METHODS FOR CONSTRUCTING ASYMPTOTICALLY NORMAL ESTIMATORS

by

Grete Usterud Penstad

University of Oslo
September 1965
0. INTRODUCTION

In this paper which is a sequel to [4], we shall give a rather general theorem on constructing asymptotically normal estimators with minimum asymptotic variances. This theorem generalizes previous work of Ferguson [5], Chiang [3] and Steene [6]. The methods of proof are similar to Chiang [3]. In fact, his methods carry over completely; thus we refer the reader to his paper for the details of proof.

The theorem is stated in Section 1. In Section 2 we use this result to give a simple treatment of the multinomial case, and in Section 3 we shall apply it to the gamma-distribution, considering estimators that can be constructed by means of the general method of Section 1.
ASSUMPTION 3. The $k \times s$-matrix

$$
\begin{pmatrix}
\frac{\partial A_1(\theta)}{\partial \theta_1} & \cdots & \frac{\partial A_s(\theta)}{\partial \theta_1} \\
\vdots & \ddots & \vdots \\
\frac{\partial A_1(\theta)}{\partial \theta_k} & \cdots & \frac{\partial A_s(\theta)}{\partial \theta_k}
\end{pmatrix}
$$

has rank $k$ for every $\theta \in \Theta$.

ASSUMPTION 4. There exists a symmetric, positive definite $s \times s$-matrix $d(\theta)$ such that

$$
\dot{A}(\theta)d(\theta)a(\theta) = \dot{A}(\theta)
$$

where $a(\theta) = (a_{ij}(\theta))$ is the covariance matrix of each $Z_\alpha$.

Assumption 4 generalizes the usual condition that $a(\theta)$ is positive definite. If $a(\theta)$ is positive definite one may choose $d(\theta) = a(\theta)^{-1}$. The generalization makes the treatment of the multinomial distribution very simple (see Section 2). The observation that Assumption 4 is sufficient is due to Ferguson [5].

In a previous paper [4; Theorem 1], we have proved that if $Z_\alpha$, $\alpha = 1, 2, \ldots$ satisfy the Assumptions 1 to 4, there exists a consistent, asymptotically normal estimator of $\theta$. This estimator depends on $Z_\alpha$, $\alpha = 1, 2, \ldots$ only through

$$
\hat{Z}_n = (\hat{Z}_{n1}, \ldots, \hat{Z}_{ns}) = \frac{1}{n} \sum_{\alpha=1}^{n} Z_\alpha
$$

on a neighborhood $S$ of $R = A(\Theta) = \{ A(\theta); \theta \in \Theta \}$. The estimator is continuously differentiable with respect to $\hat{Z}_n$ on $S$.

We now make the following
**DEFINITION.** The matrix $A$ is greater than or equal to $B$ $(A \succeq B)$ if $A - B$ is positive semi-definite.

If $A$ and $B$ are two covariance matrices of the same order and $A \succeq B$, it follows that the variances of $A$ is greater than or equal to the corresponding variances of $B$.

In the aforementioned paper it was also proved [4; Theorem 2] that any consistent estimator being a function of $\hat{Z}_n$ on $S$ and continuously differentiable on $S$ has an asymptotic covariance matrix greater than or equal to

$$
\frac{1}{n} [\Lambda(\theta) d(\theta) \Lambda(\theta)]^{-1}.
$$

Thus the problem is to construct estimators with this asymptotic covariance matrix. The following theorem gives one particular method to construct such estimators.

**THEOREM.** Let the functions $f = (f_1, \ldots, f_s)$ on $\mathbb{R}^S \times \mathbb{R}$ to $\mathbb{R}^s$, $g = (g_1, \ldots, g_s)$ on $\mathbb{R}^S$ to $\mathbb{R}^s$ and $c_{ij}$, $i,j = 1, \ldots, s$ on $\mathbb{R}^S \times \mathbb{R}$ to $\mathbb{R}^1$ satisfy the conditions below. Define a quadratic form $Q$ by

$$
Q(\hat{Z}_n, \theta) = [g(\hat{Z}_n) - f(\hat{Z}_n, \theta)] c(\hat{Z}_n, \theta) [g(\hat{Z}_n) - f(\hat{Z}_n, \theta)]'
$$

where $c(\hat{Z}_n, \theta) = (c_{ij}(\hat{Z}_n, \theta))$. Then

(i) as $n \to \infty$, there exists, with a probability tending to 1, one and only one function $\hat{\theta}(\hat{Z}_n)$ which locally minimizes the quadratic form $Q(\hat{Z}_n, \theta)$;

(ii) $\hat{\theta}(\hat{Z}_n)$ is a consistent estimator of $\theta^0$, the true parameter;
(iii) $\hat{\theta}(\hat{Z}_n)$ is continuously differentiable with respect to $\hat{Z}_n$.

