Malaysian Journal of Science **17B**: 75-76 (1996)

On Bernoulli's inequality

Kong-Ming Chong
Department of Mathematics, University of Malaya, 50603 Kuala Lumpur, Malaysia

ABSTRACT In this paper, Bernoulli's inequality and some of its generalizations are shown to be simple consequences of a spectral inequality.

ABSTRAK Dalam kertas ini, ketaksamaan Bernoulli dan beberapa pengitlakannya ditunjukkan sebagai kesimpulan mudah bagi sesuatu ketaksamaan spektral.

(Rearrangement/spectral inequality, n-tuples)

INTRODUCTION

In Ref. [1], the Bernoulli inequality

$$(1+x)^n \geq 1+nx \tag{1}$$

for $x > -1$ and integer $n \geq 1$, is proved by means of mathematical induction. In Ref. [2], a whole chapter is devoted to Bernoulli's inequality and its ramifications. In this paper, we show that Bernoulli's inequality turns out to be a rearrangement inequality which can be obtained as a simple consequence of a spectral inequality.

PRELIMINARIES

Let $\mathbf{R}^n$ denote the set of all n-tuples of real numbers. For any n-tuple $\mathbf{x} = (x_1, x_2, \ldots, x_n) \in \mathbf{R}^n$, we denote by $\mathbf{x}^* = (x_1^*, x_2^*, \ldots, x_n^*)$ the n-tuple whose components are those of $\mathbf{x}$ arranged in non-increasing order of magnitude. If $\mathbf{a} = (a_1, a_2, \ldots, a_n) \in \mathbf{R}^n$ and $\mathbf{b} = (b_1, b_2, \ldots, b_n) \in \mathbf{R}^n$, then we say that the *weak spectral inequality* $\mathbf{a} \prec\prec \mathbf{b}$ holds whenever

$$\sum_{i=1}^{k} a_i^* \leq \sum_{i=1}^{k} b_i^*, \qquad 1 \leq k \leq n \tag{2}$$

and that the *strong spectral inequality* $\mathbf{a} \prec \mathbf{b}$ holds whenever $\mathbf{a} \prec\prec \mathbf{b}$ and $\sum_{i=1}^{n} a_i = \sum_{i=1}^{n} b_i$.

SOME SPECTRAL INEQUALITIES

The following theorem gives a simple spectral inequality from which the Bernoulli inequality can be derived as a rearrangement inequality. We note that this spectral inequality is well-known (see for example, [3], Lemma 3.4 where the result is given in its most general form for measurable functions). For our purpose here, we restate and reprove it in its simplest form for n-dimensional real vectors.

1. THEOREM

Suppose that $\mathbf{a} = (a_1, a_2, \ldots, a_n) \in \mathbf{R}^n$ is any n-tuple. If $\bar{\mathbf{a}}$ is the n-tuple whose components are each equal to $\frac{1}{n}\sum_{i=1}^{n} a_i$, then

$$\bar{\mathbf{a}} \prec \mathbf{a}. \tag{3}$$

PROOF. Since $a_1^* \geq a_2^* \geq \ldots \geq a_k^* \geq \ldots \geq a_n^*$, we have

$$\sum_{i=k+1}^{n} a_i^* \leq \sum_{i=k+1}^{n} a_k^* = (n-k)a_k^* = \frac{n-k}{k}(ka_k^*) \leq \left(\frac{n}{k}-1\right)\sum_{i=1}^{k} a_i^*$$

implying that

$$\sum_{i=1}^{n} a_i^* \leq \frac{n}{k} \sum_{i=1}^{k} a_i^*$$

or

$$k\left(\frac{1}{n}\sum_{i=1}^{n} a_i\right) \leq \sum_{i=1}^{k} a_i^* \qquad \text{for } 1 \leq k \leq n.$$

But

$$\sum_{i=1}^{n} a_i = \sum_{i=1}^{n} a_i^*$$

and so the strong spectral inequality (3) holds.

2. COROLLARY
If $x > -1$ and n is a positive integer, then the following strong spectral inequality holds:

$$(1+x, 1+x, \mathrm{K} \ldots + x) \prec (1+nx, 1, 1, \ldots, 1) \qquad (4)$$

where both sides are n -tuples from R^n.

PROOF. This is a direct consequence of Theorem 1.

The above corollary can be generalized along the following direction.

3. THEOREM
If each of the real numbers $x_1, x_2, \ldots, x_n$ is greater than -1 and either all are positive or all are negative, then

$$(1+x_1, \ldots, 1+x_n) \prec (1+x_1+\ldots+x_n, 1, \ldots, 1) \qquad (5)$$

where both sides are n -tuples.

PROOF. The proof is straightforward upon considering separately the two cases as mentioned in the theorem.

In order to derive the Bernoulli inequality (1), we need to invoke the following theorem.

4. THEOREM.
If $\mathbf{a} = (a_1, a_2, \ldots, a_n)$ and $\mathbf{b} = (b_1, b_2, \ldots, b_n)$ are n-tuples of positive real numbers such that $\mathbf{a} \prec \mathbf{b}$, then

$$a_1 a_2 \ldots a_n \geq b_1 b_2 \ldots b_n \qquad (6)$$

with strict inequality unless **a** is a permutation of **b**.

PROOF. Since the function -ln is strictly convex, by [1, Theorem 2.1, p. 1327], we have

$$-\ln a_1 - \ln a_2 - \ldots - \ln a_n \leq -\ln b_1 - \ln b_1 - \ln b_2 - \ldots - \ln b_n$$

whence the result follows by virtue of the fact that the function ln is strictly increasing.

It is now easy to derive Bernoulli's inequality (1) (in view of Corollary 2) and also its generalization

$$(1+x_1) \ldots (1+x_n) > 1+x_1+ \ldots + x_n \qquad (7)$$

for $x_i > 0$ or $0 > x_i > -1$, $i = 1, 2, \ldots, n$, by virtue of Theorem 3.

REFERENCES

1 Mitrinovic, D.S. (1970) *Analytic Inequalities*. Springer Verlag: Berlin, Heidelberg, New York, p. 34.
2 Mitrinovic, D.S., Pecaric, J.E. and Fink, A.M. (1993) *Classical and New Inequalities in Analysis*. Kluwer Academic Publishers: Dordrecht, Boston, London, pp. 65-81.
3 Chong, K.M. (1974) *Can. J. Math.* **26**: 1321.

