

Chapter 2

Motivation

Abstract The appropriate notation is introduced linking the Mathematical problem with the Experimentation. The linearization of the underlying model and Fisher's information is discussed and examples are discussed.

2.1 Introduction

Experimenters working in Laboratories, without being necessarily Mathematicians, know that performing any experiment, in principle the following are needed:

- The experimental unit, appropriate for each particular field.
- The range of experimental conditions, known as design space.
- The measurements or responses (y) obtained at certain values of the explanatory or input variables (u).

Usually, in Chemical problems, the experimenter designs in blocks, and the experimental unit is the apparatus which provides the measurements. In Psychology or Medicine the experimental units are the individual (usually particular animals) under investigation, see Example 1. Thus the experiment is performed through single binary observations. In Medicine or Toxicology the binary response models are more often adopted, linking the “discrete” data, see Examples 4, 7 and 10 below. In Chemistry Kinetics, see [Sect. 4.7](#) below, the “continuous” nonlinear case is usually applied; see also Examples 8 and 13.

The Book of Science has been written with the Mathematical alphabet—do not be afraid of this globalization! It can be easily proved even for the experimentalist. Thus, in this Chapter the notation is introduced and the necessary assumptions and definitions for the mathematical formulation of the nonlinear experimental design problem. The appropriate discussion helps, especially the non-mathematicians, to understand the problem at hand.

2.2 Notation

The Euclidian space $U \subseteq \mathbb{R}^k$ in which the predictor variables (covariates or explanatory variables or independent variables, or input variables) $u = (u_1, u_2, \dots, u_k)$ take their values is known as *experimental region* or *design space*. An example from chemical kinetics is “time” and from medicine “dose”. The *parameter space* $\Theta \subseteq \mathbb{R}^p$ is the set where the involved parameters $\theta = (\theta_1, \theta_2, \dots, \theta_p)^T$ take their values. Let Ξ be the family of measures ξ such that

$$\xi(u) \geq 0, u \in U \text{ and } \int \xi(du) = 1. \quad (2.2.1)$$

Such a measure ξ is referred as a *design measure*, while the pair (U, ξ) will be called *the design*. The *support* of the design (U, ξ) , $\text{Supp}(\xi)$, is the set of points u for which the design measure is positive $\xi(u) > 0$. This is only a theoretical consideration. It might be a design point optimal, but with zero design measure, $\xi(u) = 0$, at this point u , and therefore with no practical use, as the experimenter is not taking observations at this point. Practically speaking the design measure acts as the proportion of observations devoted to the optimal design points. That is why sometimes the following way to present a design measure is adopted

$$\xi = \left\{ \begin{array}{cccc} u_1 & u_2 & \dots & u_k \\ p_1 & p_2 & \dots & p_k \end{array} \right\}, \sum p_i = 1, p_i = \frac{n_i}{n}$$

What are the optimal design points? One can certainly perform his/her experiment at different values of the input variables. The optimal values are those values which will provide a “useful predefined result”—example: minimize the variance of the parameters.

Let us denote by $\text{Mat}(v, m)$ the set of $v \times m$ size matrices and let $\theta \in \text{Mat}(p, 1)$ be the vector of parameters, $u_i \in \text{Mat}(1, k)$ the predictor variable, $i = 1, 2, \dots, n$, with n the sample size.

For the response y it is assumed that either $y \in \Psi \subseteq \mathbb{R}$ or $y \in \{0, 1, \dots, \lambda\}$, $\lambda \geq 1$. When the response y is supposed to take any value in Ψ , it is also supposed that a regression model (in general nonlinear) exists consisting in the deterministic portion $f(u, \theta)$ and the stochastic portion, e , known as error, linked through the (continuous) regression formulation

$$y_i = f(u_i, \theta) + e_i. \quad (2.2.2)$$

Let us assume that $f(u, \theta) = \theta^T g(u)$ with g being a (vector) continuous function of u , θ^T is the transpose vector of θ . Then the nonlinear problem is reduced to the so called linear problem.