(iv) $\sqrt{n}[\hat{\theta}(\hat{Z}_n) - \theta^0]$ is asymptotically normal distributed with mean 0;

(v) $\hat{\theta}(\hat{Z}_n)$ has asymptotic covariance matrix

$$\frac{1}{n} A(\theta^0) d(\theta^0) A(\theta^0)' - 1.$$

The conditions on $f$ are:

(a) $f$ may be differentiated twice with respect to $\theta$ and once with respect to $\hat{Z}_n$, and the derivatives are all simultaneously continuous in $\hat{Z}_n$ and $\theta$;

(b) $\mathbf{f}(\theta) = \begin{pmatrix}
\frac{\partial f_1(\hat{Z}_n, \theta)}{\partial \theta_1} & \ldots & \frac{\partial f_s(\hat{Z}_n, \theta)}{\partial \theta_1} \\
\frac{\partial f_1(\hat{Z}_n, \theta)}{\partial \theta_1} & \ldots & \frac{\partial f_s(\hat{Z}_n, \theta)}{\partial \theta_k} \\
\frac{\partial f_1(\hat{Z}_n, \theta)}{\partial \theta_k} & \ldots & \frac{\partial f_s(\hat{Z}_n, \theta)}{\partial \theta_k}
\end{pmatrix}$

is of rank $k$ for every $\theta \in \Theta$; and

(c) $\begin{pmatrix}
\frac{\partial f_i(\hat{Z}_n, \theta)}{\partial \hat{Z}_n} \\
\frac{\partial f_i(\hat{Z}_n, \theta)}{\partial \theta_j}
\end{pmatrix} = 0$ for every $i, j = 1, \ldots, s$

The conditions on $g$ are:

(d) $g$ is continuously differentiable with respect to $\hat{Z}_n$;

(e) $g(A(\theta)) = f(A(\theta), \theta)$ for every $\theta \in \Theta$;

(f) $g(A(\theta')) = f(A(\theta'), \theta)$ for every $\theta' \neq \theta$. 
(g) \[ \dot{g}(\theta) = \begin{pmatrix} \frac{\partial g_1(\hat{z}_n)}{\partial \hat{z}_n} & \cdots & \frac{\partial g_s(\hat{z}_n)}{\partial \hat{z}_n} \\ \frac{\partial g_1(\hat{z}_n)}{\partial \hat{z}_{ns}} & \cdots & \frac{\partial g_s(\hat{z}_n)}{\partial \hat{z}_{ns}} \end{pmatrix} \Lambda(\theta) \]

is non-singular for every \( \theta \in \mathbb{H} \); and

(h) \( f(\theta) = \dot{A}(\theta) g(\theta) \)

The conditions on \( c = (c_{ij}) \) are:

(i) \( c_{ij}, i, j = 1, \ldots, s \) may be differentiated twice with respect to \( \theta \) and once with respect to \( \hat{z}_n \), and the derivatives are simultaneously continuous in \( \hat{z}_n \) and \( \theta \);

(j) \( c \) is symmetric and positive definite for every \( (\hat{z}_n, \theta) \in \mathbb{R}^s \times \mathbb{H} \); and

(k) \( g(\theta) c(\theta) g(\theta)' = d(\theta) \) for every \( \theta \in \mathbb{H} \), where \( c(\theta) = (c_{ij}(A(\theta), \theta)) \).

The proof of this theorem is similar to the proof of Theorem 6 (and Theorem 2) in Chiang [3], and the reader may be referred to that paper for the details.

