SIS | 2022

51st Scientific Meeting
of the Italian Statistical Society

Caserta, 22-24 June

Università degli Studi della Campania *Luigi Vanvitelli*

Società Italiana di Statistica

www.unicampania.it

Book of the Short Papers

Editors: Antonio Balzanella, Matilde Bini, Carlo Cavicchia, Rosanna Verde

Città di Caserta

1222-2022 800 ANNI Università degli Studi di Padova

Dipartimento di Scienze Statistiche

Stata

sas

Università degli Studi del Sannio Benevento

Pearson

Matilde Bini (Chair of the Program Committee) - *Università Europea di Roma*
Rosanna Verde (Chair of the Local Organizing Committee) - *Università della Campania "Luigi Vanvitelli"*

PROGRAM COMMITTEE

Matilde Bini (Chair), Giovanna Boccuzzo, Antonio Canale, Maurizio Carpita, Carlo Cavicchia, Claudio Conversano, Fabio Crescenzi, Domenico De Stefano, Lara Fontanella, Ornella Giambalvo, Gabriella Grassia - Università degli Studi di Napoli Federico II, Tiziana Laureti, Caterina Liberati, Lucio Masserini, Cira Perna, Pier Francesco Perri, Elena Pirani, Gennaro Punzo, Emanuele Raffinetti, Matteo Ruggiero, Salvatore Strozza, Rosanna Verde, Donatella Vicari.

LOCAL ORGANIZING COMMITTEE

Rosanna Verde (Chair), Antonio Balzanella, Ida Camminatiello, Lelio Campanile, Stefania Capecchi, Andrea Diana, Michele Gallo, Giuseppe Giordano, Ferdinando Grillo, Mauro Iacono, Antonio Irpino, Rosaria Lombardo, Michele Mastroianni, Fabrizio Maturo, Fiammetta Marulli, Paolo Mazzocchi, Marco Menale, Giuseppe Pandolfi, Antonella Rocca, Elvira Romano, Biagio Simonetti.

ORGANIZERS OF SPECIALIZED, SOLICITED, AND GUEST SESSIONS

Arianna Agosto, Raffaele Argiento, Massimo Aria, Rossella Berni, Rosalia Castellano, Marta Catalano, Paola Cerchiello, Francesco Maria Chelli, Enrico Ciavolino, Pier Luigi Conti, Lisa Crosato, Marusca De Castris, Giovanni De Luca, Enrico Di Bella, Daniele Durante, Maria Rosaria Ferrante, Francesca Fortuna, Giuseppe Gabrielli, Stefania Galimberti, Francesca Giambona, Francesca Greselin, Elena Grimaccia, Raffaele Guetto, Rosalba Ignaccolo, Giovanna Jona Lasinio, Eugenio Lippiello, Rosaria Lombardo, Marica Manisera, Daniela Marella, Michelangelo Misuraca, Alessia Naccarato, Alessio Pollice, Giancarlo Ragozini, Giuseppe Luca Romagnoli, Alessandra Righi, Cecilia Tomassini, Arjuna Tuzzi, Simone Vantini, Agnese Vitali, Giorgia Zaccaria.

ADDITIONAL COLLABORATORS TO THE REVIEWING ACTIVITIES

Ilaria Lucrezia Amerise, Ilaria Benedetti, Andrea Bucci, Annalisa Busetta, Francesca Condino, Anthony Cossari, Paolo Carmelo Cozzucoli, Simone Di Zio, Paolo Giudici, Antonio Irpino, Fabrizio Maturo, Elvira Romano, Annalina Sarra, Alessandro Spelta, Manuela Stranges, Pasquale Valentini, Giorgia Zaccaria.