When $\Psi = \{0, 1, 2, \dots, \lambda\}$ the multiresponse problem is obtained. The most common case is $\lambda = 1$, i.e. a binary response. In this case the outcome takes only

two values, $Y_i = 1$ or 0 , and it is linked with the covariates and the parameters through a (discrete) probability model “T” with

$$P(Y_i = 1) = T(u_i, \theta), P(Y_i = 0) = 1 - T(u_i, \theta), \quad (2.2.3)$$

where u_i is the value of u going with observation Y_i , $i = 1, 2, \dots, n$

Example 1 Typical situation in Bioassays is to consider the models:

$$\begin{aligned} \text{Logit model : } T(u, \theta) &= \{1 + \exp(-\theta_1(u - \theta_2))\}^{-1}, \text{ or} \\ \text{Probit model : } T(u, \theta) &= (\sqrt{2\pi}\theta_2)^{-1} \int_{-\infty}^u \exp(-(u - \theta_1)^2 / (2\theta_2^2)) du. \end{aligned} \quad (2.2.4)$$

In both cases the parameter vector is $\theta = (\theta_1, \theta_2)$, with θ_1 the location parameter and θ_2 the scale parameter.

2.3 Assumptions

Every theoretical approach is based on a number of assumptions. It is proposed to keep the number of these assumptions limited. The following main assumptions will be considered throughout this book. Let us refer to them as Assumption 1 or 2 etc. when they are recalled. When limiting results for the sequence of estimators θ_n are considered the parameter space Θ is assumed to be compact. This is needed, as in compact set (closed and bounded set) any sequence from this set converges within this set, due to Bolzano-Weierstrass Theorem. For the errors the main assumption which is imposed is:

Assumption 1 The errors e_i are independent and identically distributed (iid) with $E(e_i) = 0$ and $V(e_i) = \sigma^2 > 0$, $i = 1, 2, \dots, n$.

Under Assumption 1 for model (2.2.2) is: $\eta = \eta(u, \theta) = E(y) = f(u, \theta)$.

And for the model (2.2.3) is: $\eta = \eta(u, \theta) = E(y) = T(u, \theta)$.

Assumption 2 The iid (independent identically distributed) errors are normally distributed with mean 0 and variance $\sigma^2 > 0$.

As far as the function f concerned is needed to be smooth in the neighborhood of θ_t , the true value of the parameter θ . That is why it is assumed:

Assumption 3 The function $f(u, \theta)$ is continuous in Θ at θ_t and the second order derivatives of f with respect to θ exist at and near θ_t .

For the introduced function T , concerning the binary response problems, recall relation (2.2.3), it is assumed that:

Assumption 4 $T(u)$ is a monotonic differentiable function.

Moreover the function $T(u)$ can be considered as a cumulative distribution function of the random variable Z defined through the random variable Y as follows

$$Y = \begin{cases} 1 & \text{if } Z \leq u \\ 0 & \text{if } Z > u. \end{cases}$$

Then: $P[Y = 1] = P[Z \leq u] = T(u)$, say, and $P[Y = 0] = P[Z > u] = 1 - T(u)$.

Thus: $E(Y) = 1 \times T(u) + 0 \times (1 - T(u)) = T(u) = \eta$ and $\text{Var}(Y) = T(u) \times (1 - T(u))$.

So the expected value and the variance of the new random variable Y are evaluated through the function $T(u)$.

Assumption 5 The assumed model used to plan the design is correct.

This Assumption is essential in practice, as the experimentalist, very often, tries to elucidate which model is appropriate, with no statistical background for such a target.

The theoretical framework is based on these assumptions, which can be easily followed by the experimenter. Therefore it is recommended that the experimentalist should see if are fulfilled.

2.4 On the Existence of Estimators

When the experimental data is collected, the question arises whether it is possible to obtain the appropriate estimates of the involved parameters. Therefore the existence of the estimators requires a particular theoretical investigation. For the (continuous) model (2.2.2) the following quantity is introduced

$$S_n(\theta) = \sum (y_i - f(u_i, \theta))^2 = \|y - f(u, \theta)\|_2^2 \quad (2.4.1)$$

where $\|\cdot\|_2$ is the l_2 -norm. An estimate $\hat{\theta}$ is the *least squares estimate* (LSE) if

$$S_n(\hat{\theta}) = \min\{S_n(\theta); \theta \in \Theta\} \quad (2.4.2)$$

Under Assumption 2 it is known that this LSE coincides with the maximum likelihood estimator (MLE). It can be proved that there exists such a LSE, under certain conditions, see Wu (1981), who relaxed the conditions imposed in the pioneering paper of Jennrich (1969).

