Bayesian T-optimal designs by simulation: a case-study on model discrimination

Rossella Berni and Federico M. Stefanini

Abstract In a case study on wine making, total anthocyanins are measured during wine pre-fermentation. An inhomogeneous Markov Chain is developed to obtain the Bayesian T-optimal design for the next year. Results are discussed in view of extensions of the utility function often needed in actual applications.

Key words: Bayesian T-optimal designs, utility function, Monte Carlo simulation

1 Introduction

Optimal design criteria received growing attention in the last ten years, both at theoretical and computational levels, in part following the increase of computational power. Since the 70s, there is a long history of seminal papers in literature on D and T-optimality, both to estimate model parameters and to discriminate among models (for example, [7],[2], [3], [1]). Each experimental point and the final optimal design are selected according to the General Equivalence Theorem (G.E.T.). Model dependency may be considered as the main disadvantage for an optimal design, thus the result depends on the hypothesized statistical model and its parameters: in presence of uncertainty on model and parameters this dependence is crucial.

More recently, this dependence was considered in a Bayesian framework [5], by introducing prior distributions on models and parameters. Later [6] extended T-optimality by adopting the Kullback-Leibler distance to address the heteroschedasticity and the non-Gaussian nature of response variables: they defined the KL-optimality.

Notwithstanding the generality achieved, in actual applications further flexibility is often needed, for example by defining a utility function in which the cost of each observation depends on the value taken by the independent variable.

Stefanini F.M. e-mail: stefanini@ds.unifi.it · Berni R. e-mail: berni@ds.unifi.it
Department of Statistics "G.Parenti", University of Florence

In this paper, a utility function based on the T-optimality criterion is defined and an inhomogeneous Markov Chain algorithm is developed [4] to perform Monte Carlo optimization.

2 Basic Theory

The class of linear and non-linear parametric models are considered as:

$$E(Y_i|x_i,\theta_j)=\eta_j(x_i,\theta_j);j=1,2;x_i\in\mathscr{X},\theta_j\in\Theta_j \tag{1}$$

with $j=1,2$ the index of the considered model. Vector $x_i=(x_{i1},..,x_{ij},..,x_{ik})$ denotes the set of k independent variables for the i-th observation $(i=1,..,n)$; for simplicity we assume there are no replicates; the two vectors of unknown parameters $\theta_j, j=1,2$ have size m_j, $\theta_j\in\Theta_j\subset\mathscr{R}^{m_j}$. We assume that $\eta_j(x_i,\theta_j)$ are continuous real functions of $(x_i,\theta_j)\in\mathscr{X}\times\Theta_j$. In an experimental design setting each x_i is the i-th trial chosen or/and controlled by the experimenter; x_i belongs to the compact set $\mathscr{X}$ which is the experimental region defined in the $\mathscr{R}^k$ field. Regarding random errors $\varepsilon_{i,j}$, we assume that $\varepsilon_{i,j}\overset{i.i.d.}{\sim}N(0,\sigma_j^2)$.

In this context we consider a discrete or exact design, i.e. a design formed by a set of n points $(x_1,..,x_i,..,x_n)$ in $\mathscr{X}$ and denoted by D_n. Furthermore, ξ is the design measure defined as a probability measure on the compact set $\mathscr{X}$ and it satisfies $\int_{\mathscr{X}}\xi(x)\,dx=1$. It must be noted that a continuous design is a design which depends only on the assumed probability measure ξ, without considering the number of experimental points. A discrete design is defined for a specific set of trials and for each design point of D_n given the total number of observations n by assigning masses p_i to each x_i: $\xi_n=\{(p_i,x_i);i=1,..,n\};\sum_i p_i=1$ where ρ is the rounded number of observations taken at x_i, with $\sum_i\rho(p_in)=n$. In [5], the theory is extended to situations in which model uncertainty is present, and it is described by an elicited prior distribution with parameters $\pi_{0j};j=1,2$. Moreover elicited prior distributions of model parameters are $p(\theta_j),j=1,2$. The central result is that the Bayesian T-optimal criterion satisfies the G.E.T. theorem while it no longer depends on unknown values of

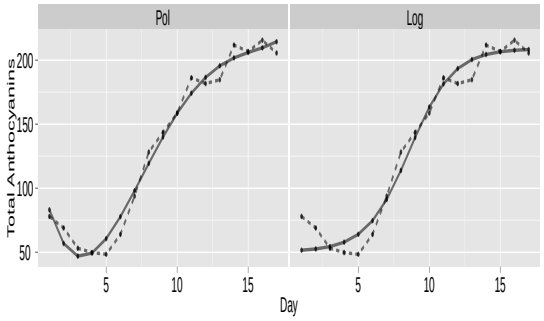

Fig. 1 Maximum likelihood estimates of polynomial (left, continuous line) and logistic (right, continuous line) models. Dashed lines join observed values.

the true model parameters. The two noncentrality parameters are:

$$\Delta_j(\xi,\theta_{\bar{j}}) = \inf_{\theta_j\in\Theta_j}\int_{\mathscr{X}}(\eta_{\bar{j}}(x,\theta_{\bar{j}}) - \eta_j(x,\theta_j))^2\xi(dx) \tag{2}$$

with $j\in\{1,2\}$ and $\bar{j}$ its complement in $\{1,2\}$, thus j is in turn the index of the true model, after [2]. A T-optimal design $\xi^\star$ maximizes the criterion:

$$\Gamma(\xi) = \sum_{j\in\{1,2\}}\pi_{0j}\int_{\Theta_j}\Delta_{\bar{j}}(x,\theta_j)p(\theta_j)d\theta_j \tag{3}$$