PUBLISHED BY PEARSON

WWW.PEARSON.COM

ISBN 9788891932310

Bias correction of the maximum likelihood estimator for Emax model at the interim analysis

Correzione per distorsione della stima di massima verosimiglianza del modello Emax nell'analisi ad interim

Caterina May and Chiara Tommasi

Abstract The Emax model is a dose-response model commonly applied in clinical trials, agriculture and environmental experiments. We consider a two-stage adaptive design for collecting "optimal" data for estimating the model parameters. At the first stage (interim analysis) a locally D-optimum design is computed to get a sample of independent observations and to produce a first-stage maximum likelihood estimate (MLE). At the second stage, the first-stage MLE is used as initial parameter-value to determine another D-optimum design and then to collect the second-stage observations.
The first-stage estimate influences the quality of the data gathered at the second stage, where a large number of observations can be collected. In real life problems, instead, the sample size of the interim analysis is usually small; therefore, the first-stage MLE should be precise enough even if based on few data. From this consideration, our guess is that if we improved the behaviour of the first-stage MLE through a bias correction, then the D-optimal design determined at the second stage would produce better experimental points. In this study we provide the analytic expression of the first-order bias correction of the MLE in the Emax model.
Abstract *Il modello Emax è un modello di risposta alla dose comunemente usato negli esperimenti clinici, agricoli e ambientali. In questo lavoro consideriamo un disegno adattivo a due stadi per raccogliere dati "ottimali" al fine di stimare i parametri del modello. Al primo stadio (analisi ad interim) viene calcolato un disegno localmente D-ottimo per raccogliere un campione di osservazioni indipendenti e ottenere una stima di massima verosimiglianza (SMV). Al secondo stadio la SMV di primo stadio viene usata come valore inziale del parametro per determinare un secondo disegno D-ottimo e raccogliere le osservazioni di secondo stadio a par-*

Caterina May
Università degli Studi del Piemonte Orientale, Department DiSEI, via Perrone, 18, 28100 Novara (Itay), 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

tire da quest'ultimo disegno sperimentale. Tali osservazioni dipendono quindi dalle risposte precedenti.
La stima di primo stadio influenza la qualità dei dati al secondo stadio. Nei problemi reali la dimensione campionaria dell'analisi ad interim è di solito bassa, quindi la SMV di primo stadio sarrebbe opportuno fosse precisa anche se basata su un campione piccolo. La nostra idea è quindi quella di migliorare la SMV di primo stadio correggendone la distorsione, così che il disegno D-ottimo al secondo stadio sia migliore. In questo lavoro determniamo l'espressione analitica al primo ordine della distorsione della SMV di primo stadio dei parametri del modello Emax.

Key words: maximum likelihood estimator, interim analysis, small sample, bias correction, D-optimality, two-stage adaptive design, Emax model

1 Introduction

The *Emax* model is well-characterized in the literature and it is frequently used for dose-response designs in clinical trials, as well as in agriculture and in environmental experiments (see, for instance, [3] and [2]). It has the form $y = \eta(x,\theta) + \varepsilon$ where y denotes a response at the dose $x \in \mathscr{X} = [a,b]$, $0 \leq a < b$; $\theta = (\theta_0, \theta_1, \theta_2)^T$ is a vector of unknown parameters; ε is a Gaussian random error; the nonlinear mean response is

$$\eta(x,\theta) = \theta_0 + \theta_1 \frac{x}{x+\theta_2}. \tag{1}$$

In equation (1), θ_0 represents the response at the dose zero; θ_1 is the maximum effect attributable to the drug; and θ_2 is the dose which produces the half of the maximum effect.

In this study, to collect observations that provide a precise estimation of θ, we consider a two-stage adaptive design. By the fact, sequential adaptive designs are quite common in clinical trials. More specifically, assume that a guessed value $\tilde{\theta} = (\tilde{\theta}_0, \tilde{\theta}_1, \tilde{\theta}_2)^T$ for θ is available, for instance from an expert opinion. At the first stage (or interim analysis) we take $n_1 < n$ observations according to a locally D-optimal design

$$\xi_1^*(\tilde{\theta}) = \arg\max_{\xi \in \Xi} |M(\xi;\tilde{\theta})|,$$

where $\xi = \begin{Bmatrix} x_{11} & \cdots & x_{1M_1} \\ \omega_{11} & \cdots & \omega_{1M_1} \end{Bmatrix}$ denotes a design, which is defined as a finite discrete probability distribution over $\mathscr{X}$, Ξ is the set of all possible designs and

