Available online at www.sciencedirect.com
SCIENCE DIRECT®

Journal of Statistical Planning and Inference 130 (2005) 63–83

journal of statistical planning and inference

www.elsevier.com/locate/jspi

Corrected confidence intervals for adaptive nonlinear regression models

D.S. Coad[a], M.B. Woodroofe[b,*]

[a]Department of Mathematics, University of Sussex, Falmer, Brighton BN1 9RF, UK
[b]Department of Statistics, University of Michigan, Ann Arbor, MI 48109-1092, USA

Received 11 August 2003; accepted 2 February 2004
Available online 29 July 2004

Abstract

A nonlinear regression model is considered in which the design variable may be a function of the previous responses. The aim is to construct confidence intervals for the parameter which are asymptotically valid to a high order. This is accomplished by using a tilting argument to construct a first approximation to a pivotal quantity, and then by using a version of Stein's identity and very weak expansions to determine the correction terms. The accuracy of the approximations is assessed by simulation for two well-known nonlinear regression models—the first-order growth or decay model and the Michaelis–Menten model, when one of the two parameters is known. Detailed proofs of the expansions are given.

Keywords: Approximately pivotal quantity; Maximum likelihood estimator; Posterior distribution; Stein's identity; Tilted approximation; Very weak expansion

1. Introduction

Consider a nonlinear regression model of the form

$$y_k = g(x_k; \theta) + \varepsilon_k, \quad k = 1, 2, \ldots,$$

where $x_k \in \mathscr{X}$ is a design variable, $\theta \in \Omega$ is an unknown parameter, $g : \mathscr{X} \times \Omega \to \mathbb{R}$ is a nonlinear function of a known form and $\varepsilon_1, \varepsilon_2, \ldots$ are independent standard normal random

* Corresponding author.
E-mail address: michaelw@umich.edu (M.B. Woodroofe).

0378-3758/$ - see front matter
doi:10.1016/j.jspi.2004.02.020

variables. Here, $\mathscr{X}$ is a compact space and Ω is a compact interval. Denote the endpoints of Ω by $\underline{\theta}$ and $\bar{\theta}$, so that $\Omega = [\underline{\theta}, \bar{\theta}]$. It is assumed that g has five derivatives in θ, that the derivatives are jointly continuous in (x, θ) and that $g'(x; \theta) > 0$ for all $x \in \mathscr{X}$ and $\theta \in \Omega$, where $'$ denotes differentiation with respect to θ. The model is adaptive in the sense that

$$x_k = x_k(y_1, \ldots, y_{k-1}), \quad k = 1, 2, \ldots .$$

The above model is quite general and includes some well-known nonlinear models. Recent surveys of adaptive nonlinear regression models are given by Ford et al. (1989) and Chaudhuri and Mykland (1993).

An adaptive normal linear model with known variance was studied by Woodroofe (1989) and Woodroofe and Coad (1997). Using a Bayesian approach and Stein's (1981) identity, Woodroofe (1989) obtained asymptotic expansions for sampling distributions and Woodroofe and Coad (1997) constructed corrected confidence sets for θ. The latter results were applied to certain three-treatment clinical trial models. The case of unknown variance was considered by Coad and Woodroofe (1998) and Woodroofe and Coad (1999), and approximations for biases, variances and confidence sets were evaluated for several other examples. A one-parameter adaptive normal nonlinear model has been studied by Woodroofe (1991), who used a Bayesian approach and a Taylor series to obtain asymptotic expansions for sampling distributions. The aim of the present paper is to obtain corrected confidence intervals for the one-parameter case above, but using Stein's identity instead of a Taylor series. Another difference between the present work and that of Woodroofe (1991) is that a tilting argument is used in order to construct a first approximation to a pivotal quantity; see, for example, Barndorff-Nielsen and Cox (1989).

It is well known that the likelihood function is not affected by the adaptive nature of the model or by the use of a stopping time; see, for example, Berger and Wolpert (1984). Thus, if $y_1, \ldots, y_n$ are observed and $\ell_n(\theta)$ denotes the log-likelihood function,

$$\ell_n(\theta) = -\frac{1}{2}\sum_{k=1}^{n}\{y_k - g(x_k;\theta)\}^2,$$

$$\ell_n'(\theta) = \sum_{k=1}^{n}\{y_k - g(x_k;\theta)\}g'(x_k;\theta),$$

$$\ell_n''(\theta) = \sum_{k=1}^{n}[\{y_k - g(x_k;\theta)\}g''(x_k;\theta) - g'(x_k;\theta)^2],$$

$$\ell_n'''(\theta) = \sum_{k=1}^{n}[\{y_k - g(x_k;\theta)\}g'''(x_k;\theta) - 3g'(x_k;\theta)g''(x_k;\theta)]$$

and

$$\ell_n^{(4)}(\theta) = \sum_{k=1}^{n}[\{y_k - g(x_k;\theta)\}g^{(4)}(x_k;\theta) - 3g''(x_k;\theta)^2 - 4g'(x_k;\theta)g'''(x_k;\theta)],$$

where $\ell_n^{(4)}$ denotes the fourth derivative of ℓ_n. Next, let $\pi_n(x)$ denote the design distributions,

$$\pi_n\{B\} = \frac{1}{n}\#\{k \leqslant n : x_k \in B\}$$

for Borel sets $B \subset \mathscr{X}$ and $n = 1, 2, \ldots,$ and suppose that the design is *stable*, so that there are distributions $\pi(\theta; .)$ for which

$$\pi_n \Rightarrow \pi(\theta; .)$$

in P_θ-probability as $n \to \infty$ for each $\theta \in \Omega$, where $\Rightarrow$ denotes weak convergence and the $\pi(\theta; .)$ depend continuously on θ. Then, as shown in Section 8,

$$\frac{1}{n}\,\ell_n''(\theta) \to -\int_{\mathscr{X}} g'(x;\theta)^2\pi(\theta; \mathrm{d}x) = -i(\theta), \quad \text{say},$$

$$\frac{1}{n}\,\ell_n'''(\theta) \to -3\int_{\mathscr{X}} g'(x;\theta)g''(x;\theta)\pi(\theta; \mathrm{d}x) = \kappa_3(\theta), \quad \text{say},$$

and

$$\frac{1}{n}\,\ell_n^{(4)}(\theta) \to \; -3\int_{\mathscr{X}} g''(x;\theta)^2\pi(\theta; \mathrm{d}x) - 4\int_{\mathscr{X}} g'(x;\theta)g'''(x;\theta)\pi(\theta; \mathrm{d}x)$$
$$= \kappa_4(\theta), \quad \text{say},$$

in P_θ-probability for all $\theta \in \Omega$ as $n \to \infty$.

Define

$$Z_n = Z_n(\theta) = \sqrt{i_n}(\theta - \hat{\theta}_n),$$

where $i_n = -\ell_n''(\hat{\theta}_n)$ is the observed Fisher information and $\hat{\theta}_n = \hat{\theta}_n(y_1, \ldots, y_n)$ is the smallest maximum likelihood estimator of θ, so that $\ell_n(\hat{\theta}_n) = \sup_{\omega\in\Omega} \ell_n(\omega)$ and $\hat{\theta}_n$ is the smallest such value, if the maximum is attained at multiple points. Then Z_n is asymptotically standard normal as $n \to \infty$ and, so, may be treated as a first approximation to a pivotal quantity. The main aims of this work are to find a tilted version of Z_n, say $W_n = W_n(\theta) = \psi_n(Z_n)$, and data-dependent quantities $\hat{\mu}_n$ and $\hat{\sigma}_n$ such that

$$W_n^* = \frac{W_n - n^{-1/2}\hat{\mu}_n}{\hat{\sigma}_n} \tag{1}$$

is asymptotically as $n \to \infty$ standard normal to third order in the very weak sense of Woodroofe (1986).

In Section 2, first-order asymptotics are developed from which the form of the tilted approximation $W_n = \psi_n(Z_n)$ is deduced. Stein's identity is then used to obtain asymptotic expansions for posterior distributions in Section 3. In Section 4, very weak expansions for $E_\theta\{h(W_n)\}$ are presented and the approximately pivotal quantities W_n^* are constructed. Two examples of nonlinear models are described in Section 5 and the accuracy of the approximations presented in Section 4 are assessed by simulation in Section 6. Some remarks and an indication of extensions to the present work are given in Section 7. Proofs of the main results are detailed in Section 8.