Malaysian Journal of Science **17B:** 77-85 (1996)

Region estimates based on local linearization in a nonlinear model

S.L. Wu and A.H. Pooi
Department of Mathematics, University of Malaya, 50603 Kuala Lumpur, Malaysia

ABSTRACT This paper is concerned with region estimates based on local linearization for the parameter vector in a nonlinear model. The quadratic approximation of the coverage probability of the region estimates is derived. The coverage probability thus derived may be used to identify the situation when region estimates have approximately the desired level of confidence.

ABSTRAK Kertas ini adalah berkenaan dengan anggaran rantau berdasarkan pelinearan tempatan bagi vektor parameter dalam model tak linear. Penghampiran kuadratik bagi kebarangkalian liputan untuk anggaran rantau diterbitkan. Kebarangkalian liputan tersebut boleh digunakan untuk mencamkan situasi bila anggaran rantau hampir-hampir mempunyai aras keyakinan yang dikehendaki.

(region estimates, local linearization)

INTRODUCTION

Consider a nonlinear regression model given by

$$y_u = \eta(\xi_u, \theta) + \varepsilon_u, \qquad u = 1,2,\ldots,n \tag{1.1}$$

where y_u is the u-th observation with mean $\eta(\xi_u, \theta)$ and random error ε_u; ξ_u is a vector of known variables θ is a $(p \times 1)$ vector of unknown parameters belonging to a parameter space Ω which is a subset of the p dimensional Euclidean space. The function $\eta(\xi_u, \theta)$ is sometimes referred to as response function and it is a known, scalar-valued function. Furthermore, $\varepsilon_1, \varepsilon_2, \ldots, \varepsilon_n$ are assumed to be identical, independent and normally distributed with zero mean and error variance σ^2.

Let $\eta(\theta) = \left(\eta(\xi_1,\theta)\ \eta(\xi_2,\theta)\ \ldots\ \eta(\xi_n,\theta)\right)^T$. The solution locus is then a subset of an n dimensional Euclidean space given by

$$\{\eta(\theta) : \theta \in \Omega\}$$

Let $\mathbf{y}$ denote a column vector of which the i-th component is y_i. For a given vector $\mathbf{y}$ of observations, the least squares estimate $\hat{\theta}$ of θ is then the value of θ which minimizes the residual sum of squares

$$S(\theta) = \sum_{u=1}^{n} \left(y_u - \eta(\xi_u, \theta)\right)^2. \tag{1.2}$$

Various methods have been proposed for constructing confidence regions for the parameter vector θ in the nonlinear regression model given by (1.1). One of the common methods is that based on likelihood ratio [1-5]. Another method is one which treats the model as if it were linear in the parameterization θ in the neighbourhood of the least squares estimate $\hat{\theta}$, and applies the usual linear model theory. The latter method shall be referred to as the method based on local linearization.

In Refs. [1, 6, 7], measures of nonlinearity were proposed for identifying the situation when an inference based on local linearization is applicable. In Ref. [8] Beale's measures of nonlinearity have been investigated. In Ref. [9] simulation has been used to obtain the coverage probability of the confidence regions based on local linearization and found that cases when the coverage probability deviates a lot from the desired value appear to be reliably predicted by Bates and Watts' parameter effects curvature diagnostic.

A natural way of finding out the cases when the coverage probability is close to the desired value is by looking at the value of the coverage probability directly. In Section 2 of this paper, we derive a quadratic approximation of the coverage probability of the region estimates based on local linearization in a nonlinear model with known error variance. Corresponding result in a model with unknown error variance is derived in Section 3. In Section 4, we verify the result in Section 3 in some nonlinear models.

CONFIDENCE REGIONS IN THE CASE WHEN THE ERROR VARIANCE IS KNOWN

As we know, when the theoretical means $\eta(\xi_u, \theta)$ are linear functions of the parameter vector θ, the regression model is linear and it can be written as follows

$$\mathbf{y} = \mathbf{X}\theta + \varepsilon$$

where $\mathbf{X}$ is an $(n \times p)$ matrix of constants. If the error variance σ^2 is known, then the $100(1-\alpha)$ % confidence regions for θ, based on the usual theory in linear models, are of the following form

$$\left\{\theta : \left(\theta - \hat{\theta}\right)^T \mathbf{X}^T\mathbf{X}\left(\theta - \hat{\theta}\right) \leq \sigma^2 \chi^2_{p,\alpha}\right\}$$

where $\chi^2_{p,\alpha}$ is defined as the $100(1-\alpha)$ percentage point of a chi-square distribution with p degrees of freedom.

Let us approximate $\eta(\xi_u, \theta)$ by the following linear function of θ

$$\eta(\xi_u, \theta) \cong \eta\left(\xi_u, \hat{\theta}\right) + \sum_{j=1}^{p} c_{uj}\left(\hat{\theta}\right)\left(\theta_j - \hat{\theta}_j\right) \tag{2.1}$$

where $$c_{uj}\left(\hat{\theta}\right) = \left.\frac{\partial \eta(\xi_u, \theta)}{\partial \theta_j}\right|_{\theta = \hat{\theta}}.$$

Then, for a given vector $\mathbf{y}$ of observations, we can obtain a confidence region $R_1(\mathbf{y})$ for θ given by

$$R_1(\mathbf{y}) = \left\{\theta : \left(\theta - \hat{\theta}\right)^T \left[\mathbf{C}\left(\hat{\theta}\right)\right]^T \left[\mathbf{C}\left(\hat{\theta}\right)\right]\left(\theta - \hat{\theta}\right) \leq \sigma^2 \chi^2_{p,\alpha}\right\} \tag{2.2}$$

where $\mathbf{C}\left(\hat{\theta}\right) = \left\{c_{uj}\left(\hat{\theta}\right)\right\}$ is the $(n \times p)$ matrix of first order partial derivatives. From now onwards, a matrix of which the (i, j) entry is m_{ij} shall be denoted by $\{m_{ij}\}$ or $\mathbf{M}$. The regions of the form given by the right hand side of (2.2) shall be referred to as the nominally-$100(1-\alpha)$ % confidence regions based on local linearization for θ in the case when σ^2 is known.