In the definition of \( Q \) and in the conditions imposed on \( f, g \) and \( c \) we have implicitly assumed that the various requirements hold for all \( \hat{z}_n \) and \( \theta \). It suffices, however, to require that \( Q \) is defined on and that (a) - (h) hold in some open set containing \( (A(\theta^0), \theta^0) \) and that (i) - (k) hold in some open set containing \( A(\theta^0) \), because \( \hat{z}_n \) converges to \( A(\theta^0) \) in probability.
and because we are only dealing with asymptotic properties.

The extension made in relation to Ferguson [5] and Chiang [3] is that we allow the function \( f \) in the quadratic form to depend on both \( \hat{Z}_n \) and \( \theta \), requiring only that \( f(A(\theta),\theta) = g(A(\theta)) \).

This added freedom gives us a convenient method for finding an \( f(\hat{Z}_n,\theta) \) in applying the theorem (this method was first proposed by Stone [6]): For simplicity, let \( g(\hat{Z}_n) = \hat{Z}_n \), then \( f(A(\theta),\theta) \) must be equal to \( A(\theta) \). Let \( \theta^*(\hat{Z}_n) \) be a consistent, continuously differentiable estimator of \( \theta \). Expanding \( A(\theta) \) in a Taylor series about \( \theta = \theta^*(\hat{Z}_n) \), we obtain for the two first terms

\[
A(\theta) \approx A(\theta^*(\hat{Z}_n)) + (\theta - \theta^*(\hat{Z}_n)) A'(\theta^*(\hat{Z}_n)).
\]

We then propose to choose

\[
f(\hat{Z}_n,\theta) = A(\theta^*(\hat{Z}_n)) + (\theta - \theta^*(\hat{Z}_n)) A'(\theta^*(\hat{Z}_n)) ,
\]

and one easily verifies that this choice of \( f \) satisfies the conditions of the theorem if \( A(\theta) \) may be differentiated twice. This particular \( f \) leads to linear equations if \( c \) depends only on \( \hat{Z}_n \).

An analogous generalization is possible for the linear form of Ferguson [5].

Let \( f \) and \( g \) satisfy the condition (a) - (h) (actually, condition (a) of \( f \) may be replaced by (a)' \( f \) is continuously differentiable with respect to \( (\hat{Z}_n,\theta) \)), and let \( b(\hat{Z}_n,\theta) = \begin{bmatrix} b_{i} \end{bmatrix} \) be a kxs-matrix with real elements satisfying the conditions

\[(1) \quad b_{i} , \quad i = 1, \ldots, k , \quad i=1,\ldots,s \quad \text{is continuously differentiable with respect to} \quad (\hat{Z}_n,\theta) ;\]
(m) \( b(\theta) f(\theta) \) is non-singular for every \( \theta \in \Theta \), where
\[
b(\theta) = b(A(\theta), \theta);
\]
and
(n) \( b(\theta) g(\theta)' = A(\theta)d(\theta) \)

Then, as \( n \to \infty \), there exists with a probability tending to 1, one and only one function \( \hat{\theta}(\hat{Z}_n) \) which locally solves the equation
\[
L(\hat{Z}_n, \theta) = b(\hat{Z}_n, \theta) \left[ g(\hat{Z}_n) - f(\hat{Z}_n, \theta) \right]' = 0.
\]
\( \hat{\theta}(\hat{Z}_n) \) is a consistent estimator of \( \theta^0 \) and continuously differentiable with respect to \( \hat{Z}_n \). Further \( \hat{\theta}(\hat{Z}_n) \) is asymptotically normal distributed with covariance matrix
\[
\frac{1}{n} \left[ A(\theta^0)d(\theta^0)A(\theta^0)' \right]^{-1}
\]

The proof of this statement may be carried out similar to the proof of the theorem. The essential point is the use of the implicit function theorem.

The definition of \( L \) and the conditions made on \( f, g \) and \( b \) need only be satisfied "locally", as the remarks made in connection with the above theorem also apply in this case.

