journal homepage: www.intl.elsevierhealth.com/journals/cmpb

Experimental design for regression analysis when the responses are subject to censoring

*Cong Han**

TAP Pharmaceutical Products Inc., 675 North Field Drive, Lake Forest, IL 60045, United States

ARTICLE INFO

Article history:
Received 29 September 2006
Received in revised form
30 April 2007
Accepted 2 May 2007

Keywords:
Optimal design
D_s-optimality
Nonlinear regression
Tobit model

ABSTRACT

Experimental design issues are investigated for regression models with possibly censored responses, arising typically from pharmacokinetic or virus dynamic experiments. Examples are provided for both locally and Bayesian optimal designs. A case study in pharmacokinetics is provided.

1. Introduction

Censored measurements arise quite often in biomedical experiments. Pharmacokinetics and virus dynamics are two examples.

Pharmacokinetic models describe the change of drug concentration over time. Since drug concentration is measured using an assay, which typically has been validated within a range, the possibility exists that certain actual measurements fall below the lowest concentration at which the validation has been performed. In such cases, it is reported that the concentration is below the quantification limit of the assay [1].

In a simplistic form, virus dynamics can be viewed as a model that describes the change of virus concentration over time. Assays that are used to measure human immunodeficiency virus (HIV) RNA and hepatitis C virus (HCV) RNA levels also have lower limits of detection or quantification [2–4].

In both pharmacokinetics and virus dynamics, the measurements are usually expressed as solutions to a system of ordinary differential equations; when the solutions exist in closed form, nonlinear regression models arise. For example, in ref. [4], the concentration of HCV RNA at time t, V(t), follows

$$V(t) = \beta[\alpha e^{-\lambda_1(t-\tau)} + (1-\alpha)e^{-\lambda_2(t-\tau)}], \quad t > \tau, \tag{1}$$

where $\lambda_{1,2} = (\gamma+\delta)/2 \pm \sqrt{(\gamma-\delta)^2 + 4(1-\omega)\gamma\delta}/2$, $\alpha = (\omega\gamma - \lambda_2)/(\lambda_1 - \lambda_2)$, β, τ, ω, γ and δ are the parameters to be estimated. A nonlinear regression model on the log scale would specify

$$y_i = \log\beta + \log[\alpha e^{-\lambda_1(t_i-\tau)} + (1-\alpha)e^{-\lambda_2(t_i-\tau)}] + \eta_i, \quad \eta_i \overset{iid}{\sim} N(0,\sigma^2), \tag{2}$$

where $N(0,\sigma^2)$ denotes a normal distribution with mean 0 and variance σ^2.

* Tel.: +1 847 582 6472.
E-mail address: cong.han@tap.com.
0169-2607/$ – see front matter © 2007 Elsevier Ireland Ltd. All rights reserved.
doi:10.1016/j.cmpb.2007.05.001

Taking into consideration a lower limit of detection at 100 copies per milliliter, the model would incorporate, in addition to the above,

$$y_i^* = \max\{\log 100, y_i\}. \tag{3}$$

More generally, the regression model in which the response may be censored can be written as

$$y_i = f(\boldsymbol{\theta}, \mathbf{x}_i) + \epsilon_i, \quad \epsilon_i \overset{\text{iid}}{\sim} N(0, \sigma^2), \quad y_i^* = \max\{c, y_i\}, \tag{4}$$

where $\boldsymbol{\theta}$ and σ^2 are the unknown parameters and c is a constant. Such a model can be fit via maximum likelihood. When the mean function for y is linear in the parameters, it reduces to the standard Tobit model [5,6], also known as the limited dependent variable model:

$$y_i = \mathbf{x}_i^{\mathrm{T}}\boldsymbol{\theta} + \epsilon_i, \quad \epsilon_i \overset{\text{iid}}{\sim} N(0, \sigma^2), \quad y_i^* = \max\{c, y_i\}, \tag{5}$$

where for linear models c is usually assumed to be 0 without loss of generality.

2. Fisher information and optimality

In the usual regression models of the form $y_i = f(\boldsymbol{\theta}, \mathbf{x}_i) + \epsilon_i$, $i = 1, \ldots, n$, where the ϵ_i's are independent and identically distributed normal random variables with mean 0 and variance σ^2, the Fisher information matrix with respect to $(\boldsymbol{\theta}^{\mathrm{T}}, \sigma^2)^{\mathrm{T}}$ can be written as

$$\begin{pmatrix} \frac{1}{\sigma^2}\sum_{i=1}^{n} \nabla f(\boldsymbol{\theta}, \mathbf{x}_i)\nabla^{\mathrm{T}} f(\boldsymbol{\theta}, \mathbf{x}_i) & 0 \\ 0 & \frac{n}{2\sigma^4} \end{pmatrix}, \tag{6}$$

where ∇ denotes the gradient with respect to $\boldsymbol{\theta}$.

Suppose in designing an experiment, the predictor $\mathbf{x}$ is constrained in a compact subset $\mathcal{X}$ of a Euclidean space. Let Ξ be the class of probability distributions on the Borel sets on $\mathcal{X}$, then any $\xi \in \Xi$ is called a design measure on $\mathcal{X}$. The concept of Fisher information can be extended so that a design measure ξ with sample size n has an information matrix

$$\begin{pmatrix} \frac{n}{\sigma^2}\int \nabla f(\boldsymbol{\theta}, \mathbf{x})\nabla^{\mathrm{T}} f(\boldsymbol{\theta}, \mathbf{x})\, d\xi(\mathbf{x}) & 0 \\ 0 & \frac{n}{2\sigma^4} \end{pmatrix}. \tag{7}$$

An exact design of size n, which can be physically implemented, can be viewed as a design measure ξ supported at $x_1, \ldots, x_k$, with probability masses $\xi_1, \ldots, \xi_k$, such that $n\xi_j \in \mathbb{N}$ for $j = 1, \ldots, k$.