For the binary response problem Silvapulle (1981) provided that conditions under which the likelihood function L ,

$$L \propto \prod \{T(u_i, \theta)\}^{y_i} \{1 - T(u_i, \theta)\}^{1-y_i} \quad (2.4.3)$$

can provide maximum likelihood estimators (Appendix A2). Roughly speaking this occurs when the intersection of the sets of values taken by the explanatory variables corresponding to successes and to failures is not the null set (see footnotes of Tables A4.1, A4.2, A4.4, in Appendix 4). This happens to be a necessary and sufficient condition for the logit and probit models. Practically, a proportion can not be estimated when only a data set of “successes” or only “failures” is available. Moreover the estimators are not appropriate if the number of “successes” (“failures”) is very small comparing to the total number of observations.

Now, having ensured that the likelihood equation can provide MLE and denoting by l the log-likelihood the following matrix is defined

$$S(\hat{\theta}, \xi_n, y) = - \left(\frac{\partial^2 \ell}{\partial \theta_i \partial \theta_j} \middle| \theta = \hat{\theta} \right) \quad (2.4.4)$$

where ξ_n is the design measure on n observations, called the *sample information matrix*.

Example 2 Maximum likelihood estimates for the logit model can be obtained through the, well known from the linear case, *normal equations*

$$\sum T_i = \sum y_i \sum u_i T_i = \sum y_i u_i,$$

with $T_i = T(u_i; \theta)$ as in (2.2.4).

One of the most important measures in Statistics is (the parametric) Fisher’s information measure. There is also the entropy type of Fisher’s information measure, but this is beyond the target of this book, see Kitsos (2011a, b). The Fisher’s information matrix is introduced in the next section.

2.5 Fisher’s Information Matrix

There are two different types of Fisher’s information measures: the parametric one, which will be presented next, and the entropy type which has been extended (Kitsos and Tavoularis 2009), and is not discussed in this monograph. Both are strongly related with the involved uncertainty in the physical phenomenon under investigation. Sometimes, due to the cost of the experiment, the experimentalist needs to run a small number of trials, and eventually estimate, adequately, the involved parameters—one can imagine how useful the theoretical limiting results are in such cases! But Fisher’s information is always there, helpful to the analysis, even with a small sample size, see Sect. 6.4.

In non-linear problems the variance depends on the unknown parameter the experimenter wants to estimate, i.e. $\sigma^2 = \sigma^2(u, \theta)$. That is σ^2 depends on the design point and the parameter vector. In the linear case it is assumed independent of the parameter vector θ . In practice it may or may not be possible to assume that

is “known”, provided a guess, or knowledge from a previous experiment. In principle, this is the main difference between the linear and the non-linear case.

Let $\nabla\eta$ denote the vector of partial derivatives (recall Assumption 1)

$$\nabla\eta = \left(\frac{\partial\eta}{\partial\theta_1}, \frac{\partial\eta}{\partial\theta_2}, \dots, \frac{\partial\eta}{\partial\theta_p} \right)^T. \quad (2.5.1)$$

Then for the exponential family of models Fisher’s information matrix is defined to be

$$I(\theta, u) = \sigma^{-2}(\nabla\eta)(\nabla\eta)^T. \quad (2.5.2)$$

Moreover in many of the nonlinear problems the covariate u and the parameter θ appear together linearly in the form $\theta^T u$. Thus the following proposition holds:

Proposition 2.1 *If $\eta = \eta(\theta^T u)$ it can be proved that $\nabla\eta = [w(\theta^T u)]^{1/2} u$ with $w(z) = [\partial\eta/\partial z]^2$, $z = \theta^T u$ and therefore:*

$$I(\theta, u) = \sigma^{-2} w(\theta^T u) u u^T.$$

This result describes that fact that if the nonlinear model is “*intrinsic linear*” as far as the parameter concerns, $\eta = \eta(\theta^T u)$, then Fisher’s information matrix can be evaluated proportional to the matrix produced only by the input vector u , i.e. equal to $\sigma^{-2} w(\theta^T u) u u^T$.