2. First-order asymptotics

Consider the signed root transformation,

$$\tilde{W}_n = \sqrt{2\{\ell_n(\hat{\theta}_n) - \ell_n(\theta)\}}\,\mathrm{sign}(\theta - \hat{\theta}_n). \tag{2}$$

A Taylor series expansion yields

$$\ell_n(\theta) - \ell_n(\hat{\theta}_n) = -\frac{1}{2} Z_n^2 + \frac{\ell_{3,n}}{6 i_n^{3/2}} Z_n^3 + \frac{\ell_{4,n}}{24 i_n^2} Z_n^4 + \mathrm{o_p}\left(\frac{1}{n}\right),$$

where $\ell_{3,n} = \ell_n'''(\hat{\theta}_n)$ and $\ell_{4,n} = \ell_n^{(4)}(\hat{\theta}_n)$. A further expansion yields $\tilde{W}_n = W_n + \mathrm{o_p}(1/n)$, where

$$W_n = Z_n - \frac{\ell_{3,n}}{6 i_n^{3/2}} Z_n^2 - \frac{1}{72}\left(\frac{3\ell_{4,n}}{i_n^2} + \frac{\ell_{3,n}^2}{i_n^3}\right) Z_n^3 = \psi_n(Z_n), \tag{3}$$

say. It is then easily seen that

$$Z_n = W_n + \frac{\ell_{3,n}}{6 i_n^{3/2}} W_n^2 + \frac{1}{72}\left(\frac{3\ell_{4,n}}{i_n^2} + \frac{5\ell_{3,n}^2}{i_n^3}\right) W_n^3 + \mathrm{o_p}\left(\frac{1}{n}\right),$$

and that

$$\ell_n(\theta) - \ell_n(\hat{\theta}_n) = -\tfrac{1}{2} W_n^2 + \log\, R_n(\theta),$$

where $\log\, R_n(\theta) = \mathrm{o_p}(1/n)$. It follows that W_n is approximately standard normal for large n for fixed θ, so that it provides a first approximation to a pivotal quantity. The transformation ψ_n is defined by (3) for moderate values of Z_n only. For technical reasons, the construction is modified for larger values following Lemma 3 in Section 8.

3. Expansions for posterior distributions

The ultimate goal is to derive a confidence interval $\mathscr{I}_n$, say, whose coverage probability differs from a nominal value γ, say, by $\mathrm{o}(1/n)$, where n is the sample size. Letting $\gamma_n(\theta) = P_\theta(\theta \in \mathscr{I}_n)$, the requirement is then that $\gamma_n(\theta) = \gamma + \mathrm{o}(1/n)$ for all θ. This result is obtained in a weak form,

$$\int_\Omega \gamma_n(\theta)\xi(\theta)\,\mathrm{d}\theta = \gamma + \mathrm{o}\left(\frac{1}{n}\right) \tag{4}$$

for a large class of smooth prior densities. Of course, the left-hand side of (4) is just the unconditional probability that $\theta \in \mathscr{I}_n$ in a Bayesian model, and that may be computed by first conditioning on the data and then integrating with respect to their marginal distribution.

Thus, consider a Bayesian model in which θ has a prior density ξ on Ω. Let E_ξ denote expectation in the Bayesian model in which θ is replaced with a random variable Θ and let E_ξ^n denote conditional expectation given $\{x_k, y_k, k=1,\ldots,n\}$. Then

$$E_\xi\{h(W_n)\} = E_\xi[E_\xi^n\{h(W_n)\}].$$

The approach here is to obtain asymptotic expansions for the posterior expectations and then to integrate them with respect to the marginal distribution of the data.

If Θ has a density ξ, then its posterior density is

$$\xi_n(\theta) \propto \mathrm{e}^{\ell_n(\theta)-\ell_n(\hat{\theta}_n)}\xi(\theta) = \mathrm{e}^{-(1/2)W_n^2} R_n(\theta)\xi(\theta).$$

So, if $n \geqslant n_0$ and B_n, defined in (17) below, occurs, then the posterior density of W_n is

$$\zeta_n(w) \propto \frac{\xi(\theta)R_n(\theta)\phi(w)}{J_n(\theta)},$$

where

$$J_n(\theta) = \frac{\mathrm{d}W_n}{\mathrm{d}\theta} = \sqrt{i_n}\psi_n'(Z_n),$$

ϕ denotes the standard normal density, and θ and w are related by (3). Thus, the posterior density of W_n is of the form

$$\zeta_n(w) = f_n(w)\phi(w),$$

where

$$f_n(w) = \frac{\xi(\theta)R_n(\theta)}{c_n J_n(\theta)},$$

and c_n is a normalising constant. Further, simple differentiation yields

$$\frac{f_n'(w)}{f_n(w)} = \frac{1}{J_n(\theta)}\left\{\frac{\xi'(\theta)}{\xi(\theta)} + \frac{R_n'(\theta)}{R_n(\theta)} - \frac{J_n'(\theta)}{J_n(\theta)}\right\}$$

and

$$\frac{f_n''(w)}{f_n(w)} = \frac{1}{J_n(\theta)^2}\left\{\frac{\xi''(\theta)}{\xi(\theta)} + 2\frac{R_n'(\theta)}{R_n(\theta)}\frac{\xi'(\theta)}{\xi(\theta)} - 3\frac{J_n'(\theta)}{J_n(\theta)}\frac{\xi'(\theta)}{\xi(\theta)} - 3\frac{R_n'(\theta)}{R_n(\theta)}\frac{J_n'(\theta)}{J_n(\theta)} + \frac{R_n''(\theta)}{R_n(\theta)} - \frac{J_n''(\theta)}{J_n(\theta)} + 3\frac{J_n'(\theta)^2}{J_n(\theta)^2}\right\},$$

where f_n' is obtained by differentiation with respect to w, and ξ', R_n' and J_n' by differentiation with respect to θ.

A version of Stein's identity (Stein, 1986) may be applied to the posterior distribution of W_n in order to obtain asymptotic expansions. Let h be a function of polynomial growth and let

$$\Phi h = \int_{-\infty}^{\infty} h(w)\phi(w)\,\mathrm{d}w.$$

Then the Stein transformation is defined as

$$Uh(w) = \mathrm{e}^{(1/2)w^2} \int_w^\infty \{h(y) - \Phi h\} \mathrm{e}^{-(1/2)y^2} \, \mathrm{d}y.$$

This is a linear transformation. Further, letting U^2 denote the composition of U with itself, it is easily seen that

$$\Phi U h = \int_{-\infty}^{\infty} w h(w) \phi(w) \, \mathrm{d}w$$

and

$$\Phi U^2 h = \frac{1}{2} \int_{-\infty}^{\infty} (w^2 - 1) h(w) \phi(w) \, \mathrm{d}w.$$

For example, if $h_i(w) = w^i$ for $i = 1, 2$, then $Uh_1(w) = 1$ and $Uh_2(w) = w$ for all $-\infty < w < \infty$. The conditional expectation $E_\xi^n\{h(W_n)\}$ may be evaluated by applying Stein's identity. Let

$$\Gamma_{n,1}(\theta) = \sqrt{n}\, \frac{f_n'(w)}{f_n(w)} \quad \text{and} \quad \Gamma_{n,2}(\theta) = n\, \frac{f_n''(w)}{f_n(w)}, \tag{5}$$

where θ and w are related by (3) again. If B_n occurs and h is a measurable function of polynomial growth, then

$$\begin{aligned} E_\xi^n\{h(W_n)\} &= \Phi h + \frac{1}{\sqrt{n}} E_\xi^n\{Uh(W_n)\Gamma_{n,1}(\Theta)\} \\ &= \Phi h + \frac{1}{\sqrt{n}} \Phi U h E_\xi^n\{\Gamma_{n,1}(\Theta)\} + \frac{1}{n} E_\xi^n\{U^2 h(W_n)\Gamma_{n,2}(\Theta)\}. \end{aligned} \tag{6}$$

Similar calculations are detailed by Woodroofe (1992) and Coad and Woodroofe (1996) in the context of a one-parameter exponential family.

