On the behaviour of the maximum likelihood estimator for exponential models under a fixed and a two-stage design

Stima di massima verosimiglianza per modelli esponenziali in presenza di un disegno fisso e di un disegno adattivo a due stadi

Caterina May and Chiara Tommasi

Abstract We consider two different design strategies for collecting "optimal" data with the aim of estimating as precisely as possible the vector parameter in a dose-response model. In particular, an exponential model with Gaussian errors is considered, and the maximum likelihood method is applied. Through a simulation study we compare the performance of the the maximum likelihood estimator (MLE) when: a) a locally D-optimum design is used to get a sample of independent observations (fixed procedure); b) a two-stage adaptive experimental procedure is applied to collect data, which are dependent since the second stage D-optimal design is determined by the responses observed at the first stage. In the latter case, the theoretical properties of the MLE are described; differently from the most of the literature, asymptotic theory is applied only in the second stage since the first stage sample size is assumed to be finite.
Abstract *Consideriamo qui due diverse strategie per raccogliere dati "ottimali" allo scopo di stimare con precisione il vettore dei parametri in un modello di risposta alla dose. Consideriamo, in particolare, un modello esponenziale con erori Gaussiani. Mediante uno studio di simulazione confrontiamo l'efficienza dello stimatore di massima verosimiglianza quando: a) si utilizza un disegno localmente D-ottimo ottenendo un campione di osservazioni indipendenti (procedura fissa); b) si utilizza una procedure adattiva a due stadi da cui si ottengono dati che sono dipendenti, dato che il disegno D-ottimo al secondo stadio è determinato dalle risposte osservate al primo stadio. In quest'ultimo caso descriviamo le proprietà teoriche dello stimatore di massima verosimiglianza; al contrario di quanto viene fatto normalmente in letteratura, la teoria asintotica viene qui applicata solo al secondo stadio mentre la dimensione campionaria el primo stadio è considerata fissa.*

Caterina May
Università degli Studi del Piemonte Orientale, Department DiSEI, via Perrone, 18, 28100 Novara (Italy), e-mail: caterina.may@uniupo.it

Chiara Tommasi
Università degli Studi di Milano, Department DEMM, via Conservatorio, 7, 20122 Milano (Italy) e-mail: chiara.tommasi@unimi.it

Key words: maximum likelihood inference, D-optimum design, two-stage adaptive design, exponential model, relative efficiency, simulations

1 Introduction

The exponential model is applied in many different contexts (medical, enviromental, pharmaceutical) to interpret dose-response relationships. A three-parameters exponential model can be written as

$$Y = \theta_0 + \theta_1 \exp(x/\theta_2) + \varepsilon, \quad \varepsilon \sim N(0;\sigma^2) \tag{1}$$

where $\theta = (\theta_0, \theta_1, \theta_2)^T$ is a vector of unknown parameters and $\eta(x,\theta) = \theta_0 + \theta_1 \exp(x/\theta_2)$ denotes the response mean at the dose $x \in \mathscr{X} = [a,b]$. The inferential goal is to estimate θ and thus efficient experimental designs are very important. An experimental design ξ can be defined as a finite discrete probability distribution over $\mathscr{X}$; the information matrix of ξ is

$$M(\xi;\theta) = \int_{\mathscr{X}} \nabla\eta(x,\theta)\nabla\eta(x,\theta)^T d\xi(x), \tag{2}$$

where $\nabla\eta(x,\theta)$ denotes the gradient of the mean response $\eta(x,\theta)$ with respect to θ. A D-optimal design $\xi^*(\theta)$ minimizes the generalized asymptotic variance of the maximum likelihood estimator (MLE) of θ, i.e.

$$\xi^*(\theta) = \arg\max_{\xi \in \Xi} |M(\xi;\theta)|, \tag{3}$$

where Ξ is the set of all designs (see [4] and [1]).

Since $\eta(x,\theta)$ is a non-linear model, the D-optimal design (3) depends on the unknown parameter θ. A common approach to tackle this problem is to use a locally optimal design, where θ in (2) is replaced by a guessed value $\tilde{\theta} = (\tilde{\theta}_0, \tilde{\theta}_1, \tilde{\theta}_2)^T$; n independent observations are collected according to this locally D-optimal design $\xi^*(\tilde{\theta})$ and then used to compute the MLE.