The concept of the *average-per-observation information matrix* will play an important role in the developed scenario concerning the nonlinear experiment design problem. It is defined for the ξ_n , the n -point design measure, to be equal:

For the discrete case:

$$M(\theta, \xi_n) = n^{-1} \sum I(\theta, u_i),$$

For the continuous case:

$$M(\theta, \xi) = \int_U I(\theta, u) \xi(du), \quad \xi \in \Xi. \quad (2.5.3)$$

The theoretical insight of Caratheodory’s Theorem (Appendix 1), so essential for the linear experiment design, can also be used for the average information matrix in nonlinear problems, see Titterington (1980a). Now, suppose that the matrix $M = M(\theta, \xi)$ is partitioned in the form:

$$M = \begin{pmatrix} M_{11} & M_{12} \\ M_{21} & M_{22} \end{pmatrix} \quad (2.5.4)$$

with $M_{11} \in \text{Mat}(s, s)$, $M_{12} \in \text{Mat}(s, p - s)$, $M_{22} \in \text{Mat}(p - s, p - s)$, $1 \leq s < p$. Then the following matrix is defined:

$$\mathbf{M}_S = \mathbf{M}_S(\theta, \xi) = \mathbf{M}_{11} - \mathbf{M}_{12}\mathbf{M}_{22}^- \mathbf{M}_{12}^T \quad (2.5.5)$$

with \mathbf{M}_{22}^- being the generalized inverse of \mathbf{M}_{22} . The information matrix \mathbf{I} , as in (2.5.2), is considered partitioned in the same fashion. This partition is helpful when interest lies in estimating the leading $s < p$ parameters in the vector θ , as it will be explained in Sect. 3.6.

2.6 Linearization of the Model

The idea of the (design) matrix \mathbf{X} , including all the input variables, is essential when one is working with linear models, see Seber and Wild (1989) among others. In nonlinear models the researcher can not define a matrix \mathbf{X} in the same fashion. This can be done only approximately through the partial derivatives of θ , with θ taking its “true” value, θ_t . Then the following $n \times p$ matrix is defined

$$\mathbf{X} = (x_{ij}) = \frac{\partial f(u_i, \theta)}{\partial \theta_j} \big|_{\theta=\theta_t}. \quad (2.6.1)$$

Then the matrix $\mathbf{X} = \mathbf{X}(\theta)$ is formed as a function of θ . The function $f(u, \theta)$, recall (2.2.2), can be “linearized” through a Taylor series expansion, in the neighborhood of θ_t as:

$$f(u, \theta) = f(u, \theta_t) + \sum (\theta_j - \theta_{tj})(\partial f(u, \theta)/\partial \theta_j) \big|_{\theta=\theta_t} \quad (2.6.2)$$

Following the pattern of linear regression models in the nonlinear regression case, an approximation to the covariance matrix, of the estimates of the parameters, can be defined as:

$$\mathbf{C} \cong [\mathbf{X}^T(\theta_t)\mathbf{X}(\theta_t)]^{-1} \sigma^2. \quad (2.6.3)$$

Moreover for all nonlinear problems a useful approximation to the covariance matrix is $\mathbf{C}^{-1} \cong n\mathbf{M}(\theta_t, \xi)$. Eventually for the average per-observation information matrix \mathbf{M} for the design measure ξ holds:

$$\mathbf{M}(\theta, \xi) = \frac{1}{n} (\nabla \eta)^T \nabla \eta = \frac{\sigma^2}{n} \mathbf{C}^{-1}(\hat{\theta}, \xi) \quad (2.6.5)$$

Relation (2.6.5) describes that (asymptotically) the covariance matrix is the inverse of the average per observation information matrix.

2.6.1 Examples

To clarify the above theoretical background, the following examples are given, helpful for the inquiring reader.

Example 3 Recall Examples 1 and 2, as far as the logit model concerns. Given that the linear approximation of the form

$$[1 + \exp(-\theta_1(u - \theta_2))]^{-1} \cong 1/2 + 1/6 \theta_1(u - \theta_2)$$

holds under the assumption $|\theta_1(u - \theta_2)| \leq 3$, then the normal equations, for the logit model, approximately are:

$$\begin{aligned} n/2 + (\theta_1/6) \sum (u_i - \theta_2) &= \sum y_i \\ (1/2) \sum u_i + (\theta_1/6) \sum u_i(u_i - \theta_2) &= \sum u_i y_i \end{aligned}$$

Solving the simultaneous equations the estimates of the parameters (θ_1, θ_2) are obtained.

In the following examples the idea of Fisher's information matrix is clarified. These examples will be reconsidered in the sequel of this development.