4. Very weak expansions

In this section, third-order very weak expansions for W_n are obtained, and an approximately pivotal quantity W_n^* is constructed by standardising. Let $\mathscr{H}_r^o$ be the class of measurable functions $h : \mathbb{R} \to \mathbb{R}$ for which $|h(z)| \leqslant 1 + |z|^r$ for all z. Further, let $\mathscr{H}_r$ be the set of h for which $h/c \in \mathscr{H}_r^o$ for some c, and let $\mathscr{H}_r^s$ be the set of symmetric $h \in \mathscr{H}_r$. Also set

$$\Gamma_1^\xi(\theta) = \frac{1}{\sqrt{i(\theta)}} \left\{ \frac{1}{3} \frac{\kappa_3(\theta)}{i(\theta)} + \frac{\xi'(\theta)}{\xi(\theta)} \right\} \tag{7}$$

and

$$\Gamma_2^\xi(\theta) = \frac{1}{i(\theta)} \left\{ \frac{\kappa_3(\theta)}{i(\theta)} \frac{\xi'(\theta)}{\xi(\theta)} + \frac{5}{12} \frac{\kappa_3(\theta)^2}{i(\theta)^2} + \frac{1}{4} \frac{\kappa_4(\theta)}{i(\theta)} + \frac{\xi''(\theta)}{\xi(\theta)} \right\}. \tag{8}$$

Theorem 1. *Suppose that the design is stable. If ξ is any twice continuously differentiable density for which*

$$\xi'(\underline{\theta}) = \xi(\underline{\theta}) = 0 = \xi(\bar{\theta}) = \xi'(\bar{\theta}), \tag{9}$$

then

$$E_\xi\{h(W_n)\} = \Phi h + \frac{\Phi U h}{\sqrt{n}} \int_\Omega \Gamma_1^\xi(\theta)\xi(\theta)\,\mathrm{d}\theta + \mathrm{o}\left(\frac{1}{\sqrt{n}}\right) \tag{10}$$

for all $h \in \mathscr{H}_2$ and

$$E_\xi\{h(W_n)\} = \Phi h + \frac{\Phi U^2 h}{n} \int_\Omega \Gamma_2^\xi(\theta)\xi(\theta)\,\mathrm{d}\theta + \mathrm{o}\left(\frac{1}{n}\right) \tag{11}$$

for each $h \in \mathscr{H}_2^s$. Moreover, (10) *and* (11) *hold uniformly with respect to $h \in \mathscr{H}_2^o$ and $h \in \mathscr{H}_2^o \cap \mathscr{H}_2^s$, respectively.*

The result may be obtained formally by observing that $\lim_{n\to\infty} \Gamma_{n,i}(\theta) = \Gamma_i^\xi(\theta)$ in P_θ-probability as $n \to \infty$ for all $\theta \in \Omega$ and $i = 1, 2$, and then formally exchanging limit and expectation. The details are presented in Section 8, and a stronger form of uniformity is established.

If $h(w) = w$, then $\Phi h = 0$, $\Phi U h = 1$ and $\Phi U^2 h = 0$. So, from (10),

$$E_\xi(W_n) \simeq \frac{1}{\sqrt{n}} \int_\Omega \Gamma_1^\xi(\theta)\xi(\theta)\,\mathrm{d}\theta.$$

If i is absolutely continuous, then the last expression may be integrated by parts, so that

$$\int_\Omega \Gamma_1^\xi(\theta)\xi(\theta)\,\mathrm{d}\theta = \int_\Omega \mu(\theta)\xi(\theta)\,\mathrm{d}\theta,$$

where

$$\mu(\theta) = \frac{1}{\sqrt{i(\theta)}} \left\{ \frac{1}{2}\,\frac{i'(\theta)}{i(\theta)} + \frac{1}{3}\,\frac{\kappa_3(\theta)}{i(\theta)} \right\}.$$

This leads to the approximation

$$E_\theta(W_n) \simeq \frac{1}{\sqrt{n}}\,\mu(\theta)$$

in the very weak sense of Woodroofe (1986, 1989). Next, let $\hat{\mu}_n = \mu(\hat{\theta}_n)$, so that

$$E_\xi^n\left\{\left(W_n - \frac{\hat{\mu}_n}{\sqrt{n}}\right)^2\right\} = E_\xi^n(W_n^2) - 2\,\frac{\hat{\mu}_n}{\sqrt{n}}\,E_\xi^n(W_n) + \frac{\hat{\mu}_n^2}{n}.$$

If both i and κ_3 are absolutely continuous, then applying (11) with $h(w) = w^2$ and another integration by parts, yields

$$E_\xi\left\{\left(W_n - \frac{\hat{\mu}_n}{\sqrt{n}}\right)^2\right\} = 1 + \frac{1}{n}\int_\Omega q(\theta)\xi(\theta)\,\mathrm{d}\theta + \mathrm{o}\left(\frac{1}{n}\right), \tag{12}$$

where

$$q(\theta)=\frac{1}{i(\theta)}\left\{\frac{1}{3}\,\frac{i'(\theta)}{i(\theta)}\,\frac{\kappa_3(\theta)}{i(\theta)}+\frac{7}{36}\,\frac{\kappa_3(\theta)^2}{i(\theta)^2}-\frac{1}{3}\,\frac{\kappa_3'(\theta)}{i(\theta)}+\frac{1}{4}\,\frac{\kappa_4(\theta)}{i(\theta)}\right\}+\mu(\theta)^2.$$

This leads to the very weak approximation

$$E_\theta\left\{\left(W_n-\frac{\hat{\mu}_n}{\sqrt{n}}\right)^2\right\}\simeq 1+\frac{q(\theta)}{n}=\sigma_n^2(\theta),$$

say. Now let $\hat{\sigma}_n^2=\sigma_n^2(\hat{\theta}_n)$ in (1). Then the main result of the paper may be stated:

Theorem 2. *Suppose that the design is stable and that both i and κ_3 have bounded piecewise continuous derivatives. If ξ is any twice continuously differentiable density for which* (9) *holds, then*

$$E_\xi\{h(W_n^*)\}=\Phi h+\mathrm{o}\left(\frac{1}{n}\right)$$

for all $h\in\mathscr{H}_2^s$ and uniformly with respect to $h\in\mathscr{H}_2^s\cap\mathscr{H}_2^o$.

The proof will be presented in Section 8.4. An alternative statement is that, with the above choices of $\hat{\mu}_n$ and $\hat{\sigma}_n$,

$$E_\theta\{h(W_n^*)\}=\Phi h+\mathrm{o}\left(\frac{1}{n}\right), \tag{13}$$

very weakly. Relation (13) may be used to construct approximate confidence intervals for θ. Given a desired confidence level $0<\gamma<1$, let

$$\mathscr{I}_n=\left\{\theta:|W_n^*(\theta)|\leqslant\Phi^{-1}\left(\frac{1+\gamma}{2}\right)\right\}, \tag{14}$$

where Φ denotes the standard normal distribution function. Then

$$P_\theta(\theta\in\mathscr{I}_n)=P_\theta\left\{|W_n^*|\leqslant\Phi^{-1}\left(\frac{1+\gamma}{2}\right)\right\}=\gamma+\mathrm{o}\left(\frac{1}{n}\right)$$

as $n\to\infty$, in the very weak sense. Thus, $\mathscr{I}_n$ is an approximate $100\gamma\%$ confidence interval for θ.

Remark 1. For the linear model with $g(x;\theta)=x\theta$, $i(\theta)=\int_{\mathscr{X}}x^2\pi(\theta;\mathrm{d}x)$, and $\kappa_3(\theta)=\kappa_4(\theta)=0$. Thus, the above formulae for $\mu(\theta)$ and $q(\theta)$ reduce to (16) and (17) of Woodroofe and Coad (1997) when $p=1$. The new results show that the non-linearity in the model contributes an extra second-order term and also several additional third-order terms.

