Chemometrics and Intelligent Laboratory Systems 95 (2009) 199–208

Contents lists available at ScienceDirect

Chemometrics and Intelligent Laboratory Systems

journal homepage: www.elsevier.com/locate/chemolab

Study of the best designs for modifications of the Arrhenius equation

Juan M. Rodríguez-Díaz *, M. Teresa Santos-Martín

Department of Statistics, Faculty of Science, University of Salamanca Pl. de los Caídos s/n, 37.008 Salamanca, Spain

ARTICLE INFO

Article history:
Received 26 May 2008
Received in revised form 23 October 2008
Accepted 24 October 2008
Available online 14 November 2008

Keywords and phrases:
Arrhenius equation
c-optimality
D-optimality
Exponential model
Maximin design
Optimal design

ABSTRACT

The Arrhenius equation is widely used to describe the relationship between the rate of a chemical reaction and the temperature. However, in some cases more precision is needed and a Modified Arrhenius (MA) model, allowing the linear parameter to be temperature-dependent, appears as the correct alternative to the plain model. Optimal designs for the Arrhenius equation have been already computed, for instance in Rodríguez-Aragón and López Fidalgo [L.J. Rodríguez-Aragón and J. López-Fidalgo (2005). Optimal designs for the Arrhenius equation. Chemometr Intell Lab Syst 77 131–138.] for independent and normally-distributed errors with constant variance and in Rodríguez-Torreblanca and Rodríguez-Díaz [C. Rodríguez-Torreblanca and J.M. Rodríguez Díaz (2007). Locally *D*- and *c*-optimal designs for Poisson and Negative Binomial regression models. Metrika 66 161–172.] for different variance structures. However, the MA model has not been studied at the same level. In this work, optimal designs for this last equation will be computed for a general design space and different optimality criteria, and their performance will be shown through convenient examples. A robustness analysis when a wrong choice of the initial values for the parameters is made or some of the hypothesis on the model are not fulfilled will be performed, in order to be able to choose the best design for each situation.

1. Introduction

From the middle of the 19th century on, several experimentally-developed equations relating the rate constant k of a chemical reaction to the temperature t have appeared in the literature. The fact that different plots gave reasonably good linear fits with the same data was due to the narrow temperature ranges usually employed in kinetic studies [22]. Eventually all these equations were dropped, mainly due to the lack of theoretical justification, except for two of them: the Arrhenius and Modified-Arrhenius ones, that were well explained theoretically.

The next subsections focus on the models and the statistical setup and notation. The main results are given in Section 2, where optimal designs are computed and compared with some others used in the literature for the Arrhenius model. Section 3 analyzes the goodness of the designs computed when the optimality conditions are not fulfilled. Lastly, there is a discussion on the model and outlines of future work.

1.1. Arrhenius models

The Arrhenius equation was first used by Svante Arrhenius in his studies of the dissociation of electrolytes [4], but nowadays is widely accepted as the right tool to describe the influence of temperature on the rates of chemical processes, as well as many other physical processes such as diffusion, thermal and electrical conductivity or viscosity.

The integrated form of the Arrhenius equation is $\ln(k)=A'-\beta/t$ where $\beta=C/R$, with C being the *activation energy* and R the *gas constant*. By taking exponentials the expected value of k can be expressed as $E[k]=A\exp(-\beta/t)$ where $A=\exp(A')>0$ is the *frequency factor*, $\beta>0$ and t is given in °K. By making the substitution $t=1/x$ it becomes the exponential model, and thus the optimal designs have been studied for instance in [16] for independent and normally-distributed errors with constant variance or in [31] for different variance structures. Optimal and compound designs specifically for the Arrhenius equation as well as a study of the efficiency of some designs used in the literature can be found in [30].

However, for the analysis of more precise rate-temperature data, particularly in studies covering a wide temperature range, it is usual to allow A' to be temperature-dependent, proportional to $-ln(t)$, or equivalently A proportional to $1/t$ raised to a power m, producing

$$E[k] = \frac{a}{t^m} e^{-\beta/t}, \tag{1}$$

where $a>0$ is now independent of the temperature and $\beta>0$; this is the so called Modified-Arrhenius (MA) model. Usually (see Laidler, 1984), the procedure employed is to use the plain Arrhenius model for data of lower precision or where the temperature range is limited, and to analyze more precise data by using model (1).