The actual coverage probability is then the probability that these confidence regions will cover the true value θ_T of the parameter vector θ, and it may be treated as a function of θ_T as follows

$$I_1(\theta_T) = P\{\theta_T \in R_1(\mathbf{y}) | \theta_T\}.$$

This actual coverage probability is usually unknown. However we can attempt to estimate its value. First, given a feasible value θ_f of θ, we obtain the coverage probability $I_1(\theta_f)$ evaluated at θ_f. We then investigate the extreme values of $I_1(\theta_f)$ over a certain subset Ω_f of Ω which we think would contain θ_T. If both the minimum and maximum values of $I_1(\theta_f)$ are not far off from $(1-\alpha)$, then we may refer to the regions of the form given by (2.2) as approximately-$100(1-\alpha)$ % confidence regions for θ.

Let us confine our attention to models satisfying the following conditions

(a) For each $\theta_f \in \Omega_f$, there exists $\delta > 0$ such that $|\theta - \theta_f| < \delta$ implies that $\theta \in \Omega_f$.
(b) If $\eta(\theta_1) = \eta(\theta_2)$, then $\theta_1 = \theta_2$.
(c) The $\eta(\xi_u, \theta)$ are functions of θ with continuous derivatives up to the third order in Ω_f.
(d) The matrix $\mathbf{C}(\theta_f)$ is of rank p for all $\theta_f \in \Omega_f$.

Supposing condition (c) is valid, then by Taylor's series expansion about θ_f, $\eta(\xi_u, \theta)$ can be written as

$$\eta(\xi_u, \theta) = \eta(\xi_u, \theta_f) + \sum_{j=1}^{p} c_{uj} t_j + \mathbf{t}^T\mathbf{C}_u\mathbf{t} + \sum_{j=1}^{p} [\mathbf{t}^T\mathbf{C}_{uj}\mathbf{t}] t_j + o(t^3), \quad u = 1, 2, \ldots, n$$

where

$$\mathbf{t} = \theta - \theta_f, \quad c_{uj_1 j_2 \ldots j_m} = \frac{1}{m!}\left[\left.\frac{\partial^m \eta(\xi_u, \theta)}{\partial \theta_{j_1} \partial \theta_{j_2} \ldots \partial \theta_{j_m}}\right|_{\theta = \theta_f}\right], \quad m = 1, 2, 3,$$

and t denotes the magnitude of $\mathbf{t}$.

We introduce an orthogonal $(n \times n)$ matrix $\mathbf{H}$ such that $\mathbf{HC}$ is an upper triangular $(p \times p)$ matrix $\mathbf{D}$ with an $((n-p) \times p)$ zero matrix beneath it [10,11]. An orthogonal transformation

$$\mathbf{H}(\mathbf{y} - \eta(\theta_f)) = \mathbf{z} \tag{2.3}$$

of coordinates in sample space is then applied so that the point $\eta(\theta_f)$ in the solution locus becomes the new origin $\mathbf{z} = \mathbf{0}$ and the tangent plane to the solution locus at $\eta(\theta_f)$ consists of all points for which $z_i = 0$ for $i = p+1, p+2, \ldots, n$. The components of $\mathbf{z}$ shall be referred to as the rotated coordinates of the sample point $\mathbf{y}$. Subsequently for a point $\eta(\theta)$ in the solution locus, its i-th rotated coordinate z_i^* can be expressed in the following form

$$z_i^* = \begin{cases} \sum_{j=1}^{p} d_{ij} t_j + \mathbf{t}^T \mathbf{D}_i \mathbf{t} + \sum_{j=1}^{p} [\mathbf{t}^T \mathbf{D}_{ij} \mathbf{t}] t_j + o(t^3), & i = 1,2,\ldots,p \\ \mathbf{t}^T \mathbf{D}_i \mathbf{t} + \sum_{j=1}^{p} [\mathbf{t}^T \mathbf{D}_{ij} \mathbf{t}] t_j + o(t^3), & i = p+1, p+2,\ldots,n \end{cases} \tag{2.4}$$

where d_{ijk} and d_{ijkl} are the i-th components of $\mathbf{Hc}_{jk}$ and $\mathbf{Hc}_{jkl}$ respectively, while $\mathbf{c}_{jk}$ and $\mathbf{c}_{jkl}$ are $(n \times 1)$ column vectors in which the u-th components are c_{ujk} and c_{ujkl} respectively. Equation (2.4) can next be transformed into

$$z_i^* = \begin{cases} \tau_i + \tau^T \mathbf{F}_i \tau + \sum_{j=1}^{p} [\tau^T \mathbf{F}_{ij} \tau] \tau_j + o(\tau^3), & i = 1,2,\ldots,p \\ \tau^T \mathbf{F}_i \tau + \sum_{j=1}^{p} [\tau^T \mathbf{F}_{ij} \tau] \tau_j + o(\tau^3), & i = p+1, p+2,\ldots,n \end{cases} \tag{2.5}$$

where

$\tau = \mathbf{Dt}$, $\mathbf{F}_i = (\mathbf{D}^{-1})^T \mathbf{D}_i \mathbf{D}^{-1}$, with $\mathbf{D}^{-1} = \{d^{ij}\}$

and $$\mathbf{F}_{ij} = (\mathbf{D}^{-1})^T \left[\sum_{m=1}^{p} d^{mj} \mathbf{D}_{im} \right] \mathbf{D}^{-1}.$$

Equation (2.5) can further be expressed as follows

$$z_i^* = \begin{cases} \phi_i, & i = 1,2,\ldots,p \\ \phi^T \mathbf{A}_i \phi + \sum_{j=1}^{p} [\phi^T \mathbf{A}_j \phi] \phi_j + o(\phi^3), & i = p+1, p+2,\ldots,n \end{cases}$$

where

$$\phi_i = \tau_i + \tau^T \mathbf{F}_i \tau + \sum_{j=1}^{p} [\tau^T \mathbf{F}_{ij} \tau] \tau_j + o(\tau^3)$$

$$\mathbf{A}_i = \mathbf{F}_i, \text{ and } \mathbf{A}_{ij} = \mathbf{F}_{ij} - 2 \sum_{k=1}^{p} f_{ijk} \mathbf{F}_k.$$