Remark 2. For a non-adaptive design, $\pi(\theta;\,.)$ does not depend on θ and $i'(\theta)/2=-\kappa_3(\theta)/3$. It follows that $\mu(\theta)=0$, so that W_n is asymptotically standard normal to second order in the very weak sense of Woodroofe (1986). Also note that, in the non-adaptive case,

$$q(\theta)=\frac{1}{4i(\theta)^2}\left[\int_{\mathscr{X}} g''(x;\theta)^2\pi(\mathrm{d}x)-\frac{1}{i(\theta)}\left\{\int_{\mathscr{X}} g''(x;\theta)g'(x;\theta)\pi(\mathrm{d}x)\right\}^2\right].$$

5. Examples

Two well-known models will be considered. Recent accounts of these and similar adaptive nonlinear models are given by Ratkowsky (1983), Bates and Watts (1988), and Seber and Wild (1989). To a first approximation, $\hat{\theta}$ has variance $1/\{ng'(x;\theta)^2\}$ (e.g. Draper and Smith, 1998). For the locally D-optimal design, the support point x_{k+1} is chosen to maximise $g'(x;\hat{\theta}_k)^2$, or, equivalently, that of $|g'(x;\hat{\theta}_k)|$; see, for example, Silvey (1980), Atkinson and Donev (1992), and Pukelsheim (1993). Below, the locally D-optimal design for each model is described. Amusingly, the adaptive designs are simpler than optimal non-adaptive designs; see, for example, Dette et al. (2003).

Example 1 (*First-order growth or decay model*). This model has been studied by, for example, Box and Lucas (1959). Here,

$$g(x;\theta)=\theta_1\exp(\theta_2 x),$$

where g is the amount of substance present, $x>0$ is time, $\theta_1>0$ is the initial amount of substance present, and θ_2 is the growth rate if $\theta_2>0$ and the decay rate if $\theta_2<0$. Suppose that θ_1 is known. Then the maximum likelihood estimator of θ_2 satisfies

$$\sum_{k=1}^{n}(y_k-\theta_1\mathrm{e}^{\hat{\theta}_{n,2}x_k})x_k\mathrm{e}^{\hat{\theta}_{n,2}x_k}=0$$

and the above criterion specifies that observation $k+1$ should be taken at $x_{k+1}=-1/\hat{\theta}_{k,2}$ for $k=1,2,\ldots$. It follows easily that

$$i(\theta)=\frac{\mathrm{e}^{-2}\theta_1^2}{\theta_2^2},\qquad \kappa_3(\theta)=\frac{3\mathrm{e}^{-2}\theta_1^2}{\theta_2^3}$$

and

$$\kappa_4(\theta)=\frac{7\mathrm{e}^{-2}\theta_1^2}{\theta_2^4}.$$

Thus, we have that $\mu(\theta)=0$ and

$$\sigma_n^2(\theta)=1+\frac{\mathrm{e}^2}{n\theta_1^2}.$$

Example 2 (*Michaelis–Menten model*). This model is of importance in enzyme kinetics; see, for example, Currie (1982). Here,

$$g(x;\theta)=\frac{\theta_1 x}{\theta_2+x},$$

where g is the reaction velocity, $x>0$ is the substrate concentration, $\theta_1>0$ is the maximum velocity of the reaction and $\theta_2>0$ is the half-saturation constant. As in Example 1, suppose that θ_1 is known. Then, here, the maximum likelihood estimator of θ_2 satisfies

$$\sum_{k=1}^{n}\left(y_k-\frac{\theta_1 x_k}{\hat\theta_{n,2}+x_k}\right)\frac{x_k}{(\hat\theta_{n,2}+x_k)^2}=0$$

and observation $k+1$ is taken at $x_{k+1}=\hat\theta_{k,2}$ for $k=1,2,\ldots$. It follows easily that

$$i(\theta)=\frac{\theta_1^2}{16\theta_2^2},\qquad \kappa_3(\theta)=\frac{3\theta_1^2}{16\theta_2^3}$$

and

$$\kappa_4(\theta)=-\frac{9\theta_1^2}{16\theta_2^4}.$$

Thus, we have that $\mu(\theta)=0$ and

$$\sigma_n^2(\theta)=1+\frac{8}{n\theta_1^2}.$$

For both models, stability of the designs follows from consistency of the maximum likelihood estimator, which is established in Section 8.1, and the measure $\pi(\theta;.)$ is degenerate at $\arg\max_x|g'(x;\theta)|$. For each model, interest lies in constructing corrected confidence intervals for θ_2. If c is the upper $100(1+\gamma)/2$ percentile of the standard normal distribution, then, from (3) and (14), an approximate $100\gamma\%$ confidence interval for θ_2 is

$$\mathscr{I}_n=\left[\theta_2:\left|(\theta_2-\hat\theta_{n,2})\left\{1-\frac{\ell_{3,n}}{6i_n}(\theta_2-\hat\theta_{n,2})-\frac{1}{72}\left(\frac{3\ell_{4,n}}{i_n}+\frac{\ell_{3,n}^2}{i_n^2}\right)(\theta_2-\hat\theta_{n,2})^2\right\}\right.\right.$$
$$\left.\left.-\frac{\hat\mu_n}{\sqrt{n}}\right|\leqslant c\,\frac{\hat\sigma_n}{\sqrt{i_n}}\right].$$

6. Simulation results

6.1. General

In order to assess the accuracy of the approximations in Section 5, a simulation study based on 10,000 replications was carried out, for selected values of the design parameters. The results are reported separately for the first-order growth or decay model and the

Table 1
Monte Carlo estimates of coverage probabilities for the first-order growth or decay model when $n = 25$

θ_1	θ_2	Z_n		W_n		W_n^*	
		$\gamma = 0.95$	$\gamma = 0.90$	$\gamma = 0.95$	$\gamma = 0.90$	$\gamma = 0.95$	$\gamma = 0.90$
15	0.1	0.947	0.894	0.948	0.896	0.948	0.896
	0.5	0.947	0.894	0.948	0.896	0.948	0.896
20	0.1	0.947	0.895	0.948	0.896	0.948	0.896
	0.5	0.947	0.895	0.948	0.896	0.948	0.896
25	0.1	0.947	0.894	0.948	0.896	0.948	0.896
	0.5	0.947	0.894	0.948	0.896	0.948	0.896

Michaelis–Menten model. In both cases, results are reported in detail when $n = 25$ and the Newton–Raphson method is used to find $\hat{\theta}_{k,2}$ for $k = 1, 2, \ldots$. Various initial estimates were used for θ_2. For comparison purposes, results are presented for the confidence intervals constructed using each of the approximately pivotal quantities Z_n, W_n and W_n^*. This will enable us to assess the increased accuracy of using a tilted approximation and the corrected version of this.

6.2. First-order growth or decay model

Monte Carlo estimates of the coverage probabilities for the first-order growth or decay model are presented in Table 1. These indicate that use of Z_n already works well, though the coverage probabilities are always less than the nominal values. Although a slight improvement is possible by using the tilted version of Z_n, no further improvement is achieved by using W_n^*. However, one explanation for the latter behaviour is that the mean correction is zero in this example.

6.3. Michaelis–Menten model

Monte Carlo estimates of the coverage probabilities for the Michaelis–Menten model are presented in Table 2. These exhibit a similar behaviour to those in Table 1. Note that the mean correction is also zero in this example.

7. Remarks

The main purpose of this paper was to extend existing work on corrected confidence intervals following adaptive linear regression models to the nonlinear case. The non-linearity in the model lead to a number of extra complications and the main finding was that a tilted version of Z_n is nearly normal in this case. Given the simulation results in Section 6, it would be interesting to consider an example in which $\mu(\theta)$ is non-zero.

We have considered a one-parameter adaptive nonlinear regression model in this paper. In follow-up work, we extend the ideas in the present paper to a two-parameter non-linear

Table 2
Monte Carlo estimates of coverage probabilities for the Michaelis–Menten model when $n = 25$

θ_1	θ_2	Z_n		W_n		W_n^*	
		$\gamma = 0.95$	$\gamma = 0.90$	$\gamma = 0.95$	$\gamma = 0.90$	$\gamma = 0.95$	$\gamma = 0.90$
15	0.1	0.939	0.890	0.940	0.890	0.941	0.890
	0.5	0.944	0.895	0.946	0.895	0.946	0.895
20	0.1	0.946	0.893	0.947	0.895	0.947	0.895
	0.5	0.947	0.894	0.948	0.896	0.948	0.896
25	0.1	0.947	0.894	0.948	0.896	0.948	0.896
	0.5	0.947	0.894	0.948	0.896	0.948	0.896

model and apply the results to a number of other well-known models. We are currently attempting to extend our results to higher dimensions and generalised linear models, but we have encountered several technical difficulties which we have not yet resolved.

