{y}_1,\bar{y}_2) = S_1 + S_2, \quad \text{where}$$

$$S_i = \nabla \mathscr{L}_{in}(\theta|\bar{y}_i) = \frac{1}{\sigma^2}\sum_{m=1}^{M_i} n_{im}\left[\bar{y}_{im}-\eta(x_{im},\theta)\right]\nabla\eta(x_{im},\theta)$$

is the score for the i-th stage, $i=1,2$.

Per-subject Fisher information matrix

- $\bar{y}_2$ depends on $\bar{y}_1$ only through ξ_2^*.
- Given $\bar{y}_1$ the second stage design ξ_2^* is completely determined.

$$\Downarrow$$

$\mathrm{E}_{\bar{y}_2|\bar{y}_1}[S_2] = 0$ and the **per-subject Fisher information matrix** is

$$\begin{aligned} \frac{1}{n}\mathrm{Cov}_{\bar{y}_1,\bar{y}_2}[S,S] &= \frac{1}{n\sigma^2}\left\{\sum_{m=1}^{M_1} n_{1m}\nabla\eta(x_{1m},\theta)\nabla\eta(x_{1m},\theta)^{\mathrm{T}}\right. \\ &+ \left.\mathrm{E}_{\bar{y}_1}\sum_{m=1}^{M_2}\left[n_{2m}\nabla\eta(x_{2m},\theta)\nabla\eta(x_{2m},\theta)^{\mathrm{T}}\right]\right\} \end{aligned}$$

Asymptotic per-subject Fisher information matrix

As $n \to +\infty$,

$$\frac{1}{n}\mathrm{Cov}_{\bar{y}_1,\bar{y}_2}[S,S] \longrightarrow \frac{1}{\sigma^2}\mathrm{E}_{\bar{y}_1}\left[\int_{\mathscr{X}}\nabla\eta(x,\theta)\nabla\eta(x,\theta)^{\mathrm{T}}d\xi_2^*(x)\right]$$

Theoretical results

Let $\widehat{\theta}_n$ be the MLE maximing the total likelihood; assume n_1 is fixed and $n_2 \to \infty$, then (under standard regularity conditions)

- **Theorem 1.**
$$\widehat{\theta}_n \xrightarrow{P} \theta^t,$$
where θ^t denotes the true unknown value of θ.
- **Theorem 2.**
$$\sqrt{n}\left(\widehat{\theta}_n - \theta^t\right) \xrightarrow{\mathscr{D}} \sigma\, M(\xi_2^*, \theta^t)^{-1/2}\, \boldsymbol{Z}, \tag{1}$$
where $\boldsymbol{Z}$ is a $(p+1)$-dimensional standard normal random vector independent of the random matrix $M(\xi_2^*, \theta^t)$.

Theoretical results

Corollary. The asymptotic variance of $\sqrt{n}\left(\widehat{\theta}_n - \theta^t\right)$ is

$$\sigma^2 \mathrm{E}_{\bar{y}_1}\left[\left(\int_{\mathscr{X}} \nabla\eta(x,\theta^t)\nabla\eta(x,\theta^t)^{\mathrm{T}}\, d\xi_2^*(x)\right)^{-1}\right]$$

- **Note:** compare this result with the asymptotic per-subject Fisher information matrix previously obtained:

  $$\frac{1}{n}\mathrm{Cov}_{\bar{y}_1,\bar{y}_2}[\mathrm{S},\mathrm{S}] \longrightarrow \frac{1}{\sigma^2}\mathrm{E}_{\bar{y}_1}\left[\int_{\mathscr{X}} \nabla\eta(x,\theta^t)\nabla\eta(x,\theta^t)^{\mathrm{T}} d\xi_2^*(x)\right],$$

  as $n \to +\infty$.

Examples

Emax model:

$$\eta(x,\theta) = \theta_0 + \theta_1 \frac{x}{x+\theta_2},$$

Exponential model:

$$\eta(x,\theta) = \theta_0 + \theta_1 \exp(x/\theta_2),$$

$x \in \mathscr{X} = [a,b]$, $0 \leq a < b$.