The residual sum of squares $S(\theta)$ can now be expressed in terms of ϕ. By minimizing $S(\theta)$ with respect to ϕ, we can find an approximation of the least squares estimate $\hat{\phi}_v$ of the parameter ϕ_v [1]. We can next obtain the least squares estimate $\hat{t}_w$ of t_w in terms of $\mathbf{z}$ as follows

$$\begin{aligned} \hat{t}_w = & \sum_{v=w}^{p} d^{wv} z_v + \sum_{v=w}^{p} d^{wv} \Bigg[2 \sum_{i=p+1}^{n} \sum_{j=1}^{p} a_{ijv} z_i z_j - \sum_{j=1}^{p} \sum_{l=1}^{p} a_{vjl} z_j z_l \\ & + 4 \sum_{i=p+1}^{n} \sum_{h=p+1}^{n} \sum_{j=1}^{p} \sum_{k=1}^{p} a_{ijv} a_{hjk} z_i z_h z_k - 2 \sum_{i=p+1}^{n} \sum_{j=1}^{p} \sum_{k=1}^{p} \sum_{l=1}^{p} a_{ijk} a_{ilv} z_j z_k z_l \\ & + 3 \sum_{i=p+1}^{n} \sum_{j=1}^{p} \sum_{l=1}^{p} a_{ijvl} z_i z_j z_l - 4 \sum_{h=p+1}^{n} \sum_{j=1}^{p} \sum_{k=1}^{p} \sum_{l=1}^{p} a_{vjl} a_{hkl} z_h z_j z_k \\ & - \sum_{j=1}^{p} \sum_{k=1}^{p} \sum_{l=1}^{p} a_{vjkl} z_j z_k z_l \Bigg] + o(z^3), \qquad w = 1,2,\ldots,p. \end{aligned} \tag{2.6}$$

Since $\mathbf{H}$ is an orthogonal matrix, the matrix $\left[\mathbf{C}\left(\hat{\theta}\right)\right]^T \left[\mathbf{C}\left(\hat{\theta}\right)\right]$ appearing in (2.2) can also be written as $\left[\mathbf{C}\left(\hat{\theta}\right)\right]^T \mathbf{H}^T \mathbf{H} \left[\mathbf{C}\left(\hat{\theta}\right)\right]$ with the (i,j) entry of $\hat{\mathbf{D}} = \mathbf{H}\left[\mathbf{C}\left(\hat{\theta}\right)\right]$ given by

$$\hat{d}_{ij} = d_{ij} + 2 \sum_{k=1}^{p} d_{ijk} \hat{t}_k + 3 \sum_{k=1}^{p} \sum_{l=1}^{p} d_{ijkl} \hat{t}_k \hat{t}_l + o\left(\hat{t}^2\right), \quad i = 1,2,\ldots,n; \; j = 1,2,\ldots,p. \tag{2.7}$$

From (2.6) and (2.7), we see that the $\hat{d}_{ij}$ can now be expressed as functions of **z**. In addition, a confidence region given by (2.2) can also be written in the following form

$$\left\{\theta : \left(\mathbf{t}-\hat{\mathbf{t}}\right)^T \hat{\mathbf{D}}^T \hat{\mathbf{D}}\left(\mathbf{t}-\hat{\mathbf{t}}\right) \le \sigma^2 \chi^2_{p,\alpha}\right\} \quad (2.8)$$

in which both $\hat{\mathbf{t}}$ and $\hat{\mathbf{D}}$ are functions of **z**.

It is noted that a region given by (2.8) will cover the parameter vector θ_f if and only if

$$\hat{\mathbf{t}}^T \hat{\mathbf{D}}^T \hat{\mathbf{D}} \hat{\mathbf{t}} \le \sigma^2 \chi^2_{p,\alpha}. \quad (2.9)$$

By substituting each expression of $\hat{d}_{ij}$ and $\hat{t}_j$ in terms of **z** into (2.9), we can express the inequality in terms of **z**. The resulting inequality may then be approximated by the following inequality

$$\sum_{i=1}^{p} z_i'^2 \le \chi^2_{p,\alpha} + \psi_1(z_1', z_2' \ldots, z_n', \mathbf{a}) \quad (2.10)$$

where $z_i' = z_i / \sigma$, $i = 1,2,\ldots,n$, **a** is a vector whose components are the $a_{ijk}\sigma$ and $f_{ijkl}\sigma^2$ where $i = 1,2,\ldots,n$; $j,k,l = 1,2,\ldots,p$, $j \le k \le l$; whereas ψ_1 ψ_1 is a function that sums up a finite number of expressions, each of which is of the form

$$\text{constant} \times \pi_1 \times \pi_2 \times \sigma^m, \qquad m = 1,2,$$

where π_1 represents the product of some components in **z**' and π_2 represents an a_{ijk}, f_{ijkl} or $a_{ijk} \cdot a_{lmv}$.

Now let $z_i' = s_i \{z_i^{(s)}\}^{1/2}$

where

$$s_i = \begin{cases} -1 & \text{if } z_i' < 0 \\ 1 & \text{if } z_i' \ge 0, \end{cases} \qquad i = 1,2,\ldots,p,$$

and apply the transformation

$$r^{(s)} = \sum_{j=1}^{p} z_j^{(s)}, \quad \bar{z}_i^{(s)} = \frac{z_i^{(s)}}{r^{(s)}}, \qquad i = 1,2,\ldots,p-1.$$

Then to the extent that the approximation given by (2.10) is adequate, we can express the coverage probability $I_1(\theta_f)$ as follows

$$I_1(\theta_f) = E_{z'_{p+1}, z'_{p+2}, \ldots, z'_n} \sum_{s_1=-1,1} \sum_{s_2=-1,1} \mathrm{K} \sum_{s_p=-1,1}$$
$$\int_{\bar{z}_1^{(s)}=0}^{1} \int_{\bar{z}_2^{(s)}=\bar{z}_1^{(s)}}^{1} \cdots \int_{\bar{z}_{p-1}^{(s)}=\bar{z}_{p-2}^{(s)}}^{1} \int_{r^{(s)} \in K_1^*}$$
$$\frac{1}{2^p}\left[\prod_{i=1}^{p-1} \chi_1^2(\bar{z}_i^{(s)})\right] \chi_1^2\left(1-\sum_{j=1}^{p-1} \bar{z}_j^{(s)}\right) \frac{\chi_p^2(r^{(s)})}{\chi_p^2(1)} dr^{(s)} d\bar{z}_{p-1}^{(s)} d\bar{z}_{p-2}^{(s)} \ldots d\bar{z}_1^{(s)}$$