As an example, consider the [20]. The NASA Panel for Data Evaluation publishes a series of evaluated sets of rate constants. The data are mainly used to model stratospheric and upper tropospheric processes, with

* Corresponding author.
E-mail addresses: juanmrod@usal.es (J.M. Rodríguez-Díaz), maysam@usal.es (M.T. Santos-Martín).

0169-7439/$ – see front matter © 2008 Elsevier B.V. All rights reserved.
doi:10.1016/j.chemolab.2008.10.011

particular emphasis on the ozone layer and its possible perturbation by anthropogenic and natural phenomena. In particular, the reaction of the hydroxyl radical OH with methane has been extensively studied. The temperature dependence of the rate coefficient of this reaction had been described mainly through the Arrhenius equation using different techniques for estimation and always for temperatures over 233°*K*. But when [14] extended the measurements of *k* to 195°*K* it appeared that the rate coefficient did not follow a strictly Arrhenius expression. They obtained a more accurate representation of the rate constant as a function of temperature by using the equation $k=2.80\times10^{-14}\,t^{0.667}\exp(.1575/t)$. This expression may be preferred for lower stratosphere and upper troposphere calculations, and it is confirmed by [5].

The MA model (as the two-parameter model $E[Y(x)]=\theta_1 x^{\nu}\exp(-\theta_2 x)$ via the change of variable $(x=1/t)$ has been studied in [10] from a Bayesian point of view. There *m* was always taken to be positive, the design space fixed to be $[0,\infty)$ and the only criterion used was *D*-optimality. Here we will allow $m<0$ and restrictions on the design space, and (locally) optimal designs will be computed with respect to different optimality criteria, as well as restricted maximin designs. Heteroscedastic cases have been also studied and they are discussed in the final section. Throughout the paper, *m* will be assumed to be known.

1.2. Optimal design of experiments

The aim is to find the values $\{t_1,t_2,\ldots\}$ at which to take samples in order to get the best estimators of the parameters of model (1), that is, the estimators with minimum variance. In other words, we look for *Optimal Designs* for this model. [6] introduced locally optimal designs for nonlinear models, the models coming from chemical kinetics. When working with non-linear models like the MA, the best design will depend on the values of the unknown parameters. To address this problem, some kind of additional information is needed, either an initial value for the parameters or a prior distribution for them. In any case, the optimal design will be a function of these initial values or distributions (locally or bayesian optimal designs). Specifically, we can write a non-linear model as $Y=\eta(t,\theta)+\epsilon$, where *Y* denotes the observation, $t\in\mathcal{T}=[t_{\min}, t_{\max}]$ is the independent variable $(t_{\min}>0)$, $\theta=(\theta_1,\theta_2,\ldots\theta_s)^T\in\Theta$ is the vector of unknown parameters, and $\eta(t,\theta)$ is a function that is non-linear respect to θ. The ε term stands for the random errors, that will be assumed to be independent and normally distributed with zero mean and variance σ^2.

A design is a collection of points of the independent variable, $\{t_1,t_2,\ldots,t_N\}$, where *N* is the size of the design. It can be written by taking the *n* distinct points (called the support points) and for each one the proportion (weight) P_i that it has in the design, and from this point of view, an *approximate design* can be defined as any probability measure in *T* with finite support.

The information matrix becomes the main tool when looking for the optimal design for the experiment. When the function $\eta(t,\theta)$ is differentiable with continuous derivative for every parameter θ_i, the information matrix for a design ξ can be written as $M(\xi,\theta)=\sum_{i=1}^{n} p_i\nabla\eta(t_i,\theta)\nabla\eta(t_i,\theta)^T$, where $\nabla\eta(t,\theta)$ is the gradient vector of $\eta(t,\theta)$. For model (1), since *m* is assumed to be known $\theta=(a,\beta)^T$, thus

$$\nabla\eta(t,\theta)=\left(t^{-m}\exp(-\beta/t),\,-at^{-m-1}\exp(-\beta/t)\right)^T \tag{2}$$

and the information matrix for the MA model is given by

$$M(\xi,\theta)=\sum_{i=1}^{n} p_i\begin{pmatrix} t^{-2m}e^{-\frac{2\beta}{t}} & -at^{-2m-1}e^{-\frac{2\beta}{t}} \\ -at^{-2m-1}e^{-\frac{2\beta}{t}} & a^2t^{-2m-2}e^{-\frac{2\beta}{t}}\end{pmatrix}.$$