One issue that we have not addressed is that of unknown variability. In this case, the nonlinear regression model becomes $y_k = g(x_k; \theta) + \sigma\varepsilon_k$ for $k = 1, 2, \ldots$, where $\sigma \in \Sigma$ is another unknown parameter and Σ is another compact interval. It is clear that the maximum likelihood estimator of σ^2 is $\hat{\sigma}_n^2 = \sum_{k=1}^{n}\{y_k - g(x_k; \hat{\theta}_n)\}^2/n$. Although the forms of Z_n and W_n will remain the same, the correction terms will now be more complicated.

Subtracting $\hat{\mu}_n/\sqrt{n}$ in (12) has the flavour of a bias correction, but is applied to the approximate pivot W_n instead of the estimator. Approximate biases for adaptive designs and linear models are considered in some detail by Coad and Woodroofe (1998). Diaconis and Zabell (1991) have used higher-order versions of Stein's identity in quite a different way and with different motivation.

8. Proofs

8.1. Consistency of maximum likelihood estimators

The following basic inequality given in Lemma 1 of Woodroofe (1991) is used repeatedly below.

Basic Inequality 1. *Let $U_k = u_k(y_1, \ldots, y_{k-1})$ be bounded measurable functions, say $|U_k| \leqslant C$ for all k, and let*

$$M_n = \sum_{k=1}^{n} U_k \varepsilon_k.$$

Then

$$E_\theta(\mathrm{e}^{tM_n}) \leqslant \mathrm{e}^{(1/2)nC^2t^2}$$

for all $t \in \mathbb{R}$, $\theta \in \Omega$ and $n \geqslant 1$.

The Basic Inequality may be seen by conditioning on $y_1, \ldots, y_{n-1}$, which leads to

$$E_\theta(\mathrm{e}^{tM_n}) = E_\theta(\mathrm{e}^{tM_{n-1}+(1/2)U_n^2t^2}) \leqslant \mathrm{e}^{(1/2)C^2t^2} E_\theta(\mathrm{e}^{tM_{n-1}}),$$

and using an induction argument. Alternatively, the inequality can be established by noting that $\exp(tM_n - t^2\sum_{k=1}^n U_k^2/2)$ is a martingale for any t. It then follows from the Basic Inequality and Bernstein's Inequality that

$$P_\theta(|M_n| \geqslant n\varepsilon) \leqslant 2\mathrm{e}^{-n\varepsilon^2/(2C^2)}$$

for all $\varepsilon > 0$.

Now let $\kappa_{j,n}(\theta, \omega)$ be the expression obtained when y_k is replaced by $g(x_k; \theta)$ in $\ell_n^{(j)}(\omega)/n$, so that

$$\kappa_{0,n}(\theta, \omega) = -\frac{1}{2n}\sum_{k=1}^n \{g(x_k; \theta) - g(x_k; \omega)\}^2,$$

$$\kappa_{1,n}(\theta, \omega) = \frac{1}{n}\sum_{k=1}^n \{g(x_k; \theta) - g(x_k; \omega)\}g'(x_k; \omega),$$

$$\kappa_{2,n}(\theta, \omega) = \frac{1}{n}\sum_{k=1}^n [\{g(x_k; \theta) - g(x_k; \omega)\}g''(x_k; \omega) - g'(x_k; \omega)^2],$$

$$\kappa_{3,n}(\theta, \omega) = \frac{1}{n}\sum_{k=1}^n [\{g(x_k; \theta) - g(x_k; \omega)\}g'''(x_k; \omega) - 3g'(x_k; \omega)g''(x_k; \omega)],$$

$$\begin{aligned}\kappa_{4,n}(\theta, \omega) = \frac{1}{n}\sum_{k=1}^n [&\{g(x_k; \theta) - g(x_k; \omega)\}g^{(4)}(x_k; \omega) - 3g''(x_k; \omega)^2 \\ &- 4g'(x_k; \omega)g'''(x_k; \omega)],\end{aligned}$$

and let

$$i_n(\theta) = -n\kappa_{2,n}(\theta, \theta) = \sum_{k=1}^n g'(x_k; \theta)^2.$$

Lemma 1. *For every* $\varepsilon > 0$ *and every* $\alpha > 0$,

$$\sup_{\theta\in\Omega} P_\theta\left\{\sup_{\omega\in\Omega} |\ell_n(\omega) - \ell_n(\theta) - n\kappa_{0,n}(\theta, \omega)| > \varepsilon n\right\} = \mathrm{o}(n^{-\alpha})$$

and

$$\sup_{\theta\in\Omega} P_\theta\left\{\sup_{\omega\in\Omega} |\ell_n^{(j)}(\omega) - n\kappa_{j,n}(\theta, \omega)| > \varepsilon n\right\} = \mathrm{o}(n^{-\alpha})$$

for $j = 1, 2, 3, 4, 5$ *as* $n \to \infty$.

Proof. For the first assertion, let $\Delta_n(\theta,\omega)=\ell_n(\omega)-\ell_n(\theta)-n\kappa_{0,n}(\theta,\omega)$. Then

$$\Delta_n(\theta,\omega)=\sum_{k=1}^{n}\{g(x_k;\omega)-g(x_k;\theta)\}\varepsilon_k$$

is of the form considered in the Basic Inequality for fixed θ and ω, and

$$|\Delta_n(\theta,\omega_2)-\Delta_n(\theta,\omega_1)|\leqslant C|\omega_2-\omega_1|\sum_{k=1}^{n}|\varepsilon_k|, \tag{15}$$

where C is an upper bound for $|g'(x;\theta)|$. Given $\varepsilon>0$, let $\delta=\varepsilon/(2C)$ and let $\underline{\theta}=\omega_0<\omega_1<\cdots<\omega_m=\bar{\theta}$ be equally spaced points for which $\omega_i-\omega_{i-1}\leqslant\delta$ for $i=1,\ldots,m$. If $|\Delta_n(\theta,\omega_i)|\leqslant n\varepsilon/2$ for $i=1,\ldots,m$ and $\sum_{k=1}^{n}|\varepsilon_k|\leqslant n$, then $\sup_{\omega\in\Omega}|\Delta_n(\theta,\omega)|\leqslant n\varepsilon$ by (15). So,

$$P_\theta\left\{\sup_{\omega\in\Omega}|\Delta_n(\theta,\omega)|\geqslant n\varepsilon\right\}\leqslant\sum_{i=1}^{m}P_\theta\left\{|\Delta_n(\theta,\omega_i)|>\frac{1}{2}n\varepsilon\right\}+P_\theta\left(\sum_{k=1}^{n}|\varepsilon_k|>n\right).$$

Here each term in the sum is $o(n^{-\alpha})$ uniformly in θ for any $\alpha>0$, by the Basic Inequality. Moreover, $P_\theta(\sum_{k=1}^{n}|\varepsilon_k|>n)$ does not depend on θ and is $o(n^{-\alpha})$ for any α, since the $|\varepsilon_k|$ have moments of all orders. The first assertion follows, and the others may be established similarly. □

Lemma 2. *There is a positive constant γ for which*

$$\sup_{\theta\in\Omega}P_\theta(|\hat{\theta}_n-\theta|\geqslant\varepsilon)\leqslant 2e^{-(1/2)n\gamma\varepsilon^2}+o(n^{-\alpha}) \tag{16}$$

for all $\varepsilon>0$ and $\alpha>0$.

Proof. Recall that $g'(x;\theta)>0$ for all x and θ. So, by compactness and continuity, there are $\delta_0>0$ and $\varepsilon_0>0$ for which $\{g(x;\omega)-g(x;\theta)\}^2>4\varepsilon_0$ whenever $|\omega-\theta|\geqslant\delta_0$ and $g'(x;\omega)^2-g''(x;\omega)\{g(x;\omega)-g(x;\theta)\}\geqslant 2\varepsilon_0$ whenever $|\omega-\theta|\leqslant\delta_0$. Then we have $|\kappa_{0,n}(\theta,\omega)|\geqslant 2\varepsilon_0$ whenever $|\omega-\theta|\geqslant\delta_0$ and $|\kappa_{2,n}(\theta,\omega)|\geqslant 2\varepsilon_0$ whenever $|\omega-\theta|\leqslant\delta_0$. For fixed θ, let $A_n=A_{n,\theta}$ be the event

$$\begin{aligned}A_n=&\left\{\sup_{\omega}|\ell_n(\omega)-\ell_n(\theta)-n\kappa_{0,n}(\theta,\omega)|\leqslant n\varepsilon_0\right\}\\&\cap\left\{\sup_{\omega\in\Omega}|\ell_n''(\omega)-n\kappa_{2,n}(\theta,\omega)|\leqslant n\varepsilon_0\right\}.\end{aligned}$$