The locally D-optimal design ξ_D^* for these models (Dette et Al., 2010) is

$$\xi_D^* = \left\{ \begin{matrix} a & x^* & b \\ 1/3 & 1/3 & 1/3 \end{matrix} \right\},$$

where

$$x^*(\theta_2) = \frac{b(a+\theta_2)+a(b+\theta_2)}{(a+\theta_2)+(b+\theta_2)} \qquad (\textit{Emax} \text{ model}),$$

$$x^*(\theta_2) = \frac{(b-\theta_2)\exp(b/\theta_2)-(a-\theta_2)\exp(a/\theta_2)}{\exp(b/\theta_2)-\exp(a/\theta_2)} \quad (\textit{Exponential} \text{ model}),$$

Fixed and adaptive designs: a comparison

Assume that a guessed value θ_0 for θ is available; take n_1 observations according to a locally D-optimal design

$$\xi_1^* = \xi^*(\theta_0),$$

then:

- in the fixed design add n_2 observations according to the same ξ_1^*, independently on the first stage;
- in the adaptive design, instead, add n_2 observations according to the locally D-optimal design

$$\xi_2^* = \xi^*(\hat{\theta}_{n_1})$$

Goal: compare the two-stage adaptive procedure with the fixed design in terms of precise estimation, for finite n.

Simulation study

We generate 10.000 experiments for each procedure with:

- $[a, b] = [0, 150]$
- $\theta_0^t = 2$ and $\theta_1^t = 0.467$
- $\theta_2^t = 50, 100$
- the nominal value $\theta_{2,0}$ ranging over $[10, 200]$
- $\sigma = 0.1$
- $n_1 = 27, 60,\ n_2 = 90, 270, 600$

$\widehat{\theta}_n^F$ and $\widehat{\theta}_n^A$ denote the MLEs obtained from the fixed and the adaptive procedures, respectively.

We compare the precision of MLEs :

- by computing their Montecarlo generalized variances $\mathrm{Det}\left[\mathrm{Var}\left(\widehat{\theta}_n^F\right)\right]$ and $\mathrm{Det}\left[\mathrm{Var}\left(\widehat{\theta}_n^A\right)\right]$
- by computing their mean square errors $\mathrm{MSE}\left(\widehat{\theta}_n^F\right)$ and $\mathrm{MSE}\left(\widehat{\theta}_n^A\right)$

Results

Emax model.

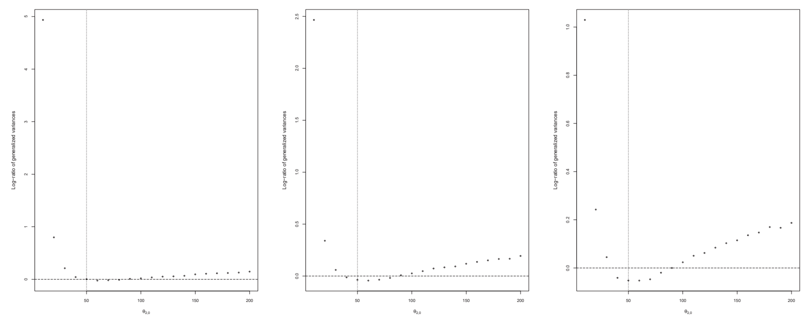

Figure: Logarithm of $\mathrm{Det}\left[\mathrm{Var}\left(\widehat{\theta}_n^F\right)\right] / \mathrm{Det}\left[\mathrm{Var}\left(\widehat{\theta}_n^A\right)\right]$ (y-axis) versus $\theta_{2,0}$ (x-axis). Vertical line: value of θ_2^t.
$n_1 = 60$; in the left-side panel $n_2 = 90$, in the center $n_2 = 270$ and in the right-hand side panel $n_2 = 600$.

Results

Exponential model.