(2.11)

where K_1^* is the set containing the values of $r^{(s)}$ which satisfy

$$r^{(s)} \le \chi^2_{p,\alpha} + \psi_1\left(s_1\left\{r^{(s)}\bar{z}_1^{(s)}\right\}^{1/2}, s_2\left\{r^{(s)}\bar{z}_2^{(s)}\right\}^{1/2}, \ldots, s_{p-1}\left\{r^{(s)}\bar{z}_{p-1}^{(s)}\right\}^{1/2},\right.$$
$$\left. s_p\left\{r^{(s)}\left(1-\sum_{i=1}^{p-1}\bar{z}_i^{(s)}\right)\right\}^{1/2}, z'_{p+1}, z'_{p+2}, \ldots, z'_n, \mathbf{a}\right)$$

(2.12)

and

$$\chi_k^2(r^{(s)}) = \frac{1}{\Gamma(k/2)2^{k/2}} \left(r^{(s)}\right)^{(k/2)-1} \exp(-r^{(s)}/2).$$

So far we have been treating the a_{ijk} and f_{ijkl} as constants. Now if we imagine that they are variables and keep θ_f and σ fixed, then $I_1(\theta_f)$ can be treated as a function of the $\overset{+}{a}_{ijk}$ and $\overset{+}{f}_{ijkl}$ given respectively by $a_{ijk}\sigma$ and $f_{ijkl}\sigma^2$. A truncated series expansion of $I_1(\theta_f)$ is then given by

$$I_1(\theta_f) \cong 1 - \alpha + I^{(1)} + I^{(2)} \quad (2.13)$$

where $I^{(1)}$ is a linear combination of the $\overset{+}{a}_{ijk}$, and $I^{(2)}$ is a linear combination of the $\overset{+}{a}_{ijk}\overset{+}{a}_{lmv}$ plus a linear combination of the $\overset{+}{f}_{ijkl}$.

The expansion in (2.13) is truncated in such a way that for a specific model and a given value of θ_f, the right hand side of (2.13) is a quadratic function of σ.

When the magnitude of **a** is sufficiently small, the set K_1^* can be written as

$$K_1^* = \left\{ r_1^{(s)} : 0 \le r^{(s)} \le r_1^{(s)*} \right\} \tag{2.14}$$

where $r_1^{(s)*}$ satisfies (2.12) with the inequality sign changed to the equal sign and $r_1^{(s)}$ changed to $r_1^{(s)*}$.

With K_1^* taking the form as shown in (2.14), the partial derivatives of $I_1(\theta_f)$ given by (2.11) can be obtained by applying the Leibnitz's rule for differentiation under the integral sign.

In deriving $I^{(1)}$ and $I^{(2)}$ which appear in (2.13), we may make use of the following result

$$\delta_{m_1,m_2} = \mathbf{E}_{z'_{p+1}, z'_{p+2}, \ldots, z'_n} \left(\chi_p^2 \left(\chi_{p,\alpha}^2 \right) \right) \left(\chi_{p,\alpha}^2 \right)^{(m_1)/2} \left(z'_i \right)^{m_2}$$

$$= \frac{2^{(m_1+m_2)/2}\,\Gamma\left(\dfrac{m_2+1}{2}\right)\Gamma\left(\dfrac{p+m_1}{2}\right)}{\sqrt{\pi}\,\Gamma\left(\dfrac{p}{2}\right)} \chi_{p+m_1}^2 \left(\chi_{p,\alpha}^2 \right),$$

$i \in \{p+1, p+2, \ldots, n\}$, $m_1 \in \mathbf{Z}^+$, and $m_2 \in \{0, 2, 4, 6, \ldots\}$.

In the present case when σ^2 is known and $p \ge 2$, the quadratic approximation of the coverage probability of the region estimates can be shown to be given as follows

$$I_1(\theta_f) \cong 1 - \alpha + I_{11}\sigma^2$$

$$- \frac{\sigma^2}{(p+2)p} \left\{ 8\left[(p+2)\delta_{2,2} + 2\delta_{4,2} \right] \sum_{i=p+1}^{n} \sum_{j=1}^{p-1} \sum_{k=j+1}^{p} a_{ijk}^2 + 12\delta_{4,2} \sum_{i=p+1}^{n} \sum_{j=1}^{p} a_{ijj}^2 \right.$$

$$\left. -8\left[(p+2)\delta_{2,2} - \delta_{4,2} \right] \sum_{i=p+1}^{n} \sum_{j=1}^{p-1} \sum_{m=j+1}^{p} a_{ijj} a_{imm} \right\}$$

where

$$I_{11} = \frac{1}{(p+4)(p+2)p} \left\{ 3\left[8(p+4)\delta_{4,0} - 5\delta_{6,0} \right] \sum_{i=1}^{p} a_{iii}^2 - 3\delta_{6,0} \sum_{i=1}^{p} \sum_{\substack{j=1 \\ j \ne i}}^{p} a_{ijj}^2 \right.$$

$$+4\left[4(p+4)\delta_{4,0} - 3\delta_{6,0} \right] \sum_{i=1}^{p} \sum_{\substack{j=1 \\ j \ne i}}^{p} a_{ijj}^2 - 4\delta_{6,0} \sum_{i=1}^{p} \sum_{\substack{j=1 \\ j \ne i}}^{p} \sum_{\substack{k=j+1 \\ k \ne i}}^{p} a_{ijk}^2$$

$$+2\left[4(p+4)\delta_{4,0} - 3\delta_{6,0} \right] \sum_{i=1}^{p} \sum_{\substack{j=1 \\ j \ne i}}^{p} a_{iii} a_{ijj} + 4\left[4(p+4)\delta_{4,0} - 3\delta_{6,0} \right] \sum_{i=1}^{p} \sum_{\substack{j=1 \\ j \ne i}}^{p} a_{iii} a_{jji}$$