Then

$$\sup_{\theta}P_\theta(A_n^c)=o(n^{-\alpha})$$

for all $\alpha>0$ by Lemma 1. Clearly, A_n implies that $\ell_n''(\omega)\leqslant -n\varepsilon_0$ for all $|\omega-\theta|\leqslant\delta_0$. It also implies that $|\hat{\theta}_n-\theta|<\delta_0$. For, if $|\hat{\theta}_n-\theta|\geqslant\delta_0$, then $0\leqslant\ell_n(\omega)-\ell_n(\theta)<-2n\varepsilon_0+\{\ell_n(\omega)-$

$\ell_n(\theta) - n\kappa_{0,n}(\theta,\omega)\}$ for some $|\omega - \theta| \geqslant \delta_0$, in which case A_n^c occurs. This establishes (16) for $\varepsilon \geqslant \delta_0$. For $0 < \varepsilon < \delta_0$, A_n and $\hat{\theta}_n - \theta > \varepsilon$ imply that $\ell_n'(\theta + \varepsilon) > 0$. Now,

$$\ell_n'(\omega) = \sum_{k=1}^{n} g'(x_k;\omega)\varepsilon_k - \sum_{k=1}^{n} g'(x_k;\omega)\{g(x_k;\omega) - g(x_k;\theta)\}.$$

The first sum is of the form considered in the Basic Inequality. When $\omega = \theta + \varepsilon$, the second is at least $n\gamma'\varepsilon$ for some positive γ'. That there is a γ for which

$$P_\theta[A_n \cap \{\ell_n'(\theta+\varepsilon) > 0\}] \leqslant \mathrm{e}^{-(1/2)n\gamma\varepsilon^2}$$

for all θ and $0 < \varepsilon < \delta_0$ now follows from the Basic Inequality, and $P_\theta(\hat{\theta}_n - \theta < -\varepsilon)$ may be handled similarly to complete the proof. $\square$

Remark 3. The proof of Lemma 2 shows that $P_\theta(A_n, |\hat{\theta}_n - \theta| > \varepsilon) \leqslant 2\exp(-\frac{1}{2}n\gamma\varepsilon^2)$.

8.2. Towards the expansions: more lemmas

Clearly, $i_n(\theta) \geqslant 2n\varepsilon_0$ for all $\theta \in \Omega$, since $\kappa_{2,n}(\theta,\omega) \geqslant 2\varepsilon_0$ when $|\theta - \omega| \leqslant \delta_0$. Moreover, $\kappa_{j,n}$ are uniformly bounded, say $|\kappa_{j,n}(\theta,\omega)| \leqslant C_1$ for all $\theta, \omega \in \Omega$, $n = 1, 2, \ldots$ and $j = 2, 3, 4, 5$. Now let $\ell_{j,n} = \ell_n^{(j)}(\hat{\theta}_n)$ and let B_n be the event

$$B_n = \{\underline{\theta} < \hat{\theta}_n < \bar{\theta}, i_n \geqslant n\varepsilon_0\} \cap \{|\ell_{j,n}| \leqslant (C_1+1)n,\ j = 2, 3, 4, 5\}. \tag{17}$$

Lemma 3. *If ξ is a twice continuously differentiable density on Ω for which* (9) *holds, then*

$$P_\xi(B_n^c) = \mathrm{O}(n^{-3/2})$$

and

$$\int_\Omega P_\theta(B_n^c)|\xi'(\theta)|\,\mathrm{d}\theta = \mathrm{O}\left(\frac{1}{n}\right).$$

Proof. By Lemma 1,

$$P_\theta(B_n^c) \leqslant P_\theta[A_n \cap \{\hat{\theta}_n = \underline{\theta} \text{ or } \bar{\theta}\}] + \frac{1}{n^2}$$

for all $\theta \in \Omega$ and large n. So, for large n,

$$\begin{aligned} n^{3/2}P_\xi(B_n^c) &\leqslant 2n^{3/2}\int_\Omega \{\mathrm{e}^{-(1/2)n\gamma(\theta-\underline{\theta})^2} + \mathrm{e}^{-(1/2)n\gamma(\theta-\bar{\theta})^2}\}\xi(\theta)\,\mathrm{d}\theta + \frac{2}{\sqrt{n}} \\ &\leqslant 2\int_0^\infty n\left\{\xi\left(\underline{\theta} + \frac{z}{\sqrt{n}}\right) + \xi\left(\bar{\theta} - \frac{z}{\sqrt{n}}\right)\right\}\mathrm{e}^{-(1/2)\gamma z^2}\,\mathrm{d}z + \frac{2}{\sqrt{n}} \\ &\to \{\xi''(\underline{\theta}) + \xi''(\bar{\theta})\}\int_0^\infty z^2\mathrm{e}^{-(1/2)\gamma z^2}\,\mathrm{d}z \end{aligned}$$

by Remark 3 and the Dominated Convergence Theorem. This establishes the first assertion and the second may be established similarly. $\square$

Let $\alpha_n = \log^2 n$ and let $\psi_n(z)$ be a three times continuously differentiable function for which

$$\psi_n(z) = z - \frac{\ell_{3,n}}{6i_n^{3/2}} z^2 - \frac{1}{72}\left(\frac{3\ell_{4,n}}{i_n^2} + \frac{\ell_{3,n}^2}{i_n^3}\right) z^3$$

for $|z| \leqslant \alpha_n$, and

$$\psi_n''(z) = \begin{cases} \{\psi_n''(-\alpha_n) + \psi_n'''(-\alpha_n)(z-\alpha_n)\}\mathrm{e}^{-(1/2)(z-\alpha_n)^2} & \text{if } z \leqslant -\alpha_n, \\ \{\psi_n''(\alpha_n) + \psi_n'''(\alpha_n)(z-\alpha_n)\}\mathrm{e}^{-(1/2)(z-\alpha_n)^2} & \text{if } z \geqslant \alpha_n. \end{cases}$$

Then

$$\psi_n'(z) = 1 - \frac{\ell_{3,n}}{3i_n^{3/2}} z - \frac{1}{24}\left(\frac{3\ell_{4,n}}{i_n^2} + \frac{\ell_{3,n}^2}{i_n^3}\right) z^2,$$

$$\psi_n''(z) = -\frac{\ell_{3,n}}{3i_n^{3/2}} - \frac{1}{12}\left(\frac{3\ell_{4,n}}{i_n^2} + \frac{\ell_{3,n}^2}{i_n^3}\right) z$$

and

$$\psi_n'''(z) = -\frac{1}{12}\left(\frac{3\ell_{4,n}}{i_n^2} + \frac{\ell_{3,n}^2}{i_n^3}\right)$$

for $|z| \leqslant \alpha_n$. It then follows that there are a constant C_2 and an integer n_0 for which

$$\frac{1}{2} \leqslant \psi_n'(z) \leqslant 2, \quad |\psi_n''(z)| \leqslant \frac{C_2}{\sqrt{n}}, \quad |\psi_n'''(z)| \leqslant \frac{C}{n} \tag{18}$$

for all z whenever $n \geqslant n_0$ and B_n occurs.

Remark 4. It follows from Lemma 2 that $\sup_n E_\xi(|Z_n|^r 1_{B_n}) < \infty$ for any ξ, and then from (18) that $\sup_n E_\xi(|W_n|^r 1_{B_n}) < \infty$ for any ξ.

Recall that $J_n(\theta) = \sqrt{i_n}\psi'(Z_n)$, so that $J_n'(\theta) = i_n \psi_n''(Z_n)$ and $J_n''(\theta) = i_n^{3/2}\psi_n'''(Z_n)$.

Lemma 4. *If the design is stable, then*

$$\frac{J_n(\theta)}{\sqrt{n}} \to \sqrt{i(\theta)}, \qquad \frac{J_n'(\theta)}{J_n(\theta)} \to -\frac{\kappa_3(\theta)}{3i(\theta)}$$

and

$$\frac{J_n''(\theta)}{J_n(\theta)} \to -\frac{1}{12}\left\{\frac{3\kappa_4(\theta)}{i(\theta)} + \frac{\kappa_3(\theta)^2}{i(\theta)^2}\right\}$$

in P_θ-probability for all $\theta \in \Omega$ as $n \to \infty$. Further, there is a constant C for which

$$\left\{\left|\frac{J_n'(\theta)}{J_n(\theta)}\right| + \left|\frac{J_n''(\theta)}{J_n(\theta)}\right|\right\} 1_{B_n} \leqslant C$$

for all sufficiently large n.

