Chemometrics and Intelligent Laboratory Systems 154 (2016) 150–161

Contents lists available at ScienceDirect

Chemometrics and Intelligent Laboratory Systems

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

Review

D-optimal experimental designs for a mixture of Adair models

J. López-Fidalgo[1], M.M. Rodríguez-Hernández*

University of Castilla-La Mancha, Department of Mathematics, Institute of Mathematics applied to Science and Engineering, Avda. Camilo José Cela 3, 13071 Ciudad Real, Spain

ARTICLE INFO

Article history:
Received 4 March 2015
Received in revised form 9 March 2016
Accepted 12 March 2016
Available online 28 March 2016

Keywords:
Adair model
Covariance matrix
D-optimality
EM algorithm
Information matrix
Mixture of distributions

ABSTRACT

In a Pharmacokinetic reaction a ligand may be bound to different types of macro-molecules, each with different number of binding sites. They are frequently involved in certain diseases diagnostics. Adair equation is used very often to model the reactions of biological macro-molecules with a ligand. This equation relates the saturation ratio with the free ligand concentration when the equilibrium is reached and it depends on some association constants of the chemical reaction, which have to be estimated.

The main problem considered in this paper is the computation of optimal experimental designs for a mixture of Adair models when different types of macromolecules are mixed in the experiment. The main contribution of this work is obtaining the Fisher Information Matrix for a model with a mixture of probability distributions. Since this is not anymore in the Exponential family the expectation cannot be obtained analytically. Then the computation of optimal designs through the information matrix cannot be done with traditional methods. In data analysis when one has the data this expectation can be computed empirically from the data. But in experimental design data are not available when the experiment is being scheduled. Assuming nominal values of the parameters, as is usually done for nonlinear models, simulations were performed for each point in a suitable discretized design space. The number of simulations and the sample size used in each simulation were empirically tuned for both.

A sensitivity analysis was performed for different possible true values of the parameters. Since this meant an important computational burden fractional designs were used to cover a reasonable neighborhood of the nominal values of the parameters. In order to display the results in a friendly way for the practitioner, "safe" neighborhoods of the optimal designs are provided.

© 2016 Elsevier B.V. All rights reserved.

Contents

* Corresponding author. Tel.: +34 926 295300.
E-mail addresses: jesus.lopezfidalgo@uclm.es (J. López-Fidalgo), mmercedes.rodriguez@uclm.es (M. Rodríguez-Hernández).
[1] Tel.: +34 926 295212; fax: +34 926 295361.

http://dx.doi.org/10.1016/j.chemolab.2016.03.021
0169-7439/© 2016 Elsevier B.V. All rights reserved.

1. Introduction

Adair's equation is used to model reactions of biological macromolecules with a ligand. This equation, which depends on the association constants of the chemical reaction, relates the saturation rate and the concentration of free ligand when the chemical equilibrium is reached. Monomers and dimers are chemical compounds of much interest in the fields of biochemistry and medicine, where they are involved in the diagnosis of certain diseases. Frequently one ligand binds to different types of macromolecules. A well known example of a mixture of two macro-molecules with different binding sites is the binding of oxygen to myoglobin (monomer) and hemoglobin (tetramer). This paper provides optimal experimental designs for a mixture of two Adair models depending on the free ligand. Since the free ligand is a random variable that is not under the control of the experimenter it is necessary to find a relationship between the free ligand and another variable that will be controlled experimentally. This variable is the initial ligand set at the beginning of the experiment. Then, the initial ligand is chosen in such a way the final free ligand will be optimal from the point of view of the experimental design theory.

The Adair model is used at the moment the chemical reaction of the macro-molecules and ligands reaches the equilibrium. These reactions are usually reversible. Adair [1] proposed a model of how oxygen molecules bind to the binding sites of hemoglobin. For more details see [12].

In this paper we consider an experiment consisting of a container with an initial ligand concentration, $[L_0]$, in it. Then a semipermeable dialysis bag with concentrations of two types of macro-molecules (proteins) with N binding sites, $[M_N^{(0)}]$ and $[M_N^{(1)}]$ is introduced into the container. They could be two types of macro-molecules, two types of proteins or a mixture of a macro-molecule and a protein. Hereafter we will refer to macro-molecules for simplicity. Fig. 1 (left) shows the initial situation of the two macro-molecules just before unions occur. The initial ligand molecules begin to enter the dialysis bag, but the two types of macro-molecules cannot leave the bag. The equilibrium is reached when there is the same unbound ligand concentration inside and outside the bag, called free ligand, $[L]$. Fig. 1 (right) displays the situation, in a naïf way, for both types of macro-molecules when the equilibrium is reached. From top to bottom we see both types of bound macro-molecules, free ligand and the two types of unbound macro-molecules.