$$+8\left[(p+4)\delta_{4,0} - \delta_{6,0} \right] \sum_{i=1}^{p-1} \sum_{l=i+1}^{p} \sum_{\substack{j=1 \\ j \ne l \\ j \ne i}}^{p} a_{iij} a_{llj} + 4\left[2(p+4)\delta_{4,0} - \delta_{6,0} \right] \sum_{i=1}^{p} \sum_{\substack{j=1 \\ j \ne i}}^{p} \sum_{\substack{m=1 \\ m \ne j \\ m \ne i}}^{p} a_{iij} a_{jmm}$$

$$+4\left[8(p+4)\delta_{4,0} - 3\delta_{6,0} \right] \sum_{i=1}^{p} \sum_{\substack{j=1 \\ j \ne i}}^{p} a_{iij} a_{jii} - 2\delta_{6,0} \sum_{i=1}^{p} \sum_{\substack{j=1 \\ j \ne i}}^{p-1} \sum_{\substack{m=j+1 \\ m \ne i}}^{p} a_{ijj} a_{imm}$$

$$\left. +8\left[3(p+4)\delta_{4,0} - \delta_{6,0} \right] \sum_{i=1}^{p-1} \sum_{j=i+1}^{p} \sum_{\substack{k=1 \\ k \ne j \\ k \ne i}}^{p} a_{ijk} a_{jki} - 12(p+4)\delta_{4,0} \sum_{i=1}^{p} \sum_{j=1}^{p} f_{iijj} \right\}.$$

CONFIDENCE REGIONS IN THE CASE WHEN THE ERROR VARIANCE IS UNKNOWN

In Section 2, the error variance σ^2 is assumed to be known. In this section we shall consider the case when σ^2 is unknown.

Suppose $\eta(\xi_u, \theta)$ is approximated by the linear function of θ given by (2.1). Then, for the given vector **y** of observations, we can obtain a confidence region $R_2(\mathbf{y})$ for θ given by

$$R_2(\mathrm{y}) = \left\{ \theta : \left(\theta - \hat{\theta} \right)^T \left[\mathrm{C}(\hat{\theta}) \right]^T \left[\mathrm{C}(\hat{\theta}) \right] \left(\theta - \hat{\theta} \right) \le \frac{p\mathrm{S}(\hat{\theta})}{\mathrm{n-p}} F_{p,n-p,\alpha} \right\} \tag{3.1}$$

where $F_{p,n-p,\alpha}$ is the $100(1-\alpha)$ percentage point of an F- distribution with p and $(n-p)$ degrees of freedom. The regions of the form given by the right hand side of (3.1) shall be referred to as the nominally - $100(1-\alpha)\%$ confidence regions based on local linearization for θ in the case when σ^2 is unknown.

The probability that the confidence regions of the form given by (3.1) will cover the true value θ_T of the parameter vector θ is a value of interest to us. This probability may be treated as a function of θ_T and the true value σ_T of σ, as follows

$$I_2(\theta_T, \sigma_T) = P\{\theta_T \in R_2(\mathbf{y}) | \theta_T, \sigma_T\}$$

Though this actual coverage probability is usually unknown, we can attempt to estimate its value. First, given a feasible value θ_f of θ, and a value of σ, we obtain the coverage probability $I_2(\theta_f, \sigma)$ evaluated at θ_f and σ. We then investigate the extreme values of $I_2(\theta_f, \sigma)$ over a certain subset Ω_f of Ω which we think would contain θ_T, and an interval $(0,\sigma_f)$ which we think would cover σ_T. If both the minimum and maximum values of $I_2(\theta_f, \sigma)$ are not far off from $(1-\alpha)$, then we may refer to the regions of the form given by (3.1) as approximately - $100(1-\alpha)\%$ confidence regions for θ when σ^2 is unknown.

Following the same procedure as described in Section 2, we can express the confidence region $R_2(\mathbf{y})$ in the following form

$$R_2(\mathbf{y}) = \left\{\theta : (\theta - \hat{\theta})^T \hat{\mathbf{D}}^T \hat{\mathbf{D}} (\theta - \hat{\theta}) \le d^{*2} S(\hat{\theta})\right\}$$

where $$d^{*2} = \frac{p}{n-p} F_{p,n-p,\alpha}.$$

The confidence region given by (3.1) will cover θ_f if and only if

$$\hat{\mathbf{t}}^T \hat{\mathbf{D}}^T \hat{\mathbf{D}} \hat{\mathbf{t}} \le d^{*2} S(\hat{\theta}). \tag{3.2}$$

The minimum residual sum of squares $S(\hat{\theta})$ can be expressed as a quartic function of $\mathbf{z}[1]$. By substituting the expressions of $\hat{t}_w, \hat{d}_{ij}$ and $S(\hat{\theta})$ in terms of $\mathbf{z}$ into the inequality in (3.2), we can obtain an inequality in terms of $\mathbf{z}$. This inequality can then be approximated by an inequality of the following form

$$\sum_{i=1}^{p} z_i'^2 \le d^{*2} \sum_{j=p+1}^{n} z_j'^2 + \psi_2(z_1', z_2', \ldots, z_n', \mathbf{a}) \tag{3.3}$$

where ψ_2 is a function similar to ψ_1 in (2.10).

Using a procedure similar to that in Section 2, we can derive from (3.3) a truncated series expansion of $I_2(\theta_f, \sigma)$ similar to that for $I_1(\theta_f)$ [cf. (2.13)].