2. APPLICATION TO THE MULTINOMIAL DISTRIBUTION

To illustrate the theorem of Section 1, we shall give an application to the multinomial distribution. Consider a multinomial experiment of \( s \) mutually exclusive events \( R_1, \ldots, R_s \). For each \( i \) the probability of \( R_i \) is \( p_i(\theta), p_i(\theta) > 0 \). Let
\[
Z_{\alpha}^i = \begin{cases} 
1 & \text{if } R_i \text{ occurs in the } \alpha \text{-th trial} \\
0 & \text{if } R_i \text{ does not occur in the } \alpha \text{-th trial}
\end{cases}
\]
\( i = 1, \ldots, s \), \( \alpha = 1, 2, \ldots \)
Using the notation already introduced we have

\[ Z_\alpha = (Z_\alpha^1, \ldots, Z_\alpha^s) \quad \alpha = 1, 2, \ldots \]

\[ A(\theta) = (p_1(\theta), \ldots, p_s(\theta)), \quad \sum_{i=1}^s p_i(\theta) = 1 \]

\[ a(\theta) = (\sum_{i=1}^s p_i(\theta) - p_1(\theta)p_j(\theta)) \]

Because of the condition \( \sum_{i=1}^s p_i(\theta) = 1 \) we have that \( a(\theta) \) is singular, but

\[ d(\theta) = \begin{pmatrix} p_1(\theta)^{-1} & 0 \\ \vdots & \ddots \\ 0 & \ddots & p_s(\theta)^{-1} \end{pmatrix} \]

satisfies the Assumption 4.

It is tacitly understood that Assumptions 2 and 3 are satisfied.

If we generate estimators of \( \theta \) by minimizing \( Q(Z_n, \theta) \) or solving \( L(Z_n, \theta) = 0 \) we get consistent, asymptotically normal estimators with covariance matrix

\[ \frac{1}{n} [A(\theta)d(\theta)A(\theta)]^{-1} = \frac{1}{n} \left[ \sum_{i=1}^s \frac{1}{p_i(\theta)} \frac{\partial p_i(\theta)}{\partial \theta_{\lambda}} \frac{\partial p_i(\theta)}{\partial \theta_{\alpha}} \right]^{-1} \]

which is also the Cramér-Rao lower bound for estimators of \( \theta \), viz.

\[ \frac{1}{n} \left[ \mathbb{E} \left( \frac{\partial \ln p(Z_{\alpha}; \theta)}{\partial \theta_{\lambda}} \frac{\partial \ln p(Z_{\alpha}; \theta)}{\partial \theta_{\alpha}} \right) \right]^{-1} \]

where

\[ p(Z_{\alpha}; \theta) = \prod_{i=1}^s p_i(\theta) Z_{\alpha_i} \]
(It has been shown ( [1] and [4] ) that if the common probability density of \( Z_x, \alpha = 1, 2, \ldots \), \( p(\cdot; \theta) \), satisfies some additional regularity conditions, estimators generated by one of these methods have asymptotic covariance matrix equal to the Cramér-Rao lower bound if and only if \( p(\cdot; \theta) \) is of the exponential type, i.e.

\[
p(z; \theta) = \exp \left[ \xi_0(\theta) + \Phi(z) + \sum_{i=1}^{s} \xi_i(\theta) z_i \right].
\]

We may, if \( p_i(\theta), i = 1, \ldots, s \) is twice continuously differentiable, for example choose

\[
\begin{align*}
  f_i(\hat{Z}_n, \theta) &= p_i(\theta) \\
g_i(\hat{Z}_n) &= \hat{Z}_{ni} \\
c_{ij}(\hat{Z}_n, \theta) &= \frac{\partial p_i(\theta)}{\partial \theta} \bigg|_{\theta = \theta} \quad i, j = 1, \ldots, s
\end{align*}
\]

in the quadratic form \( Q \), or

\[
\begin{align*}
  f_i(\hat{Z}_n, \theta) &= p_i(\theta) \\
g_i(\hat{Z}_n) &= \hat{Z}_{ni} \\
c_{ij}(\hat{Z}_n, \theta) &= \sum_{i,j} \hat{Z}_{nj}^{-1} \quad i, j = 1, \ldots, s
\end{align*}
\]

getting respectively the minimum-\( \chi^2 \) - estimators and the minimum-modified-\( \chi^2 \) - estimators.

This is easily generalized to the case of several independent multinomial experiments depending on the same basic parameter \( \theta \).

In this case special choices of \( f, g \) and \( c \) will give the result of Taylor [3; Theorem p.88] and thus, as pointed out by Taylor, Berkson's minimum logit \( \chi^2 \)-method [2].
3. APPLICATION TO THE GAMMA-DISTRIBUTION

We shall use the results of Section 1 to propose some new estimators of the unknown parameters in the gamma-distribution.