In general, if the macro-molecules have N binding sites the ligand can bind them through $0, 1, 2, \ldots, N$ sites. The concentration of the two types, $s = 0, 1$, of macro-molecules bound to the ligand will be denoted by $[M_N^{(s)i}] i = 1, 2, \ldots, N$ while the unbound macro-molecules will be denoted by $[M_N^{(s)0}]$.

The models for $s = 0, 1$ can be formulated as:

$$y^{(s)} = \frac{K_1^{(s)}[L] + 2K_1^{(s)}K_2^{(s)}[L]^2 + \ldots + NK_1^{(s)}\ldots K_N^{(s)}[L]^N}{N\left(1 + K_1^{(s)}[L] + K_1^{(s)}K_2^{(s)}[L]^2 + \ldots + K_1^{(s)}\ldots K_N^{(s)}[L]^N\right)} + \varepsilon. \quad (1)$$

For more details on the procedure see Section 2 of [9].

In order to estimate the parameters of the model different values of the initial ligand concentration are tried in different experiments. When the equilibrium is reached, the ligand that is out of the bag is measured, $[L]$. Inside the bag, the saturation of each of the two types of macro-molecules, y_i, is also measured from a sample.

The aim of this paper is to compute optimal experimental designs for the saturation ratio for a mixture of two types of macro-molecules in the following cases:

- one binding site (two monomers),
- one and two binding sites (one monomer and one dimer),
- two binding sites (two dimers).

An optimal design will provide both, good estimates and savings in the use of experimental resources [5]. The theory of optimal experimental design provides tools for computing good designs from different points of view. A model relating uncorrelated observations (responses), $y_1, \ldots, y_n$ with some explanatory variables with the corresponding values $[L]_1, \ldots, [L]_n$ that are under the control of the experimenter is stated by giving a probability density function (pdf),

$$f(y_i, [L]_i; \theta),\ i = 1, \ldots, n,$$

where θ is the vector of parameters to be estimated. The collection of values $[L]_1, \ldots, [L]_n$ from a design space χ are usually called *exact experimental design of size n*. Some of these values may be repeated, e.g. there are k different values, say $[L]_1, \ldots, [L]_k$ without loss of generality. This suggests the definition of a finite discrete probability distribution assigning to each point a weight proportional to the number of replicates. Using this idea Kiefer [8] introduced the concept of *approximate* design, ξ, as any probability measure defined on a design space χ, where the design points may be chosen. The Fisher Information Matrix (FIM) associated to an approximate design ξ is,

$$M(\xi; \theta) = \int_\chi I([L]; \theta)\xi(d[L]),\ I([L]; \theta) = \mathrm{E}\left(\frac{-\partial^2 L(\theta)}{\partial \theta^2}\right), \quad (2)$$

where $L(\theta)$ is the log-likelihood function. For non-linear models this matrix depends on the parameters.

Caratheodory's theorem allows restricting to finite designs,

$$\xi = \begin{Bmatrix} [L]_1 & \ldots & [L]_k \\ p_1 & \ldots & p_k \end{Bmatrix},$$

where $\xi([L]_i) = p_i$ define the mass probability function. A realizable design has to be "approximated" as $n_i \approx n \times p_i$ replicates for the i-th experiment.

D-optimality is one of the most popular optimality criteria. A D-optimal design maximizes the determinant of the Information Matrix, $\det[M(\xi; \theta)]$. The general equivalence theorem establishes that a design, ξ^*, is D-optimal if

$$tr\left(M^{-1}(\xi^*, \theta)I([L]; \theta)\right) \le m, \quad [L] \in \chi, \quad (3)$$

where m is the number of parameters. The left term is usually referred as the sensitivity function. Moreover, the equality is reached at the design points.

For more details on optimal design see Section 2 of [9] and the references provided there.

The main issue in this paper is that a mixture of distributions is not in the Exponential family. Thus, it is not possible to find an analytic expression for the FIM since the expectation in Eq. (2) cannot be computed analytically. It needs to be approximated in some way. In this paper it has been approximated through simulations at different points of the design space, once some nominal values of the parameters are established. As is well known (see e.g [3], [11] or [10]), computing Maximimum Likelihood Estimates (MLE) directly for models with mixture of distributions becomes rather tedious for many families of distributions. This problem may be solved by using the EM algorithm for missing data.

Download English Version:

https://daneshyari.com/en/article/1180320

Download Persian Version:

https://daneshyari.com/article/1180320

Daneshyari.com