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

is the information matrix of ξ. Moreover $\nabla\eta(x,\theta)$ denotes the gradient of the mean response $\eta(x,\theta)$ with respect to θ (see for instance, [8] or [1] as references in optimal design of experiments). The design $\xi^*(\tilde{\theta})$ is said locally optimal because it depends on a guessed parameter value $\tilde{\theta}$ due to the non-linearity of $\eta(x,\theta)$. Since a D-optimal design minimizes the generalized asymptotic variance of the MLE for θ, it should improve the precision of the parameter estimates. Let $\hat{\theta}_{n_1}$ be the first-stage MLE based on the n_1 observations gathered during the interim analysis. At the second stage, the available data information can be used to improve the choice of the experimental points. Therefore, $n_2 = n - n_1$ 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 $\tilde{\theta}$.

The final MLE is computed employing the whole sample of $n = n_1 + n_2$ data, which are dependent because the second-stage data depend on the first-stage responses through $\hat{\theta}_{n_1}$. In [9] and in [10] the theoretical properties of this final MLE are described.

The sample size of the interim analysis is usually small and thus $\hat{\theta}_{n_1}$ might be affected by the bias which converges to zero as n increases to infinity. On the other hand, the second-stage D-optimal design depends on $\hat{\theta}_{n_1}$ and at this stage a larger number of observations are collected. If these n_2 responses are observed at bad design points (or in bad proportions), they might produce an unreliable final MLE. On the other hand, if we improve the behaviour of the first-stage MLE through a bias correction, then the D-optimal design determined at the second stage should produce better experimental points. In Section 2 we provide the analytic expression of the first-order bias correction of the MLE in the *Emax* model.

2 Simulations of MLEs efficiencies and first order bias correction

As explained in the introduction, we need to understand how the MLE at the first stage influences the D-optimal design at the second stage. From the analytical expression of the locally D-optimal design ξ_D^* for the *Emax* model (provided by [7]):

$$\xi_D^*(\theta) = \xi_D^*(\theta_2) = \begin{Bmatrix} a & x^*(\theta_2) & b \\ 1/3 & 1/3 & 1/3 \end{Bmatrix}, \tag{3}$$

where the interior support point is

$$x^*(\theta_2) = \frac{b(a+\theta_2) + a(b+\theta_2)}{(a+\theta_2)+(b+\theta_2)}, \tag{4}$$

we have that the second stage D-optimal design depends only on θ_2.

The behaviour of $x^*(\theta_2)$ when $[a,b] = [0,150]$ is plotted in Figure 1. Let θ_2^t denotes the "true" value of θ_2. From Figure 1 we can note that the derivative of $x^*(\theta_2)$ is a positive decreasing function of θ_2 and thus the effect on $x^*(\theta_2)$ is larger for the values $\theta_2 < \theta_2^t$.

The following theorem provides the expression for the bias of the first stage MLE of θ_2, which is herein denoted by $\hat{\theta}_{n_1,2}$.

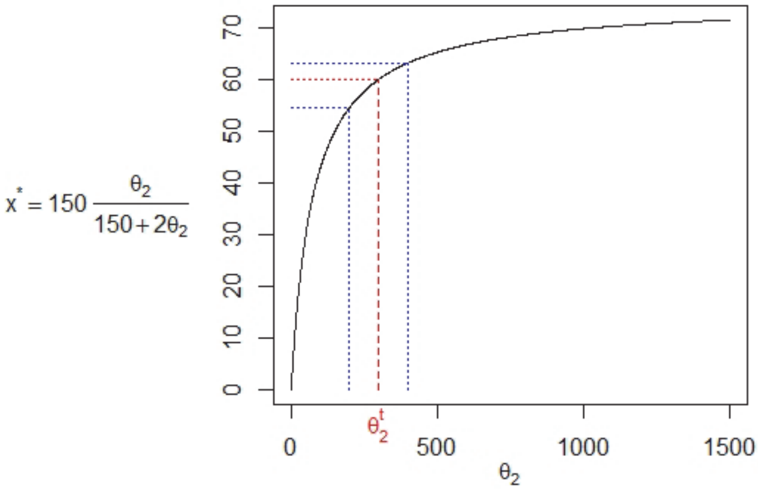

Fig. 1 D-optimum middle dose $x^*(\theta_2)$ for the Emax model

Proposition 1. *Let $\theta_{2,0}$ be a nominal value for θ_2. If n_1 first stage observations are taken according to the local D-optimal design (3), with equal numbers treated at the experimental points a, $x^*(\theta_{2,0})$ and b, then the bias of the first stage MLE of θ_2 is*