Under the usual regularity conditions, the asymptotic covariance matrix of the maximum likelihood estimate of $\boldsymbol{\theta}$ is proportional to $\sigma^2\left[\int \nabla f(\boldsymbol{\theta}, \mathbf{x})\nabla^{\mathrm{T}} f(\boldsymbol{\theta}, \mathbf{x})\, d\xi(\mathbf{x})\right]^{-1}$. It is hence conventional to assume σ^2 known and equal to 1, as well as $n = 1$; the Fisher information matrix (with respect to $\boldsymbol{\theta}$ only) then becomes $\int \nabla f(\boldsymbol{\theta}, \mathbf{x})\nabla^{\mathrm{T}} f(\boldsymbol{\theta}, \mathbf{x})\, d\xi(\mathbf{x})$. The usual optimality criteria are real-valued (or extended real-valued) functions of this information matrix [7].

In the presence of censoring, the block-diagonality of the Fisher information matrix does not hold any more. Assuming that the primary interest remains in $\boldsymbol{\theta}$, and σ^2 is treated as a nuisance parameter, the suitable optimality criterion is D_s-optimality defined as follows. For design ξ, let $l(\boldsymbol{\theta}, \sigma^2, \mathbf{x})$ be the log-likelihood function evaluated at $\mathbf{x}$ and let $M(\xi; \boldsymbol{\theta}, \sigma^2)$ denote the Fisher information matrix for design ξ:

$$-\int \left[E\begin{pmatrix} \frac{\partial^2}{\partial\boldsymbol{\theta}\partial\boldsymbol{\theta}^{\mathrm{T}}} l(\boldsymbol{\theta}, \sigma^2, \mathbf{x}) & \frac{\partial^2}{\partial\boldsymbol{\theta}\partial\sigma^2} l(\boldsymbol{\theta}, \sigma^2, \mathbf{x}) \\ \frac{\partial^2}{\partial\sigma^2\partial\boldsymbol{\theta}^{\mathrm{T}}} l(\boldsymbol{\theta}, \sigma^2, \mathbf{x}) & \frac{\partial^2}{\partial(\sigma^2)^2} l(\boldsymbol{\theta}, \sigma^2, \mathbf{x}) \end{pmatrix} \right] d\xi(\mathbf{x}), \tag{8}$$

where the expectation is with respect to y^*. Let $A^{\mathrm{T}} = (I_s\ 0)$, a $s \times (s+1)$ matrix, where s is the dimension of $\boldsymbol{\theta}$, then the D_s-optimality criterion for estimating $\boldsymbol{\theta}$ (a subset of the parameters, which characterize the mean function) is defined as

$$\varphi(M(\xi; \boldsymbol{\theta}, \sigma^2)) = -\log\det[A^{\mathrm{T}} M^{-1}(\xi; \boldsymbol{\theta}, \sigma^2) A]. \tag{9}$$

For given values of $\boldsymbol{\theta}$ and σ^2, the locally D_2-optimal design ξ^*, with $\varphi(\cdot)$ defined as above, is such that

$$\varphi(M(\xi^*; \boldsymbol{\theta}, \sigma^2)) = \max_{\xi \in \Xi} \varphi(M(\xi; \boldsymbol{\theta}, \sigma^2)). \tag{10}$$

When a prior distribution $\pi(\boldsymbol{\theta}, \sigma^2)$ is specified for the unknown parameters, a Bayesian D_s-optimal design maximizes the Bayesian D_s-optimality:

$$-\int \log\det[A^{\mathrm{T}} M^{-1}(\xi; \boldsymbol{\theta}, \sigma^2) A]\, d\pi(\boldsymbol{\theta}, \sigma^2). \tag{11}$$

For notational simplicity, let $f_i = f(\boldsymbol{\theta}, \mathbf{x}_i)$; define $\delta_i = 1$ if $y_i^* > c$ and $\delta_i = 0$ otherwise. Then, the log-likelihood function for case i is

$$(1 - \delta_i)\log\Phi\left(\frac{c - f_i}{\sigma}\right) - \frac{\delta_i}{2\sigma^2}(y_i - f_i)^2 - \frac{\delta_i}{2}\log(2\pi\sigma^2), \tag{12}$$

where $\Phi(\cdot)$ is the cumulative distribution function of a standard normal random variable.

Let $\phi(\cdot)$ be the density function of a standard normal random variable and define $k_i = (c - f_i)/\sigma$ and $g_i = \partial f_i/\partial\boldsymbol{\theta}$. After tedious algebra and calculus, it can be seen that up to a weight determined by the design ξ, an observation i contributes

$$\begin{pmatrix} \frac{1}{\sigma^2}\left[k_i\phi(k_i) + \frac{\phi^2(k_i)}{\Phi(k_i)} + \Phi(-k_i)\right] g_i g_i^{\mathrm{T}} & \frac{1}{2\sigma^3}\phi^2(k_i)\left[1 + k_i^2 + k_i\frac{\phi(k_i)}{\Phi(k_i)}\right] g_i \\ \frac{1}{2\sigma^3}\phi^2(k_i)\left[1 + k_i^2 + k_i\frac{\phi(k_i)}{\Phi(k_i)}\right] g_i^{\mathrm{T}} & \frac{k_i}{4\sigma^4}\phi(k_i) + \frac{k_i^3}{4\sigma^4}\phi(k_i) + \frac{k_i^2}{4\sigma^4}\frac{\phi^2(k_i)}{\Phi(k_i)} + \frac{1}{2\sigma^4}\Phi(-k_i) \end{pmatrix} \tag{13}$$