Example 4 Consider that model, in which $P(y = 1) = T(\theta^T u)$. That is the model T is “intrinsic linear”: the model can be non-linear, but the parameters and the input variable are related linearly.

Let: $\theta_1 + \theta_2 u_1 = z = \theta^T u$ and $T'(z) > 0$, $\theta = (\theta_1, \theta_2)$, $u = (1, u_1)$.

Then the log-likelihood ℓ , will be $\ell = \log\{T(z)^Y [1 - T(z)]^{1-Y}\} + \text{const.}$

Therefore Fisher's information matrix equals to

$$I(\theta, \xi) = E\{(\nabla \ell)(\nabla \ell)^T\} = \alpha(\theta) u u^T, \text{ with } \alpha(\theta) = T'^2[T(1 - T)].$$

Application: T might be either the logit or probit model.

Example 5 For the nonlinear regression model $\eta = \theta_1 - \exp(-\theta_2 u)$, $u \in [-1, 1]$

Fisher's information matrix is evaluated equal to

$$I(\theta, \xi) = (\nabla \eta)(\nabla \eta)^T = \begin{pmatrix} 1 & u \exp(-\theta_2) \\ u \exp(-\theta_2) & u^2 \exp(-\theta_2) \end{pmatrix}$$

Note that $I(\theta, \xi)$ is a function only of θ_2 , $I(\theta, \xi) = I(\theta_2, \xi)$, therefore the covariance matrix, recall relation (2.6.5), depends on the parameters, but only on the component θ_2 of the parameter vector (θ_1, θ_2) . Notice that in linear models the covariance matrix does not depend on the parameters.

Example 6 The nonlinear regression model used to describe growth phenomena is

$$\eta = \theta_1 \exp(\theta_2 u), u \in [a, b]$$

The x_{ij} vectors $i = 1, 2, \dots, n$, $j = 1, 2$ can be formed according to (2.6.1) as

$$x_{i1} = \partial\eta/\partial\theta_1 = \exp(\theta_2 u_i), \quad x_{i2} = \partial\eta/\partial\theta_2 = \theta_1 u_i \exp(\theta_2 u_i)$$

The quantity $S_n(\theta) = \sum (y_i - \theta_1 \exp(\theta_2 u_i))^2$ is formed. In order to find the Least Square Estimator $\hat{\theta}$, which minimizes $S_n(\theta)$, the partial derivatives of $S_n(\theta)$ are needed, which provide the “normal equations”

$$\begin{aligned} \sum (y_i - \theta_1 \exp(\theta_2 u_i)) \exp(\theta_2 u_i) &= 0 \\ \sum (y_i - \theta_1 \exp(\theta_2 u_i)) \theta_1 u_i \exp(\theta_2 u_i) &= 0. \end{aligned}$$

Moreover, in this case, from the *Hessian matrix* and its’ expected value it is eventually evaluated (recall (2.6.5)) that

$$\sigma^{-2} n M(\theta, \xi_n) = \begin{pmatrix} \sum \exp(2\theta_2 u_i) & \theta_1 \sum u_i \exp(2\theta_2 u_i) \\ \theta_1 \sum u_i \exp(2\theta_2 u_i) & \theta^2 \sum u^2 \exp(2\theta_2 u_i) \end{pmatrix}.$$

Also, from (2.4.4) is evaluated that

$$\sigma^{-2} S(\hat{\theta}, \xi_n, y) = \begin{pmatrix} \sum \exp(2\hat{\theta}_2 u_i) & \hat{\theta}_1 \sum u_i \exp(2\hat{\theta}_2 u_i) \\ \hat{\theta}_1 \sum u_i \exp(2\hat{\theta}_2 u_i) & A(y, u, \hat{\theta}) \end{pmatrix} \quad (2.7.1)$$

where $-A(y, u, \hat{\theta}) = \hat{\theta}_1 \sum u_i^2 \exp(\hat{\theta}_2 u_i) \{y_i - 2\hat{\theta}_1 \exp(\hat{\theta}_2 u_i)\}$.

Example 7 Dose-Finding in Phase I/II Clinical Trials

The goals of a clinical trial, Thall and Russel (1998), are:

1. To perform an experiment that satisfies specific safety and efficacy requirements,
2. To consider the trial early as it is likely that no dose is both safe and efficacious,
3. To estimate, with some risk, the rates of the events at the pre-decided level of dose.