$$E(\hat{\theta}_{n_1,2} - \theta_2) = \frac{b_2(\theta)}{n_1} + O(n_1^{-2}),$$

where $b_2(\theta) > 0$ is given by

$$\begin{aligned} b_2(\theta) = & \frac{1}{(a-b)^4\theta_1^2\theta_2^2(a+\theta_{2,0})^2(b+\theta_{2,0})^2} \\ & \cdot\{3\sigma^2(a+\theta_2)^2(b+\theta_2)^2[2ab+(a+b)\theta_{2,0}+\theta_2(a+b+2\theta_{2,0})]^2 \\ & [3ab(a+b)+(a^2+10ab+b^2)\theta_{2,0}+3(a+b)\theta_{2,0}^2 \\ & +2\theta_2(a^2+ab+b^2+3(a+b)\theta_{2,0}+3\theta_{2,0}^2)]\}. \end{aligned} \tag{5}$$

Proof. Cox and Snell (1968) introduced the $O(n^{-1})$ formula for the bias of the MLE in the case of n observations not being identically distributed. Cordeiro and Klein (1994) proposed a matrix expression for this bias, which is herein specialized for the Emax model and the D-optimal design $\xi_D^*(\theta_{2,0})$. Calculations are available by the authors upon request.

This result justifies the fact that, when $\theta_{2,0} < \theta_2^t$ the fixed procedure (that consists in collecting all the n observations according to the initial (first-stage) D-optimal design) seems to have a worst performance than a two-stage design; in facts, $\hat{\theta}_{n_1,2}$

has a positive bias and thus it takes (on average) larger values, and thus we expect that $x^*(\hat{\theta}_{n_1,2})$ is closer to $x^*(\theta_2^t)$ than $x^*(\theta_{2,0})$ does.

3 Conclusions

In this paper we have presented the idea of improving the two-stage adaptive design proposed in [9] by introducing a bias correction. An analytic form of the first-order bias correction under the Emax model has been provided. We also justify the non-symmetric performance of the adaptive procedure in comparison with a fixed one, that results from the simulations in [9].

As a future work, we aim to apply the bias correction to the two-stage procedure in order to investigate its possible improvement. We aim at exploring also other types of bias corrections, in particular the ones obtained by modifying the score function (see, for instance, [11]).

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. M. Baker, J. L. Hobman, C. Dodd, S. J. Ramsden, D. J. Stekel. Mathematical modelling of antimicrobial resistance in agricultural waste highlights importance of gene transfer rate. *FEMS Microbiology Ecology*, 92(4), 2016.
3. F. Bretz, H. Dette, and J. Pinheiro. Practical considerations for optimal designs in clinical dose finding studies. *Statistics in medicine*, 29(7-8), 731–742, 2010.
4. B. Bornkamp, J. Pinheiro, and F. Bretz. *DoseFinding: Planning and Analyzing Dose Finding Experiments*, 2018. R package version 0.9-16.
5. D.R. Cox and E.J. Snell. A general definition of residuals. *Journal of the Royal Statistical Society (B)*, 30 (2):248-275, 1968.
6. G.M. Cordairo and R. Klain. A general definition of residuals. *Statistics and Probability Letters*, 19:169-176, 1994.
7. 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.
8. V. Fedorov. *Theory of Optimal Experiments*. Academic Press, New York, 1972.
9. N. Flournoy, C. May, and C. Tommasi. The effects of adaptation on maximum likelihood inference for nonlinear models with normal errors. *Journal of Statistical Planning and Inference*, 214:139-150, 2021.
10. C. May, and C. Tommasi. On the behaviour of the maximum likelihood estimator for exponential models under a fixed and a two-stage design. *Book of Short Papers SIS 2020*, Pearson, 2020.
11. 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.