Another possibility to obtain the data is to adopt a two-stage procedure. At the first stage a locally D-optimal design $\xi^*(\tilde{\theta})$ is applied to collect n_1 responses (with $n_1 < n$), which are used to estimate the unknown parameter. Let $\hat{\theta}_{n_1}$ be the MLE of θ based on first-stage responses. Then, at the second stage, $n_2 = n - n_1$ additional responses are collected according to another locally D-optimal design, $\xi_2^*(\hat{\theta}_{n_1})$, where $\hat{\theta}_{n_1}$ is used in (2) instead of $\tilde{\theta}$. Finally, the MLE is computed employing the whole sample of $n = n_1 + n_2$ data. Let us note that $\xi_2^*(\hat{\theta}_{n_1})$ is a random probability distribution as it depends on the first-stage observations through $\hat{\theta}_{n_1}$; as a consequence, the second-stage observations are not independent on the first-stage ones. Given $\xi_2^*(\hat{\theta}_{n_1})$, however, the second-stage observations are conditionally independent on the first-stage data, and hence it can be proved that the likelihood function is the same in both the following experimental settings (see Sect. 2.1):

1) n independent observations obtained according to $\xi^*(\tilde{\theta})$ (fixed procedure or one-stage);
2) n_1 independent observations accrued according to $\xi^*(\tilde{\theta})$ and then, given $\hat{\theta}_{n_1}$, other $n_2 = n - n_1$ independent responses coming from $\xi_2^*(\hat{\theta}_{n_1})$ (two-stage procedure).

Let $\hat{\theta}_n^{1S}$ and $\hat{\theta}_n^{2S}$ denote the MLEs when the one-stage and the two-stage procedures, respectively, are adopted to collect the data. In this paper we develop a simulation study to compare the performance of$\hat{\theta}_n^{1S}$ and $\hat{\theta}_n^{2S}$.

2 Theoretical properties of the two-stage design

2.1 Likelihood

The total likelihood is

$$\begin{aligned}\mathscr{L}(\theta;\mathbf{y}_{n_2},\mathbf{x}_{n_2},\mathbf{y}_{n_1},\mathbf{x}_{n_1}) &= \mathscr{L}(\theta;\mathbf{y}_{n_2}|\mathbf{x}_{n_2},\mathbf{y}_{n_1},\mathbf{x}_{n_1})\cdot\mathscr{L}(\mathbf{x}_{n_2}|\mathbf{y}_{n_1},\mathbf{x}_{n_1})\cdot\mathscr{L}(\theta;\mathbf{y}_{n_1}|\mathbf{x}_{n_1})\cdot\mathscr{L}(\mathbf{x}_{n_1})\\ &= \mathscr{L}(\theta;\mathbf{y}_{n_2}|\mathbf{x}_{n_2})\cdot\mathscr{L}(\mathbf{x}_{n_2}|\mathbf{y}_{n_1},\mathbf{x}_{n_1})\cdot\mathscr{L}(\theta;\mathbf{y}_{n_1}|\mathbf{x}_{n_1})\cdot\mathscr{L}(\mathbf{x}_{n_1})\\ &\propto \mathscr{L}(\theta;\mathbf{y}_{n_2}|\mathbf{x}_{n_2})\cdot\mathscr{L}(\theta;\mathbf{y}_{n_1}|\mathbf{x}_{n_1}) \qquad (4)\end{aligned}$$

From the (4) we can see that the total likelihood for the dependent data of the two-stage design is the same as the likelihood with independent data of the fixed design.

2.2 Asymptotics

In order to obtain the consistency and the asymptotic distribution of $\hat{\theta}_n^{2S}$, the classical approach in the literature is to assume that both the sample sizes n_1 and n_2 grow to infinity (see [6]).

A different approach is considered in [5], where n_1 is fixed and only n_2 goes to infinity; this assumption is more realistic in many experimental situations and, in addition, fixing n_1 should improve the approximation of the finite distribution with the asymptotic one. Despite the second stage observations depend on the first-stage data through $\hat{\theta}_{n1}$, the MLE $\hat{\theta}_n^{2S}$ maintains good properties, as stated in the following theorems (see [5] for the proofs).

Theorem 1. *As $n_2 \to \infty$, $\hat{\theta}_n^{2S}$ converges in probability to the true value θ of the parameter.*