The dose response curve, for the logit model (recall Example 1), is the cumulative odds model, to describe and to approach, the unknown dose-response function. So the underlying model, describing the experiment, is nonlinear. A binary indicator describes the levels of severe toxicity. The appropriate Clinical Trial imposes a target to estimate that optimal level of the dose, among different candidate dose levels, that satisfies both the efficacy and toxicity criteria.

For examples of non-linear models in Cancer Bioassays see Kitsos et al. (2009).

Example 8 Oxidation of benzene.

This model is, among the chemical reaction models, one with the most necessary needed information: Includes 4 input variables and 4 parameters:

$$\text{Model: } \eta = \frac{\theta_1 \exp(-\theta_3 u_3) \theta_2 \exp(-\theta_4 u_3) u_1 u_2}{\theta_1 \exp(-\theta_3 u_3) u_1 + u_4 \theta_2 \exp(-\theta_4 u_3) u_2}$$

The input variables and the design space U are defined below, while the four involved parameters belong to the parameter space $\Theta \subseteq R^4$,

$$u = (u_1, u_2, u_3, u_4) \in U = \Delta_1 \times \Delta_2 \times \Delta_3 \times \{5.75\} \subseteq R^4, \text{ with} \\ \Delta_1 = [10^{-3}, 16 \times 10^{-3}], \Delta_2 = [10^{-3}, 4 \times 10^{-3}], \Delta_3 = [623, 673].$$

Moreover the introduced notation is:

η	The initial reaction rate
u_1	The concentration of oxygen
u_2	The concentration of benzene
$u_3 = 1/T - 0.0015428$	T is the absolute temperature of the reaction
u_4	The observed stoichiometric number
$\theta_1, \theta_2, \theta_3, \theta_4$	Model parameters arising from Arrhenius' law.

Usually the real situation is not such complicated, see [Sect. 4.7](#), i.e. indeed the parameters involved are 2 or 3 and usually no particular knowledge concerning one of them is needed. For the n -th order chemical kinetic models see Manerswamy et al. (2009), while for a compilation of the Chemical Kinetics models and the corresponding design theory see Kitsos and Kolovos (2013).

The above examples provide evidence that the nonlinear optimal experimental design theory covers a broad range of applications as it will be also clarified in the sequence.

References

- Jennrich RJ (1969) Asymptotic properties of non-linear least-squares estimators. *Ann Math Stat* 40:633–643
- Kitsos CP (2011a) Invariant canonical form for the multiple logistic regression. *Math Eng Sci Aerosp (MESA)* 2(3):267–275
- Kitsos CP (2011b) On Fisher's information measures. *e-Proceedings of digital technologies 2011*, Zilina, Slovakia, 10–11 Nov 2011, pp 178–183
- Kitsos CP, Kolovos KG (2013) A compilation of the D-optimal designs in chemical kinetics. *Chemical Engin Comm* 200:1–20
- Kitsos CP, Tavoularis NK (2009) Logarithmic Sobolev inequalities for information measures. *IEEE Trans Inform Theory* 55(6):2554–2561
- Kitsos CP, Toulas TL, Lolas G (2009) Modeling cancer bioassays. *Biometrie und Medizinische Informatik Greifswalder Seminarberichte*, Heft, vol 15, pp 75–90
- Manerswamy A, Munson-McGee S, Steiner R, Johnson CI (2009) D-optimal design of DSC experiments for n th order kinetics. *J Therm Anal Calorim* 97:895–902
- Seber CAF, Wild CJ (1989) *Nonlinear regression*. Wiley, London
- Silvapulle MJ (1981) On the existence of the maximum likelihood estimators for the Binomial response models. *J.R.S.S.B.* 43:310–313
- Thall PF, Russel K (1998) A strategy for dose-finding and safety monitoring based on efficacy and adverse outcomes in phase I/II clinical trials. *Biometrics* 54:251–264
- Titterton DM (1980) Aspects of optimal design in dynamic systems. *Technometrics* 22:287–299
- Wu CFJ (1981) Asymptotic theory of nonlinear least squares estimates. *Ann Stat* 9:501–513

<http://www.springer.com/978-3-642-45286-4>

Optimal Experimental Design for Non-Linear Models
Theory and Applications

Kitsos, C.P.

2013, IX, 102 p. 8 illus., Softcover

ISBN: 978-3-642-45286-4