Let $X_j, j=1,2,...$ be independent, identically distributed random variables with probability density

$$p(x; \alpha, \sigma) = \frac{1}{\Gamma(\alpha) \sigma^\alpha} x^{\alpha-1} e^{-\frac{x}{\sigma}}, \quad x > 0, \quad \alpha, \sigma > 0$$

with respect to the Lebesque measure.

The maximum likelihood estimators of $\alpha$, $\sigma$ are found as the solution of the equations

$$\psi(\alpha^*) + \ln \sigma^* = \hat{z}_{n2}$$

$$\alpha^* \sigma^* = \hat{z}_{n1}$$

where

$$\hat{z}_{n1} = \frac{1}{n} \sum_{j=1}^{n} X_j, \quad \hat{z}_{n2} = \frac{1}{n} \sum_{j=1}^{n} \ln X_j$$

and $\psi'(\alpha) = \frac{\Gamma'(\alpha)}{\Gamma(\alpha)}$.

$(\alpha^*, \sigma^*)$ is asymptotically normal distributed with mean $(\alpha, \sigma)$ and covariance matrix

$$\frac{1}{n} \frac{1}{\alpha \psi'(\alpha)-1} \left( \begin{array}{cc} \alpha & -\sigma \\ -\sigma & \psi'(\alpha) \sigma^2 \end{array} \right).$$

(See for example Sverdrup [7; Ch. XIII, p.120]).
We get another set of estimators \((\hat{\alpha}, \hat{\sigma})\) by solving the equations

\[
EX = \frac{1}{n} \sum_{j=1}^{n} x_j = \bar{x}
\]

\[
\text{var } X = \frac{1}{n} \sum_{j=1}^{n} (x_j - \bar{x})^2 = S^2.
\]

One easily finds

\[
EX = \alpha \sigma \quad \text{and} \quad \text{var } X = \alpha \sigma^2,
\]

thus we get

\[
\hat{\alpha} = \frac{\bar{x}^2}{S^2}, \quad \hat{\sigma} = \frac{S^2}{\bar{x}}
\]

It may be shown that \((\hat{\alpha}, \hat{\sigma})\) is asymptotically normal distributed with mean \((\alpha, \sigma)\) and covariance matrix

\[
\frac{1}{n} \begin{pmatrix}
2\alpha (\alpha + 1) & -2(\alpha + 1)\sigma \\
-2(\alpha + 1)\sigma & 2\alpha + 3 - \sigma^2
\end{pmatrix}
\]

The asymptotic variances of \(\alpha^*\) and \(\sigma^*\) are smaller than the asymptotic variances of \(\hat{\alpha}\) and \(\hat{\sigma}\) (this will be shown later). This suggests that \((\alpha^*, \sigma^*)\) should be preferred as estimator since we do not know any non-asymptotic properties of \((\alpha^*, \sigma^*)\) or \((\hat{\alpha}, \hat{\sigma})\). However, a Monte-Carlo experiment has suggested for some finite \(n\) and special values of \((\alpha, \sigma)\) that the estimator \((\hat{\alpha}, \hat{\sigma})\) gives values closer to the real parameter than \((\alpha^*, \sigma^*)\).

Introducing

\[
\hat{z}_{n3} = \frac{1}{n} \sum_{j=1}^{n} x_j^2
\]

we may write

\[
(2) \quad \hat{\alpha} = \frac{\hat{z}_{n1}}{\hat{z}_{n3}}, \quad \hat{\sigma} = \frac{\hat{z}_{n3} - \hat{z}_{n1}}{\hat{z}_{n1}}
\]
Thus \((\alpha^*, \sigma^*)\) depends on \((\hat{Z}_{n1}, \hat{Z}_{n2})\) and \((\alpha, \sigma)\) depends on 
\((\hat{Z}_{n1}, \hat{Z}_{n3})\). We shall propose an estimator depending on 
\(\hat{Z}_n = (\hat{Z}_{n1}, \hat{Z}_{n2}, \hat{Z}_{n3})\), hoping it will have the asymptotic properties of 
\((\alpha^*, \sigma^*)\) and the finite properties of \((\alpha, \sigma)\).