Proof. By Lemmas 1 and 2, $i_n/n \to i(\theta)$, $\ell_{3,n}/n \to \kappa_3(\theta)$ and $\ell_{4,n}/n \to \kappa_4(\theta)$ in P_θ-probability as $n \to \infty$, and Z_n is stochastically bounded in P_θ-probability for each $\theta \in \Omega$. So, using Lemma 1 again, $\psi_n'(Z_n) \to 1$, $\sqrt{i_n}\psi_n''(Z_n) \to -\kappa_3(\theta)/\{3i(\theta)\}$ and $i_n\psi_n'''(Z_n) \to \{3\kappa_4(\theta)/i(\theta) + \kappa_3(\theta)^2/i(\theta)^2\}/12$ in P_θ-probability as $n \to \infty$. The first three assertions of the lemma follow directly, and the fourth is a direct consequence of (18). □

Lemma 5. *Suppose that the design is stable and let ξ be as in Lemma* 3. *Then*

$$\lim_{n\to\infty} \int_{B_n} \left\{ n\left|\frac{R_n'(\theta)}{R_n(\theta)}\right|^2 + \left|\frac{R_n''(\theta)}{R_n(\theta)}\right|^2 \right\} \mathrm{d}P_\xi = 0.$$

Proof. Recalling that $\log\, R_n(\theta) = \ell_n(\theta) - \ell_n(\hat\theta_n) + W_n^2/2$,

$$\frac{R_n'(\theta)}{R_n(\theta)} = \ell_n'(\theta) + \sqrt{i_n}\psi_n'(Z_n)W_n.$$

If B_n occurs, $\ell_n'(\theta)$ is expanded in a Taylor series about $\hat\theta_n$, and $\psi_n'(Z_n)W_n$ is expanded as a polynomial in $(\theta - \hat\theta_n)$, the linear, quadratic and cubic terms all vanish. From the definition of B_n, it then follows that there is a constant C for which

$$\left|\frac{R_n'(\theta)}{R_n(\theta)}\right| \mathbf{1}_{B_n} \leqslant \begin{cases} C\alpha_n^5/n & \text{if } |Z_n| \leqslant \alpha_n, \\ Cn & \text{if } |Z_n| \geqslant \alpha_n. \end{cases}$$

That

$$\lim_{n\to\infty} \int_{B_n} n\left|\frac{R_n'(\theta)}{R_n(\theta)}\right|^2 \mathrm{d}P_\xi = 0$$

follows, using Lemma 2. The term involving R_n'' is simpler analytically, if more complicated algebraically. □

Recall the definitions (5), (7) and (8) of $\Gamma_{n,1}(\theta)$, $\Gamma_{n,2}(\theta)$, $\Gamma_1^\xi(\theta)$ and $\Gamma_2^\xi(\theta)$.

Lemma 6. *If the design is stable and ξ is as in Lemma* 3, *then*

$$\lim_{n\to\infty} \int_{B_n} (|\Gamma_{n,1} - \Gamma_1^\xi| + |\Gamma_{n,2} - \Gamma_2^\xi|)\, \mathrm{d}P_\xi = 0.$$

Proof. This follows directly from Lemmas 4 and 5. To see how, observe that

$$|\Gamma_{n,1} - \Gamma_1^\xi| \leqslant \left|\frac{\sqrt{n}}{J_n} - \frac{1}{\sqrt{i}}\right| \left|\frac{\xi'}{\xi}\right| + \frac{\sqrt{n}}{J_n}\left|\frac{R_n'}{R_n}\right| + \left|\frac{\sqrt{n}J_n'}{J_n^2} - \frac{\kappa_3}{3i^{3/2}}\right|.$$

By Lemmas 4 and 5, the right-hand side converges to zero in P_ξ-probability and is bounded on B_n by a constant multiple of $1 + |\xi'/\xi| + |R_n'/R_n|$, which is uniformly integrable. The assertions concerning $\Gamma_{n,1}$ now follow from the Dominated Convergence Theorem, as extended in Problem 16.4 of Billingsley (1985), and $\Gamma_{n,2}$ may be analysed similarly. □

Recall that $\mathscr{H}_r^o$ is the class of measurable functions $h:\ \mathbb{R} \to \mathbb{R}$ for which $|h(z)| \leqslant 1+|z|^r$ for all z, and let

$$\Delta_{i,n} = \operatorname*{essup}_{h\in\mathscr{H}_i^o} |E_\xi^n\{h(W_n)\} - \Phi h|$$

for $i = 0, 1$, and observe that $\Delta_{0,n} \leqslant 2\min(2, \Delta_{1,n})$.

Lemma 7. *If the design is stable, then* $\lim_{n\to\infty} E_\xi(\Delta_{1,n} 1_{B_n}) = 0$.

Proof. If $h \in \mathscr{H}_1^o$, then $|Uh| \leqslant 4$, by Lemma 1 of Woodroofe (1992). So, by Stein's identity,

$$|E_\xi^n\{h(W_n)\} - \Phi h| 1_{B_n} = \frac{1}{\sqrt{n}} E_\xi^n\{|Uh(W_n)\Gamma_{n,1}|\} 1_{B_n} \leqslant \frac{4}{\sqrt{n}} E_\xi^n(|\Gamma_{n,1}|) 1_{B_n},$$

which is independent of h and approaches zero in the mean as $n \to \infty$, by Lemma 6. □

8.3. Proof of Theorem 1

It suffices to establish the theorem for $h \in \mathscr{H}_2^o$. If $h \in \mathscr{H}_2^o$, then

$$\int_{B_n^c} |h(W_n)|\, \mathrm{d}P_\xi \leqslant P_\xi(B_n)^{3/4} [E_\xi\{(1+W_n^2)^4\}]^{1/4} = \mathrm{o}\left(\frac{1}{n}\right),$$

$$E_\xi\{h(W_n)\} = \int_{B_n} h(W_n)\, \mathrm{d}P_\xi + \mathrm{o}\left(\frac{1}{n}\right),$$

$$\int_{B_n} h(W_n)\, \mathrm{d}P_\xi = \int_{B_n} E_\xi^n\{h(W_n)\}\, \mathrm{d}P_\xi$$

and

$$E_\xi^n\{h(W_n)\} = \Phi h + \frac{1}{\sqrt{n}} \Phi U h I_n + \frac{1}{n} II_n(h),$$

where

$$I_n = E_\xi^n\{\Gamma_{n,1}(\Theta)\}$$

and

$$II_n(h) = E_\xi^n\{U^2 h(W_n) \Gamma_{n,2}(\Theta)\}$$

on B_n, as in (17). It is clear from Lemma 6 that

$$\lim_{n\to\infty} \int_{B_n} I_n\, \mathrm{d}P_\xi = E_\xi(\Gamma_1^\xi) = \int_\Omega \Gamma_1^\xi(\theta)\xi(\theta)\, \mathrm{d}\theta.$$

So, it suffices to show that

$$\lim_{n\to\infty} \int_{B_n} \operatorname*{essup}_{h\in\mathscr{H}_2^o} |II_n(h) - (\Phi U^2 h) E_\xi(\Gamma_2^\xi)|\, \mathrm{d}P_\xi = 0. \tag{19}$$