Theorem 2. *As $n_2 \to \infty$, $\sqrt{n}\left(\hat{\theta}_n^{2S} - \theta\right)$ converges in distribution to*

$$\sigma\, M[\xi_2^*(\hat{\theta}_{n_1}),\theta]^{-1/2}\,\mathbf{Z},$$

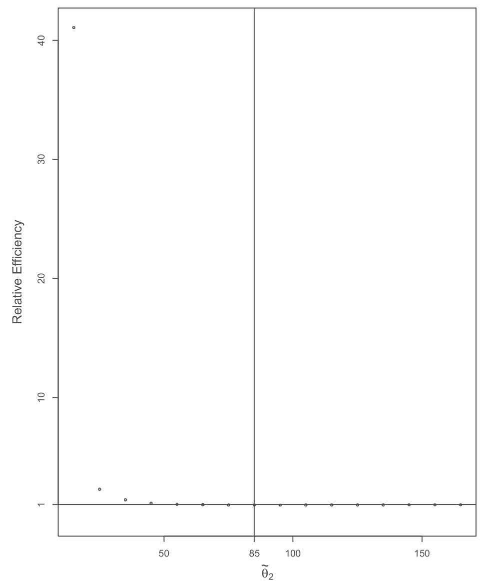

Fig. 1 Relative efficiency $\mathrm{MSE}(\hat{\theta}_n^{1S})/\mathrm{MSE}(\hat{\theta}_n^{2S})$ (on the y-axes) versus different nominal values $\tilde{\theta}_2$ (on the x-axes). The model is exponential with $\theta_0 = -0.08265$, $\theta_1 = 0.08265$, $\theta_2 = 85$ and $\sigma = 0.1$. The vertical line represents the true value of θ_2.

where $\mathbf{Z}$ *is a* 3*-dimensional standard normal random vector independent of the random matrix* $M(\xi_2^*(\hat{\theta}_{n_1}), \theta)$.

Theorem 3. *As* $n_2 \to \infty$,*, the asymptotic variance of* $\sqrt{n}\left(\hat{\theta}_n^{2S} - \theta\right)$ *is*

$$\sigma^2 E\left[\left(\int_{\mathscr{X}} \nabla\eta(x,\theta)\nabla\eta(x,\theta)^{\mathrm{T}}\, d\xi_2^*(\hat{\theta}_{n_1})(x)\right)^{-1}\right] \tag{5}$$

The expression of the asymptotic variance of $\hat{\theta}_n^{2S}$ provided in (5) justifies the use of a D-optimal design to collect the second stage data.

3 Simulations

The goal of the simulation study is to compare the two-stage adaptive procedure with the fixed design in terms of precision of the MLEs, $\hat{\theta}_n^{2S}$ and $\hat{\theta}_n^{1S}$.

The D-optimum design $\xi^*(\theta)$ for the exponential model is provided in [3]. It is a three point design equally supported at the extremes of the experimental domain $\mathscr{X} = [a;b]$ and at

$$x^* = \frac{(b-\theta_2)\exp(b/\theta_2) - (a-\theta_2)\exp(a/\theta_2)}{\exp(b/\theta_2) - \exp(a/\theta_2)}. \tag{6}$$

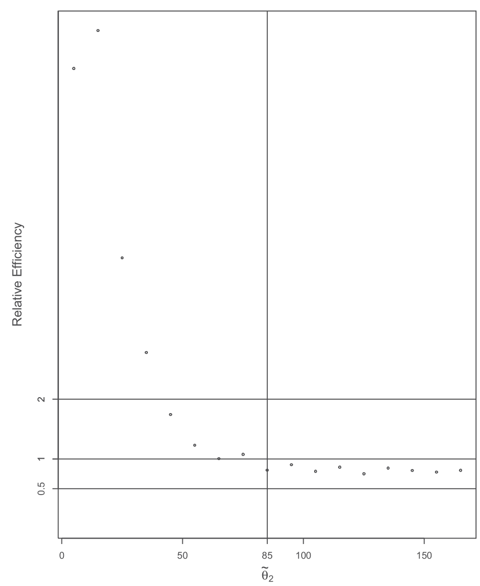

Fig. 2 Relative efficiency $\mathrm{MSE}(\hat{\theta}_n^{1S})/\mathrm{MSE}(\hat{\theta}_n^{2S})$ (on the y-axes) versus different nominal values $\tilde{\theta}_2$ (on the x-axes). The model is exponential with $\theta_0 = -0.08265$, $\theta_1 = 0.08265$, $\theta_2 = 85$ and $\sigma = 0.25$. The vertical line represents the true value of θ_2.