Put
\[ Z_j = (X_j, \ln X_j, X_j^2), \]
then \(Z_j, j=1,2,\ldots\) is a sequence of independent, identically distributed random vectors. In order to apply the theorem of Section 1 we need the first and second order moments of \(Z = (X, \ln X, X^2)\) (deleting the index \(j\)). We know that
\[
\Gamma (\alpha) \sigma^\alpha = \int_0^\infty x^{\alpha-1} e^{-x/\sigma} \, dx,
\]
differentiating on both sides gives
\[
\Gamma' (\alpha) \sigma^\alpha + \Gamma (\alpha) \sigma^\alpha \ln \sigma = \int_0^\infty \ln x \, x^{\alpha-1} e^{-x/\sigma} \, dx,
\]
and once more
\[
\Gamma'' (\alpha) \sigma^\alpha + 2 \Gamma' (\alpha) \sigma^\alpha \ln \sigma + \Gamma (\alpha) \sigma^\alpha (\ln \sigma)^2 = \int_0^\infty (\ln x)^2 \, x^{\alpha-1} e^{-x/\sigma} \, dx.
\]
Substituting \(\alpha + r\) for \(\alpha\) in (1) and dividing by \(\Gamma (\alpha) \sigma^\alpha\) gives
\[
EX^r = \frac{\Gamma(\alpha+r)}{\Gamma(\alpha)} \sigma^r.
\]
In the same way we obtain from (4)
\[
EX^r \ln X = \left(\frac{\Gamma' (\alpha+r)}{\Gamma (\alpha)} + \frac{\Gamma (\alpha+r)}{\Gamma (\alpha)} \ln \sigma\right) \sigma^r,
\]
and from (5)
\[
EX^r (\ln X)^2 = \left(\frac{\Gamma'' (\alpha+r)}{\Gamma (\alpha)} + 2 \frac{\Gamma' (\alpha+r)}{\Gamma (\alpha)} \ln \sigma + \frac{\Gamma (\alpha+r)}{\Gamma (\alpha)} (\ln \sigma)^2\right) \sigma^r.
\]
These formulae give us

\[ A(\alpha, \sigma) = (\alpha \sigma, \psi(\alpha) + \ln \sigma, \alpha(\alpha+1)\sigma^2), \]

and

\[ a(\alpha, \sigma) = \begin{pmatrix} \alpha \sigma^2 & \sigma & 2\alpha(\alpha+1)\sigma^3 \\ \sigma & \psi'(\alpha) & (2\alpha+1)\sigma^2 \\ 2\alpha(\alpha+1)\sigma^3 & (2\alpha+1)\sigma^2 & 2\alpha(\alpha+1)(2\alpha+3)\sigma^4 \end{pmatrix}. \]

Differentiating \( A(\alpha, \sigma) \) we get

\[ \dot{A}(\alpha, \sigma) = \begin{pmatrix} \sigma & \psi'(\alpha) & (2\alpha+1)\sigma^2 \\ \sigma & 1/\sigma & 2\alpha(\alpha+1)\sigma \end{pmatrix}. \]

Assumptions 1 to 3 of Section 1 are easily verified to hold. The covariance matrix \( a(\alpha, \sigma) \) is positive definite for all \((\alpha, \sigma)\); if not there would have been a linear relation between \( X, \ln X \) and \( X^2 \) with probability 1 (Sverdrup [7; Ch. XII, p.16]). We may therefore choose \( d(\alpha, \sigma) = a(\alpha, \sigma)^{-1} \) in Assumption 4. Thus, applying one of the methods in Section 1, the constructed estimators will have asymptotic covariance matrix

\[ \frac{1}{n} \left[ \dot{A}(\alpha, \sigma)d(\alpha, \sigma)\dot{A}(\alpha, \sigma)' \right]^{-1}. \]

To evaluate this matrix it is necessary to calculate

\[ d(\alpha, \sigma) = a(\alpha, \sigma)^{-1}: \]

\[ \frac{1}{|a(\alpha, \sigma)|} \begin{vmatrix} [2\alpha(\alpha+1)(2\alpha+3)\psi'(\alpha) - (2\alpha+1)^2] & -4\alpha(\alpha+1)\sigma^5 & [(-2\alpha(\alpha+1)\psi'(\alpha) + (2\alpha+1)] \sigma^4 \\ -4\alpha(\alpha+1)\sigma^5 & 2\alpha^2(\alpha+1)\sigma^6 & \alpha\sigma^4 \\ (-2\alpha(\alpha+1)\psi'(\alpha) + (2\alpha+1)]\sigma^3 & \alpha\sigma^4 & [(\alpha\psi'(\alpha) - 1)\sigma^2 \end{vmatrix} \]
where $|a(\alpha, \sigma)| = \alpha[2\alpha(\alpha+1)\psi'(\alpha) - (2\alpha+3)]\sigma^6$. This gives