The following result may be used in deriving the truncated series expansion

$$\beta_{m_1,m_2} = E_{z'_{p+1},\ldots,z'_n}\left(\chi^2_{p,\alpha}(d^{+2})\right)(d^{+2})^{(m_1)/2}(z_j')^{m_2}$$
$$= \frac{2^{(m_1+m_2)/2}\Gamma\left(\frac{m_2+1}{2}\right)\Gamma\left(\frac{m_1+m_2+n-2}{2}\right)(d^{*2})^{(m_1+p-2)/2}}{2\sqrt{\pi}\Gamma\left(\frac{p}{2}\right)\Gamma\left(\frac{m_2+n-p}{2}\right)(1+d^{*2})^{(m_1+m_2+n-2)/2}}$$

where

$$m_2 \in \{0,2,4,6,\ldots\}; m_1 \ge 3 - m_2 - n; j \in \{p+1, p+2, \ldots, n\},$$

and $$d^{+2} = d^{*2} \sum_{j=p+1}^{n} z_j'^2.$$

For a specific model and a given value of θ_f, the coverage probability of the region estimates may be treated as a function of σ. The quadratic approximation of the coverage probability, treated as a function of σ, can be shown to be

$$
\begin{aligned}
I_2(\theta_f,\sigma) &\cong 1-\alpha+I_{21}\sigma^2 \\
&+\frac{\sigma^2}{(p+2)p}\Bigg\{4\Big[(p+2)\big(d^{*4}+2d^{*2}-2\big)\beta_{2,2}+d^{*2}\beta_{4,0}-\big(2+d^{*2}\big)^2\beta_{4,2}\Big]\sum_{i=p+1}^{n}\sum_{j=1}^{p-1}\sum_{k=j+1}^{p}a_{ijk}^2 \\
&+\Big[(p+2)d^{*2}\big(8+3d^{*2}\big)\beta_{2,2}+3d^{*2}\beta_{4,0}-3\big(2+d^{*2}\big)^2\beta_{4,2}\Big]\sum_{i=p+1}^{n}\sum_{j=1}^{p}a_{ijj}^2 \\
&+2\Big[(p+2)\big(2+d^{*2}\big)^2\beta_{2,2}+d^{*2}\beta_{4,0}-\big(2+d^{*2}\big)^2\beta_{4,2}\Big]\sum_{i=p+1}^{n}\sum_{j=1}^{p-1}\sum_{m=j+1}^{p}a_{ijj}a_{imm}\Bigg\}
\end{aligned}
$$

where I_{21} is derived from I_{11} by replacing all the $\delta_{i,0}$ in I_{11} by $\beta_{i,0}$.

VERIFICATION OF RESULTS

Consider a simple nonlinear model given by

$$y_1=-\theta_1+\varepsilon_1,\ y_2=-\theta_2-a_{222}\theta_2^2+\varepsilon_2,\ y_3=\varepsilon_3 \quad (4.1)$$

which has three observations, two parameters and only one nonlinear term a_{222}. Suppose that σ^2 is unknown, $\theta_f=0$ and $\alpha=0.05$. An orthogonal transformation of the type as given in (2.3) will transform the model given in (4.1) to the following

$$z_1=t_1+\varepsilon_1',\ z_2+a_{222}t_2^2+\varepsilon_2',\ z_3=\varepsilon_3'$$

where ε_i' is the i-th component of $\varepsilon'=\mathbf{H}\varepsilon$.

For a given value of $a_{222}^+\in(-0.05,0.05)$, we obtain the coverage probability of the confidence regions for θ by integrating a triple integral numerically. In performing the integration, the values of z_2/σ are restricted to lie in the interval (-5,5). The first reason for such a restriction is that the probability that the absolute value of a standardised normal variable will be greater than 5 is about 5.7 x 10^{-7} which is too small to have a significant effect on the final result of the integration. The second reason is that $\hat{t}_2$ will then be always unique. In Fig. 1 we present a graph of the *exact* coverage probability obtained by numerical integration and the coverage probability obtained by quadratic approximation. This figure shows that quadratic approximation is excellent when $|a_{222}^+|\le 0.027$. When $|a_{222}^+|>0.027$, deviation from the *exact* value begins to be discernible.

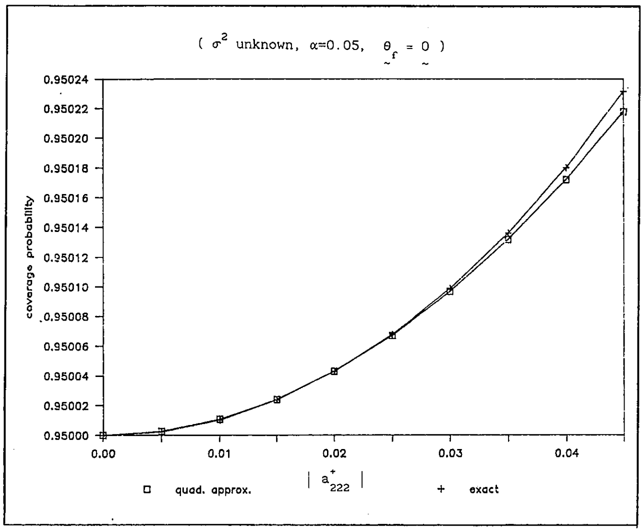

Figure 1. Coverage probability versus a_{222}^+ for the simple nonlinear model: $y_1=-\theta_1+\varepsilon_1,\ y_2=-\theta_2-a_{222}\theta_2^2+\varepsilon_2,\ y_3=\varepsilon_3$.

For a vector $\mathbf{a}$ in which the component a^+ is nonzero while all the other components are zero, let R_{q,a^+} be an interval containing all the values of a^+ such that

$$\frac{|\text{exact c.p. - quadratic c.p.}|}{|\text{exact c.p. - }(1-\alpha)|}\times 100\%\le q\% \quad (4.2)$$

where c.p. stands for coverage probability evaluated at $\theta_f=0$ and q is some positive real number. For $q=$ 1,3,5, the corresponding intervals R_{q,a^+} for $a^+=a_{222}^+$ are shown in Table 1.

Results for other two-parameter models with n=3, σ^2 unknown and only one nonlinear term a_{ijk} and f_{ijkl} are shown in Tables 1 and 2.

Table 1 shows that among the three types of nonlinear terms given by $\{a_{122}^+,a_{211}^+\}$, $\{a_{112}^+,a_{212}^+\}$ and $\{a_{111}^+,a_{222}^+\}$, the first type will have to assume a much larger abso-

Table 1.