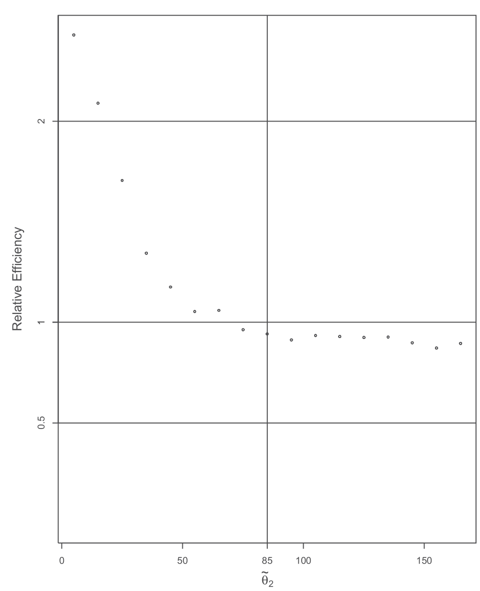

Fig. 3 Relative efficiency $\mathrm{MSE}(\hat{\theta}_n^{1S})/\mathrm{MSE}(\hat{\theta}_n^{2S})$ (on the y-axes) versus different nominal values $\tilde{\theta}_2$ (on the x-axes). The model is exponential with $\theta_0 = -0.08265$, $\theta_1 = 0.08265$, $\theta_2 = 85$ and $\sigma = 0.5$. The vertical line represents the true value of θ_2.

Hence, at each stage, $1/3$ of the observations are taken at a, b and x^*. Note that x^*, and hence $\xi^*(\theta)$, depends only on the non-linear parameter θ_2 of model (1). Herein, we take $a = 0$ and $b = 150$. From model (1) with $\theta_0 = -0.08625$, $\theta_1 = 0.08625$ and $\theta_2 = 85$ and 3 different values for $\sigma = 0.1; 0.25; 0.5$,

1. we generate $n_1 = 30$ independent observations according to $\xi^*(\tilde{\theta}_2)$ to compute the first-stage MLE, $\hat{\theta}_{n_1}$, where $\tilde{\theta}_2 \in (0; 150)$ is a nominal value for θ_2;
2. we generate further $n_2 = 300$ independent observations according to $\xi^*(\hat{\theta}_{n_1})$ to obtain $\hat{\theta}_n^{2S}$;
3. we generate further $n_2 = 300$ independent observations according to $\xi^*(\tilde{\theta}_2)$ to obtain $\hat{\theta}_n^{1S}$.

For each choice of σ and $\tilde{\theta}_2$ we repeat the computation of $\hat{\theta}_n^{1S}$ and $\hat{\theta}_n^{2S}$ 5000 times, to get their Monte Carlo MSEs. Simulations are realized with R package in [2]. Figures 1, 2 and 3 displays the relative efficiency $\mathrm{MSE}(\hat{\theta}_n^{1S})/\mathrm{MSE}(\hat{\theta}_n^{2S})$ for different choices of the nominal value $\tilde{\theta}_2$, and for $\sigma = 0.1; 0.25; 0.5$, respectively.

4 Conclusions

The simulations in Sect.3 show that the two-stage procedure outperforms the one-stage (or fixed) procedure whenever the assumed nominal value $\tilde{\theta}_2$ is much inferior to the true value of θ_2 (in this example $\theta_2 = 85$). For the other values of $\tilde{\theta}_2$, the relative efficiency of the two-stage procedure is around one (never less then 0.5). This behaviour appears to be more pronounced as σ increases. An explanation can be seen in the slope of $x^* = x^*(\theta_2)$ in (6), which is larger for small value of θ_2. Hence, $x^*(\tilde{\theta}_2)$ is far away from the true optimal dose if $\tilde{\theta}_2 < 85$ and replacing $\tilde{\theta}_2$ with the first stage estimate may improve the results.

In conclusion, we suggest to apply the two-stage procedure to collect the data if we do not have enough knowledge about the true value of θ_2, which is often the case in real-life problems.

References

1. A. C. Atkinson, A. N. Donev, and R. D. Tobias. *Optimum experimental designs, with SAS*, volume 34 of *Oxford Statistical Science Series*. Oxford University Press, Oxford, 2007.
2. B. Bornkamp, J. Pinheiro, and F. Bretz. *DoseFinding: Planning and Analyzing Dose Finding Experiments*, 2018. R package version 0.9-16.
3. H. Dette, C. Kiss, M. Bevanda, and F. Bretz. Optimal designs for the emax, log-linear and exponential models. *Biometrika*, 97(2):513–518, 2010.
4. V. Fedorov. *Theory of Optimal Experiments*. Acadimic Press, New York, 1972.
5. N. Flournoy, C. May, and C. Tommasi. The effects of adaptation on maximum likelihood inference for non-linear models with normal errors. arXiv:1812.03970, Submitted, 2019.
6. L. Pronzato and A. Pázman. *Design of experiments in nonlinear models*, volume 212 of *Lecture Notes in Statistics*. Springer, New York, 2013. Asymptotic normality, optimality criteria and small-sample properties.