$$\frac{1}{n} \left[ \hat{A}(\alpha, \sigma) \hat{d}(\alpha, \sigma) \hat{A}(\alpha, \sigma)' \right]^{-1} = \frac{1}{\alpha \psi'(\alpha)} \begin{pmatrix} -\sigma & \sigma \\ -\sigma & \psi'(\alpha) \sigma^2 \end{pmatrix},$$

which is the same asymptotic covariance matrix as the one given above for the maximum likelihood estimators.

**Remark.** We are now able to show, as previously stated, that the asymptotic variances of $\hat{\alpha}$ and $\hat{\sigma}$ are greater than the asymptotic variances of $\alpha^*$ and $\sigma^*$. Since $a(\alpha, \sigma)$ is positive definite, it follows that

$$|a(\alpha, \sigma)| = \alpha[2\alpha(\alpha+1)\psi'(\alpha) - (2\alpha+3)]\sigma^6 > 0,$$

and therefore

$$\psi'(\alpha) > \frac{2\alpha+3}{2\alpha(\alpha+1)}.$$

Thus the ratios between the asymptotic variances are

$$\frac{1}{n} \frac{2\alpha(\alpha+1)}{\alpha} = 2\alpha(\alpha+1)\psi'(\alpha)-2(\alpha+1)>(2\alpha+3)-2(\alpha+1) = 1$$

and

$$\frac{1}{n} \frac{2\alpha+3}{\alpha} \sigma^2 = (2\alpha+3) - \frac{2\alpha+3}{\psi'(\alpha)} > (2\alpha+3) - 2(\alpha+1) = 1.$$

At this point we have to choose the functions $f$ and $g$ satisfying the conditions (a) - (h), and the matrix $c$ satisfying (i) - (k) if we want to use the quadratic form $Q$, or the matrix $b$ satisfying (l) - (n) if we want to use the linear form $L$. Each possible choice of these functions will result in estimators which are asymptotically normal distributed with mean $(\alpha, \sigma)$ and covariance matrix (6).
We shall give some examples of specific choices of \( f, g, c \) and \( b \): 

Let 
\[
g(\hat{Z}_n) = \hat{Z}_n \\
f(\hat{Z}_n, \alpha, \sigma) = A(\alpha, \sigma),
\]

then 
\[
\dot{g}(\alpha, \sigma) = I \quad \text{and} \quad \dot{f}(\alpha, \sigma) = \dot{A}(\alpha, \sigma).
\]

It is easily proved that \( f \) and \( g \) satisfy conditions (a) - (h).

To apply \( L \) we need a 2x3 matrix \( b(\hat{Z}_n, \alpha, \sigma) \) satisfying (n):

\[
b(A(\alpha, \sigma), \alpha, \sigma) = \dot{b}(\alpha, \sigma) = \dot{A}(\alpha, \sigma)d(\alpha, \sigma)(\dot{g}(\alpha, \sigma)')^{-1} = \begin{pmatrix} 0 & 1 & 0 \\ \frac{1}{\sigma^2} & 0 & 0 \end{pmatrix}
\]

We simply choose 
\[
b(\hat{Z}_n, \alpha, \sigma) = \begin{pmatrix} 0 & 1 & 0 \\ \frac{1}{\sigma^2} & 0 & 0 \end{pmatrix}
\]

The conditions (1) and (m) are now satisfied. The solution of

\[
L(\hat{Z}_n, \alpha, \sigma) = b(\hat{Z}_n, \alpha, \sigma)[\hat{Z}_n - A(\alpha, \sigma)]' = 0
\]

with respect to \((\alpha, \sigma)\) are estimators with the stated properties.