Upper limit $(\times 10^{-2})$ of the interval R_{q,a^+} when $\alpha = 0.05$

(lower limit = -upper limit)

a^+ \ q	1	3	5
a^+_{122}, a^+_{211}	14.21	21.92	30.00
a^+_{112}, a^+_{212}	4.42	6.90	9.64
a^+_{111}, a^+_{222}	2.33	3.38	4.28
f^+_{1111}, f^+_{2222}	0.12	0.26	0.47

Table 2.

Upper limit $(\times 10^{-2})$ of the interval R_{q,a^+} when $\alpha = 0.05$

(lower limit = -upper limit)

a^+ \ q	5	10	15
a^+_{322}, a^+_{311}	0.426	0.579	0.712
a^+_{312}	0.470	0.611	0.759

lute value in order to produce the same degree of inadequacy of quadratic approximation as given by the left side of (4.2). Table 2 shows that when the two types of nonlinear terms given by $\{a^+_{322}, a^+_{311}\}$ and a^+_{312} have the same absolute value, they give rise to about the same degree of inadequacy of quadratic approximation. The two tables also show that the a^+_{3jk} terms need to assume only a very much smaller absolute value in comparison with a^+_{1jk} or a^+_{2jk} in order to produce the same adverse effect on the adequacy of quadratic approximation. We suspect that similar sort of results might also be true for two-parameter models with $n > 3$ observations.

Now consider a quadratic model with $n = 3$ and the values of the f_{ijk} and f_{ijkl} as given in the appendix. Assuming that σ^2 is unknown, and choosing $\theta_f = 0$ and $\alpha = 0.05$, we obtain the coverage probability of the confidence regions for θ by numerical integration for a number of small values of σ. In Table 3, the coverage probabilities obtained by numerical integration and quadratic approximation are presented. The table shows that for $\sigma \le 0.05$, the probabilities based on quadratic approximation agree fairly well with the *exact* values obtained by numerical integration.

Table 3. Coverage probability of the region estimates in the quadratic model.

σ	Coverage Probability -0.95	
	Quadratic Approximation (x10^{-6})	Numerical Integration (x10^{-6})
0.01	0.664	0.665
0.02	2.658	2.664
0.03	5.980	6.009
0.04	10.631	10.721
0.05	16.611	16.823

Next consider the exponential model $y_u = \theta_1 \exp\{-\xi_u \theta_2\} + \varepsilon_u, u = 1,2,3$, with $\xi_1 = -0.2692635$, $\xi_2 = -0.3987761$, $\xi_3 = -0.4768550$. Suppose that σ^2 is unknown, $\theta_f = (0.7689, 3.8600)^T$ and $\alpha = 0.05$. We obtain the coverage probabilities of the confidence regions for θ by numerical integration for a number of values of σ. In Table 4, the coverage probabilities obtained by numerical integration and quadratic approximation are presented. This table shows that for $\sigma \le 0.03$, the probabilities based on quadratic approximation differ very slightly from the *exact* values obtained by numerical integration.

Table 4. Coverage probability of the region estimates in the exponential model.

σ	Coverage Probability -0.95	
	Quadratic Approximation (x10^{-5})	Numerical Integration (x10^{-5})
0.005	-0.560	-0.599
0.010	-2.242	-2.240
0.015	-5.043	-5.037
0.020	-8.966	-8.947
0.025	-14.010	-13.964
0.030	-20.174	-20.059

Finally, consider the model

$$y_u = \frac{1}{\theta_1 \xi_{u1} + \theta_2 \xi_{u2} + \theta_3 \xi_{u3}} + \varepsilon_u, \quad u = 1,2,3,4$$

in which the inverse of the theoretical mean is a linear function. Let ξ_{uj} be given by

ξ_{uj}			
	j		
u	1	2	3
1	80.06	18.91	1.00
2	74.99	22.99	2.00
3	69.99	27.99	2.00
4	64.99	31.99	3.00

Suppose that σ^2 is unknown, $\theta_f = (0.015, 0.045, 0.5)^T$ and $\alpha = 0.05$. In Table 5, the coverage probabilities obtained by numerical integration and quadratic approximation are presented. This table shows that for the values of σ considered the probabilities based on quadratic approximation are in good agreement with the *exact* values obtained by numerical integration.

Table 5. Coverage probability of the region estimates in the model in which the inverse of the theoretical mean is a linear function.

$\sigma(\times 10^{-3})$	Coverage Probability -0.95	
	Quadratic Approximation (x10^{-7})	Numerical Integration (x10^{-7})
1.0	-4.704	-4.704
1.1	-5.687	-5.691
1.2	-6.963	-6.773

REFERENCES

1 Beale, E.M.L. (1960) *J. R. Statist. Soc. B* **22**: 41.
2 Hamilton, D. and Wiens, D. (1987) *Biometrika* **74**: 423.
3 Johansen, S. (1983) *Scand. J. Statist.* **10**: 161.
4 Pooi, A.H. (1989) *Bull. Malaysian Math. Soc.* **12**: 73.
5 Pooi, A.H. (1991) *Bull. Malaysian Math. Soc.* **14(2)**: 73.
6 Bates, D.M. and Watts D.G. (1980) *J. R. Statist. Soc. B* **42**: 1.
7 Linssen, H.N. (1975) *Statist. Neerland.* **29**: 93.
8 Guttman, I. and Meeter, D.A. (1965) *Technometrics* **7**: 623.
9 Donaldson, J.R. and Schnabel, R.B. (1986) *Computer Science and Statistics : The Interface, D.M. Allen (ed.)*, Elsevier Science Publishers B.V., North Holland.
10 Businger, P. and Golub, G.H. (1965) *Num. Math.* **7**: 269.
11 Golub, G.H. (1965) *Num. Math.* **7**: 206.

Appendix

The values of f_{ijk} and f_{ijkl} used in Section 4.

i	j	k	l	f_{ijk}	f_{ijkl}
1	1	1	1	0.04	0.0042
1	1	1	2		0.0040
1	1	2	2	0.08	0.0037
1	2	2	2	0.30	0.0041
2	1	1	1	0.30	0.0038
2	1	1	2		0.0042
2	1	2	2	0.090	0.00360
2	2	2	2	0.040	0.00430
3	1	1	1	0.008	0.00020
3	1	1	2		0.00010
3	1	2	2	0.007	0.00015
3	2	2	2	0.008	0.00030