3 A Monte Carlo algorithm for wine making

The kinetic of total anthocyanins (TAs) during a vinification procedure which is quite popular in Tuscany is considered. Year 2010 data were collected by daily sampling during maceration just after a pumping over. One hundred *ml* were withdrawn from the sampling valve and the UV/VIS spectra of the supernatants were recorded after centrifugation. Two models were considered (Figure 1): a polynomial (η_1) which is motivated by the possible role played by sulfitation, a logistic curve (η_2) which is based on chemical considerations about the presence of TAs in solution. The expected values are:

$$\eta_1(x,\theta_1) = \theta_{1,0}+\theta_{1,1}x+\theta_{1,2}x^2+\theta_{1,3}x^3+\theta_{1,4}x^4 \tag{4}$$

$$\eta_2(x,\theta_2) = \theta_{2,2}+((\theta_{2,1}-\theta_{2,2})/(1+exp((\theta_{2,3}-x)/\theta_{2,4}))) \tag{5}$$

thus discrimination is performed between two non-linear functions in coefficients and/or factors with only one independent variable ($k=1$).

The prior probability values on models are $\pi_{0,1}=0.1$ and $\pi_{0,2}=0.9$. Discretized posterior distributions on a grid of 25 points given observed data d were derived after calculating Laplace approximations under weakly informative priors, respectively $p(\theta_1\mid d)$ and $p(\theta_2\mid d)$. Following [4], we defined the inhomogeneous Markov Chain to optimize the utility function:

$$u(\xi,\theta_1,\theta_2) = 1\cdot 10^{-10}+\pi_{01}\Delta_2(\xi,\theta_1)+\pi_{02}\Delta_1(\xi,\theta_2) \tag{6}$$

with a multivariate normal jump distribution $g(\tilde{\xi}\mid\xi)$ defined on suitably transformed coordinates and weights $(h_x(x_i),h_p(p_i)), i=1,\ldots,5$. Given the current state (ξ,v) of this chain, the steps of our algorithm are:

1. Generate a candidate design $\tilde{\xi}$ given the current state using g.
2. Generate J points $(\theta_1^{(j)},\theta_2^{(j)})_j$ from the distributions of model parameters $p(\theta_1\mid d)$ and $p(\theta_2\mid d)$. Calculate

$$\tilde{v} = J^{-1}\sum_{j=1}^{J} log\left(u(\tilde{\xi},\theta_1^{(j)},\theta_2^{(j)})\right)$$

and set the candidate $(\tilde{\xi}, \tilde{v})$.
3. Calculate $\alpha_J = min(1, exp(J\tilde{v} - Jv)$, and with probability α_J accept the candidate $(\tilde{\xi}, \tilde{v})$, otherwise let v unchanged.
4. Increase the current value of J according to a suitable cooling schedule $\{J_t : t = 1, 2, \ldots\}$ with $J_{t+1} \geq J_t$, e.g. every fifty steps increase J by 25.
5. Go to step (1) up to convergence.

Among the designs made by six distinct points, we found the optimum at,(x_i, p_i): $(1.00, 0.48), (3.40, 0.35), (3.26, 0.11), (8.87, 0.02), (11.64, 0.02), (16.06, 0.02)$.

4 Discussion

The proposed algorithm is based on a utility function which is remarkably close to what optimized in the classic framework (up to a small constant to make it positive). Nevertheless the implementation does not change much if equation (6) is changed to incorporate the cost of observations, that may eventually depend on the value taken by x. Similarly, the total number of observations could be explicitly accounted for.

From the standpoint of the quality of calculations, approximated prior distributions were considered to better compare results of simulations to the solution provided by strictly following [5]. In [4] it is presented a more general algorithm to simultaneously obtain the posterior distribution and the optimized decision. In the literature, it has been already hypothesized that the prior distribution on models could be updated given the observed data, as we performed for model parameters, even though a substantial increase of computational burden is expected after such extension.

References

1. Atkinson, A.C., Cox, D.R.: Planning experiments for discriminating between models (with discussion). J. R. Stat. Soc. ser. B, **36**, 321–348 (1974)
2. Atkinson, A.C., Fedorov, V.V.: The design of experiments for discriminating two rival models. Biometrika, **62**, 57–70 (1975)
3. Atkinson, A.C., Fedorov, V.V.: The design of experiments for discriminating between several models. Biometrika, **62**, 289–303 (1975)
4. Müller, P., Sansó, B., De Iorio, M.: Optimal Bayesian design by inhomogeneous Markov chain simulation. J. Am. Stat. Assoc., **99**, 788–798 (2004)
5. Ponce De León, A.C., Atkinson, A.C.: Optimum experimental design for discriminating between two rival models in the presence of prior information. Biometrika, **78**, 601–618 (1991)
6. Tommasi, C., Lopez-Fildago, J.: Bayesian optimum designs for discriminating between models with any distribution. Comput. Stat. Data An., **54**, 143–150 (2010)
7. Wynn, H.P.: The sequential generation of D-optimum experimental designs. Ann. Math. Stat., **41**, 1655–1664 (1970)