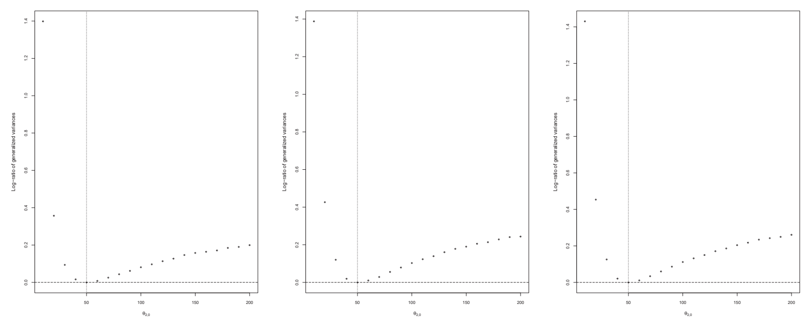

Figure: Logarithm of $\mathrm{Det}\left[\mathrm{Var}\left(\widehat{\theta}_n^F\right)\right]/\mathrm{Det}\left[\mathrm{Var}\left(\widehat{\theta}_n^A\right)\right]$ (y-axis) versus $\theta_{2,0}$ (x-axis). Vertical line: value of θ_2^t.
$n_1 = 60$; in the left-side panel $n_2 = 90$, in the center $n_2 = 270$ and in the right-hand side panel $n_2 = 600$.

Results

Emax model:

		Fixed procedure			Adaptive Procedure		
$\theta_{2,0}$	n_2	MSE($\widehat{\theta_0}$)	MSE($\widehat{\theta_1}$)	MSE($\widehat{\theta_2}$)	MSE($\widehat{\theta_0}$)	MSE($\widehat{\theta_1}$)	MSE($\widehat{\theta_2}$)
20	90	0.00020	0.00316	581.78	0.00019	0.00215	291.50
50	90	0.00019	0.00189	226.85	0.00019	0.00196	229.43
80	90	0.00020	0.00210	226.23	0.00020	0.00204	228.69
20	270	0,00009	0.00090	149.44	0,00009	0.00087	105.13
50	270	0,00009	0.00078	86.916	0,00009	0.00082	90.744
80	270	0,00009	0.00090	93.424	0,00009	0.00086	94.743
20	600	0.00004	0.00041	61.907	0.00004	0.00042	47.247
50	600	0.00004	0.00038	41.219	0.00004	0.00040	43.405
80	600	0.00004	0.00043	43.679	0.00004	0.00040	44.218

Table: Performance of fixed and adaptive designs in terms of MSE; $n_1 = 60$ and three nominal values: $\theta_{2,0} = \theta_2^t$ and $\theta_{2,0} = \theta_2^t \pm 30$, with $\theta_2^t = 50$.

Results

Eponential model:

		Fixed procedure			Adaptive Procedure		
$\theta_{2,0}$	n_2	MSE($\widehat{\theta}_0$)	MSE($\widehat{\theta}_1$)	MSE($\widehat{\theta}_2$)	MSE($\widehat{\theta}_0$)	MSE($\widehat{\theta}_1$)	MSE($\widehat{\theta}_2$)
20	90	0.21824	0.00016	0.18700	0.21822	0.00012	0.13232
50	90	0.21828	0.00010	0.11565	0.21827	0.00010	0.11566
80	90	0.21822	0.00012	0.12668	0.21822	0.00011	0.11915
20	270	0.21807	0.00007	0.08492	0.21807	0.00005	0.05506
50	270	0.21837	0.00005	0.05293	0.21836	0.00005	0.05293
80	270	0.21811	0.00005	0.05864	0.21810	0.00005	0.05422
20	600	0.21806	0.00004	0.04338	0.21806	0.00002	0.02733
50	600	0.21815	0.00002	0.02642	0.21814	0.00002	0.02642
80	600	0.21807	0.00003	0.02840	0.21807	0.00002	0.02594

Table: Performance of fixed and adaptive designs in terms of MSE; $n_1 = 60$ and three nominal values: $\theta_{2,0} = \theta_2^t$ and $\theta_{2,0} = \theta_2^t \pm 30$, with $\theta_2^t = 50$.

Conclusions

1. The simulations suggest that unless a researcher can trust in a guess which is very close to the true value, it is beneficial to adapt
2. The size of n_1 relative to n is also important; we recommend calculating a locally optimal stage 1 sample size as in Lane et al. (2014)
3. The asymptotic approximation for the MLE distribution in the two-stage adaptive procedure following from this study might be closer to the finite distribution than in the standard approach (as showed by Lane et al. (2014) by Lin et al. (2020) for the one-parameter case)

Reference: Flournoy N., May C., Tommasi C. (2018). The Effects of Adaptation on Inference for Non-Linear Regression Models with Normal Errors. *Arxiv preprint 1812.03970: https://arxiv.org/abs/1812.03970*