The inverse of the information matrix is (asymptotically) proportional to the covariance matrix of the parameter estimators of the model. For this reason, optimal experimental designs typically minimize some convex function of the inverse of the information matrix. The most used criterion is *D*-optimality, that focuses in the determinant of the information matrix. A design ξ is *D*-optimal if it maximizes this determinant, which is equivalent to minimizing that of the covariance matrix. A *D*-optimal design minimizes the volume of the confidence ellipsoid of the parameters.

The General Equivalence Theorem is a useful tool for checking whether a design is optimal, which happens when the directional derivative is greater than 0 in every direction (design). In the case of *D*-optimality, it can be stated as follows: if *s* is the number of parameters of the model and $\theta=\theta^0$ is the vector of initial values of the unknown parameters, then the design ξ^* is locally *D*-optimal if and only if

$$g_{\xi^*}(t)=s-\nabla\eta\left(t,\theta^0\right)^T M^{-1}\left(\xi^*,\theta^0\right)\nabla\eta\left(t,\theta^0\right)\geq 0\quad\forall t\in T \tag{3}$$

and the inequality becomes equality at the support points of ξ^*.

If we are interested in looking for the best estimator for a linear combination of parameters, then *c*-optimality is the criterion of choice. It is specially used for vectors (1,0,…,0),…,(0,…,0,1); when considering these individually, optimal designs for the estimation of each parameter can be obtained. For $c\in R^k$, a design ξ is c-optimal if it maximizes $-c^TM(\xi,\theta)^{-1}c$, which is equivalent to minimizing $Var(c^T\theta)$, and the General Equivalence Theorem can be used to check whether a design is *c*-optimal. However, *c*-optimal designs can be constructed geometrically using Elfving's Theorem [11]; the procedure will be detailed in Section 2.3. A discussion on support points and weights for *c*-optimality for some non-linear two-parameter models can be found in [13]. A general overview on optimal designs can be found for example in the books of [27] or the recent of [3].

2. Computations

Through this section, general formulas for computing *D*- and *c*-optimal designs for the MA model will be provided, and optimal temperatures will be found for explicit values of *m* and the parameter β. An example checking optimality of the proposed designs and a study comparing the efficiency of some other designs found in the literature for the Arrhenius model will illustrate the behaviour of the computed designs. Since *m* is known, the parameters to be estimated are *a* and β. It is not difficult to check that the optimal designs do not depend on the error variance σ^2, thus $\sigma^2=1$ will be assumed without loss of generality. And it can be seen that the (local) *D*-, c_1- and c_2-optimal designs (with $c_1=(1,0)^T$, $c_2=(1,0)^T$) do not depend on the initial value of the linear parameter *a*. Thus from now on we will assume $a=1$.

2.1. D-optimal designs

Since we have two parameters we need at least two points in the design. Let us begin by considering a two-point design, say supported in $\{t_1,t_2\}$ with $t_{\min}\leq t_1<t_2\leq t_{\max}$. Given that the number of design points is the number of parameters of the model, both points should have the same weight for the design to be *D*-optimal ([32], Lemma 5.1.3), and the determinant of the information matrix is (proportional) to

$$Det(t_1,t_2)=e^{-2\beta\left(\frac{1}{t_1}+\frac{1}{t_2}\right)}t_1^{-2(m+1)}(t_1-t_2)^2t_2^{-2(m+1)}. \tag{4}$$

The points $\{t_1,t_2\}$ maximizing Eq. (4) for different conditions on T and *m* are given in Table 1, where the weight of every point is always 1/2 and cases (a) and (b) assume $m\geq 0$. From now on, *D*-optimal designs with two support points will be given just by those points, $\{t_1,t_2\}$, since the optimal weights will be 1/2.