Because of the simple form of \( b(\hat{Z}_n, \alpha, \sigma) \), the equation above is equivalent to

\[
\hat{Z}_{n2} = A_2(\alpha, \sigma) \\
\hat{Z}_{n1} = A_1(\alpha, \sigma)
\]

Thus we get the estimators \((\hat{\alpha}, \hat{\sigma})\) as the solution of

\[
\psi(\hat{\alpha}) + \ln \hat{\sigma} = \hat{Z}_{n2} \\
\hat{\alpha} \quad \hat{\sigma} = \hat{Z}_{n1}
\]

which is seen to be the maximum likelihood estimators (1).
Another possibility is to choose
\[ g(\hat{Z}_n) = \hat{Z}_n, \]
\[ f(\hat{Z}_n, \alpha, \sigma) = A(\hat{\alpha}, \hat{\sigma}) + (\alpha - \hat{\alpha}, \sigma - \hat{\sigma}) \dot{A}(\hat{\alpha}, \hat{\sigma}). \]

\[ f(\hat{Z}_n, \alpha, \sigma) \]

is obtained as mentioned after the theorem in Section 1; we have evaluated \( A(\alpha, \sigma) \) "about" the estimator (2). One may verify that the conditions (a) - (h) are satisfied, and that
\[ c(\hat{Z}_n, \alpha, \sigma) = \hat{d}(\hat{\alpha}, \hat{\sigma}) \]
satisfies conditions (i) - (k). Differentiating
\[ Q(\hat{Z}_n, \alpha, \sigma) = [\hat{Z}_n - f(\hat{Z}_n, \alpha, \sigma)] \hat{d}(\hat{\alpha}, \hat{\sigma}) [\hat{Z}_n - f(\hat{Z}_n, \alpha, \sigma)]' \]
with respect to \((\alpha, \sigma)\) and setting the derivatives equal to 0 we get
\[ -2 \ddot{A}(\hat{\alpha}, \hat{\sigma}) \hat{d}(\hat{\alpha}, \hat{\sigma}) [\hat{Z}_n - f(\hat{Z}_n, \alpha, \sigma)]' = 0. \]

(In fact we would have obtained an equivalent equation using \( b(\hat{Z}_n, \alpha, \sigma) = \dot{A}(\hat{\alpha}, \hat{\sigma}) \hat{d}(\hat{\alpha}, \hat{\sigma}) \) in the linear form L.)

As
\[ \dot{A}(\hat{\alpha}, \hat{\sigma}) \hat{d}(\hat{\alpha}, \hat{\sigma}) = \begin{pmatrix} 0 & 1 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 0 \end{pmatrix}, \]
the equation is equivalent to
\[ \hat{Z}_{n2} = f(\hat{Z}_n, \alpha, \sigma) \]
\[ \hat{Z}_{n1} = f(\hat{Z}_n, \alpha, \sigma) \]

Using the fact that \( \hat{\alpha} \hat{\sigma} = \hat{Z}_{n1} \), we get the estimator \((\hat{\alpha}, \hat{\sigma})\) defined by
\[ \hat{\alpha} = \hat{\alpha} \left[ 1 + \frac{\hat{Z}_{n2} - \psi(\hat{\alpha}) - \ln \hat{\sigma}}{\hat{\alpha} \psi'(\hat{\alpha}) - 1} \right] \]
\[ \hat{\sigma} = \hat{\sigma} \left[ 1 - \frac{\hat{Z}_{n2} - \psi(\hat{\alpha}) - \ln \hat{\sigma}}{\hat{\alpha} \psi'(\hat{\alpha}) - 1} \right]. \]
Substituting \((\hat{\alpha}^*, \hat{\sigma}^*)\) for \((\hat{\alpha}, \hat{\sigma})\) in the expression

\[
P(\hat{Z}_{n1}', \hat{Z}_{n2}', \hat{Z}_{n3}') = \frac{\hat{Z}_{n2}' - \psi(\hat{\alpha}) - \ln \hat{\sigma}}{\alpha \psi'(\alpha)}
\]

we see that \(P = 0\) in this case. Generally it is to be expected that \(P\) is small (at least asymptotically, \(P \to 0\)), thus one may hope that \((\hat{\alpha}, \hat{\sigma})\) is an estimator with finite properties similar to those of \((\hat{\alpha}, \hat{\sigma})\), having at the same time the best possible asymptotic properties. However, the only way to obtain a more definite conclusion concerning the finite properties of the estimators considered in this section, is through numerical calculations.
REFERENCES