For this, write

$$II_n(h) = II_{n,1}(h) + II_{n,2}(h) + II_{n,3}(h),$$

where

$$II_{n,1}(h) = \Phi U^2 h \bar{\Gamma}_{n,2},$$
$$II_{n,2}(h) = E_\xi^n[\{U^2 h(W_n) - \Phi U^2 h\}\bar{\Gamma}_{n,2}],$$
$$II_{n,3}(h) = E_\xi^n[U^2 h(W_n)\{\Gamma_{n,2}(\theta) - \bar{\Gamma}_{n,2}\}]$$

and

$$\bar{\Gamma}_{n,2} = E_\xi^n\{\Gamma_{n,2}(\Theta)\}.$$

From Lemma 6, it is clear that

$$\int_{B_n} \operatorname*{essup}_{h\in\mathscr{H}_2^o} |II_{n,1}(h) - (\Phi U^2 h)E_\xi(\Gamma_2^\xi)|\,\mathrm{d}P_\xi \leqslant C\int_{B_n} |\Gamma_{n,2} - \Gamma_2^\xi|\,\mathrm{d}P_\xi \to 0$$

and

$$\int_{B_n} \operatorname*{essup}_{h\in\mathscr{H}_2^o} |II_{n,3}(h)|\,\mathrm{d}P_\xi \leqslant C\int_{B_n} |\Gamma_{n,2} - \bar{\Gamma}_{n,2}|\,\mathrm{d}P_\xi \to 0$$

as $n \to \infty$, where C denotes an upper bound for $|U^2 h|$. Similarly, using Lemma 7,

$$\int_{B_n} \operatorname*{essup}_{h\in\mathscr{H}_2^o} |II_{n,2}(h)|\,\mathrm{d}P_\xi \leqslant C\int_{B_n} |\Delta_{0,n}||\bar{\Gamma}_{n,2}|\,\mathrm{d}P_\xi \to 0$$

as $n \to \infty$. Relation (19) follows and provides more uniformity in h than was asserted in the theorem. □

8.4. Proof of Theorem 2

One additional lemma is needed for the proof of Theorem 2. If h is a function of polynomial growth, $v \in \mathbb{R}$ and $\sigma > 0$, let

$$h^*(w) = h_{v,\sigma}(w) = h\left(\frac{w - v}{\sigma}\right).$$

Then the following lemma may be obtained by specialising Lemma 1 of Woodroofe and Coad (1997).

Lemma 8. *There is a constant C for which*

$$|\Phi h^* - [\Phi h + (\Phi U^2 h)\{v^2 - (\sigma^2 - 1)\}]| \leqslant C(|v|^3 + |\sigma^2 - 1|^{3/2}),$$

$$|\Phi U h^* + 2(\Phi U^2 h)v| \leqslant C(|v|^2 + |\sigma^2 - 1|)$$

and

$$|\Phi U^2 h^* - \Phi U^2 h| \leqslant C(|v| + |\sigma^2 - 1|)$$

for all $h \in \mathscr{H}_2^s \cap \mathscr{H}_2^o$, $|v| \leqslant 1$, *and* $|\sigma^2 - 1| \leqslant \frac{1}{2}$.

For the proof of Theorem 2, let $v = \hat{\mu}_n/\sqrt{n}$ and $\sigma^2 = 1 + \hat{q}_n/n$ in the definition of h^*, where $\hat{q}_n = q(\hat{\theta}_n)$. Then

$$E_\xi\{h(W_n^*)\} = E_\xi\{h^*(W_n)\} = \int_{B_n} E_\xi^n\{h^*(W_n)\}\,\mathrm{d}P_\xi + \mathrm{o}\left(\frac{1}{n}\right)$$

uniformly with respect to $h \in \mathscr{H}_2^o$, as in the proof of Theorem 1. If $h \in \mathscr{H}_2^o \cap \mathscr{H}_2^s$, then $\Phi U h^*$ is of order $1/\sqrt{n}$, and, from the proof of Theorem 1,

$$E_\xi^n\{h^*(W_n)\} = \Phi h^* + \frac{1}{\sqrt{n}}(\Phi U h^*)E_\xi^n(\Gamma_1^\xi) + \frac{1}{n}(\Phi U^2 h^*)E_\xi^n(\Gamma_2^\xi) + \frac{1}{n}\Delta_n,$$

where $E_\xi(|\Delta_n| 1_{B_n}) \to 0$ as $n \to \infty$. Then, from Lemma 8,

$$E_\xi\{h(W_n^*)\} = \Phi h + \frac{1}{n}(\Phi U^2 h)\int_\Omega \{\mu(\theta)^2 - q(\theta) - 2\mu(\theta)\Gamma_1^\xi(\theta) + \Gamma_2^\xi(\theta)\}\xi(\theta)\,\mathrm{d}\theta + \mathrm{o}\left(\frac{1}{n}\right)$$

uniformly with respect to $h \in \mathscr{H}_2^o \cap \mathscr{H}_2^s$. Consider the integrand in the coefficient of $1/n$. After collecting and taking advantage of cancellation, it may be written

$$\frac{1}{3}\left(\frac{\kappa_3 \xi}{i^2}\right)' + \left(\frac{\xi'}{i}\right)',$$

the integral of which vanishes by (9). The calculation is similar to that which led to (12). □

Acknowledgements

Part of this work was carried out while the first author was a visiting scholar at the University of Michigan during August and September 2000, and in receipt of Overseas Travel Grant GR/N37568 from the U.K. Engineering and Physical Sciences Research Council. The second author's research for this paper was supported by the U.S. Army and National Science Foundation. The authors also wish to thank the two referees for their comments and suggestions.

References

Atkinson, A.C., Donev, A.N., 1992. Optimal Experimental Designs, Clarendon Press, Oxford.
Barndorff-Nielsen, O.E., Cox, D.R., 1989. Asymptotic Techniques for Use in Statistics, Chapman & Hall, London.

Bates, D.M., Watts, D.G., 1988. Nonlinear Regression Analysis and its Applications, Wiley, New York.
Berger, J.O., Wolpert, R.L., 1984. The Likelihood Principle, Institute of Mathematical Statistics, Hayward, CA.
Billingsley, P., 1985. Probability and Measure, 2nd Edition. Wiley, New York.
Box, G.E.P., Lucas, H.L., 1959. Design of experiments in nonlinear situations. Biometrika 49, 77–90.
Chaudhuri, P., Mykland, P.A., 1993. Nonlinear experiments: optimal design and inference based on likelihood. J. Amer. Statist. Assoc. 88, 538–546.
Coad, D.S., Woodroofe, M.B., 1996. Corrected confidence intervals after sequential testing with applications to survival analysis. Biometrika 83, 763–777.
Coad, D.S., Woodroofe, M.B., 1998. Approximate bias calculations for sequentially designed experiments. Sequential Anal. 17, 1–31.
Currie, D.J., 1982. Estimating Michaelis–Menten parameters: bias, variance and experimental design. Biometrics 38, 907–919.
Dette, H., Melas, V.B., Pepelyshev, A., 2003. Standardised maximin E-optimal designs for the Michaelis–Menten model. Statist. Sinica 13, 1147–1163.
Diaconis, P., Zabell, S., 1991. Closed form summation for classical distributions: variations on a theme of de Moivre. Statist. Sci. 6, 284–302.
Draper, N.R., Smith, H., 1998. Applied Regression Analysis, 3rd Edition. Wiley, New York.
Ford, I., Titterington, D.M., Kitsos, C.P., 1989. Recent advances in nonlinear experimental design. Technometrics 31, 49–60.
Pukelsheim, F., 1993. Optimal Design of Experiments, Wiley, New York.
Ratkowsky, D.A., 1983. Nonlinear Regression Modelling: A Unified Practical Approach, Marcel Dekker, New York.
Seber, G.A.F., Wild, C.J., 1989. Nonlinear Regression, Wiley, New York.
Silvey, S.D., 1980. Optimal Design: An Introduction to the Theory for Parameter Estimation, Chapman & Hall, London.
Stein, C., 1981. Estimation of the mean of a multivariate normal distribution. Ann. Statist. 9, 1135–1151.
Stein, C., 1986. Approximate Computation of Expectations, Institute of Mathematical Statistics, Hayward, CA.
Woodroofe, M., 1986. Very weak expansions for sequential confidence levels. Ann. Statist. 14, 1049–1067.
Woodroofe, M., 1989. Very weak expansions for sequentially designed experiments: linear models. Ann. Statist. 17, 1087–1102.
Woodroofe, M., 1991. Corrected confidence levels for adaptive nonlinear regression. Amer. J. Math. Man. Sci. 11, 79–93.
Woodroofe, M., 1992. Integrable expansions for posterior distributions for one-parameter exponential families. Statist. Sinica 2, 91–111.
Woodroofe, M., Coad, D.S., 1997. Corrected confidence sets for sequentially designed experiments. Statist. Sinica 7, 53–74.
Woodroofe, M., Coad, D.S., 1999. Corrected confidence sets for sequentially designed experiments: examples. In: Ghosh, S. (Ed.), Multivariate Analysis, Design of Experiments and Survey Sampling: A Tribute to Jagdish N. Srivastava. Marcel Dekker, New York, pp. 135–161 (Reprinted in Sequential Anal. 21, 191–218 (2002)).