CHAPTER II

ESTIMATING FUNCTIONS FOR
ADDITIVE LATENT VARIABLES
MODELS ASSUMING SMALL ERROR
VARIANCE
2.1 INTRODUCTION

This chapter is devoted to the estimation of the structural parameter $\theta$ in the additive latent variables models, $X = Y + Z$, treating $\eta$ as a nuisance parameter. In Section 2.2, we give some basic definitions and results from the theory of estimating functions, which are used extensively in the thesis. For a brief review of Estimating Function Theory we refer to Godambe and Kale (1991). In Section 2.3, we consider the loss of information about the parameters of interest $\theta$ ignoring $\eta$, due to the unobservability of the latent variable $Y$. We extend the results of Kale (1962a) where he assumes $g(y, \theta)$ belongs to the multi-parameter exponential family and $Z \sim N(0, \eta)$ with $\eta$ known. Our results are true for the general additive latent variables models, where we do not restrict $g(y, \theta)$ to the exponential family and $Z$ not restricted to a normal variable. Further, the results in this section hold good for other latent variables models, where the observable variable is $X = \psi(Y, Z)$ with some mild regularity conditions on $\psi$. Section 2.4 deals with the question of obtaining optimal estimating equation for $\theta$ ignoring $\eta$, when $g(y, \theta)$ belongs to the exponential family and $Z$ is such that $E(Z) = 0$, $Var(Z) = \eta$ and $\eta$ is small but unknown. Since $f(x, \omega)$, the pdf of $X$ is not available in closed form, we use the small order error variance approximation to $f(x, \omega)$ as considered in Chesher (1991). Approximate optimal estimating function for $\theta$ is obtained using the approximation to $f(x, \omega)$. In Section 2.5,
we assume a semi-parametric model for Y in which each of the first four cumulants of Y is specified function of θ and Z remains as in the previous section. Following Godambe and Thompson (1989), we obtain approximate optimal estimating function for θ ignoring η. Results of this chapter are reported in Kale and Sebastian (1992).

2.2 PRELIMINARIES

Let X be a random variable with pdf belonging to the class \( p(x, \omega), x \in X, \omega \in \Omega \) with respect to a \( \sigma \)-finite measure \( \nu \). Note that while X itself may be an nx1 random vector of independent and identically distributed random variables, it can also be, in general, any type of random variable. We assume that \( \omega = (\theta', \eta')', \) where \( \theta = (\theta_1, ..., \theta_m)' \) and \( \eta = (\eta_1, ..., \eta_r)' \).

A vector valued function \( u(x, \omega) = (u_1, ..., u_{m+r}) \) defined on \( X \times \Omega \) is a statistical estimating function for \( \omega \), if \( u(x, \omega) \) is a Borel measurable function for each \( \omega \in \Omega \). An estimating function \( u(x, \omega) \) is a regular unbiased estimating function, if \( E_{\omega}(u) = 0 \) and if \( u(x, \omega) \) satisfies the regularity conditions given in Kale (1985). The equation \( u(x, \omega) = 0 \) is known as a regular unbiased statistical estimating equation and a solution of this equation provides an estimate of \( \omega \). The standardized form of an estimating function \( u(x, \omega) \) is defined as \( u^* (x, \omega) = D_u^{-1} u \), where \( D_u = E \left( \frac{\partial u}{\partial \omega} \right) \), so that when \( \omega \) is scalar valued (i.e., \( r = 0 \) and \( m = 1 \)), we have \( u^* (x, \omega) = u(x, \omega) / E \left( \frac{\partial u}{\partial \omega} \right) \). An estimating function \( u^* (x, \omega) \) is said to be optimal in the class \( U \) of all regular
unbiased estimating functions, if \( u^* \in U \) and \( M(u^*) - M(u^*) \) is nonnegative definite (ndd), \( \forall \omega \in \Omega \) and \( \forall u \in U \), where \( M(u^*) \) denotes the variance-covariance matrix of \( u^* \).

Godambe (1960) showed that the score function \( u^* = \frac{\partial \log p(x, \omega)}{\partial \omega} \) for a real valued parameter \( \omega \) is an optimal estimating function with \( \text{Var}(u^*) = \frac{1}{I(\omega)} \) pointing out that the likelihood equation is an optimal estimating equation for the estimation of \( \omega \), where \( I(\omega) \) denotes the Fisher information contained in the sample. Kale (1962b) extended these results for vector valued \( \omega \) and showed that \( u^* = \frac{\partial \log p(x, \omega)}{\partial \omega} \), the vector score function is an optimal estimating function with variance-covariance matrix \( I^{-1}(\omega) \), where \( I(\omega) \) is the Fisher information matrix.

Definition 2.2.1

An \( m \)-dimensional estimating function \( u(x, \theta) = (u_1, ..., u_m)' \) from \( X \times \Theta \) is said to be a regular unbiased estimating function for \( \theta \) ignoring \( \eta \), if \( u(x, \theta) \) is a Borel measurable function for each fixed \( \theta \in \Theta \), \( E_{\omega}(u) = 0 \) for every \( \omega \in \Omega \), \( u \) depends on \( \omega \) only through \( \theta \) and \( u \) satisfies the regularity conditions given in Chandrashekatar (1983).

Let \( U_m \) denote the class of all regular unbiased estimating functions for \( \theta \) ignoring \( \eta \). The standardized form of \( u \) is defined as
We note that \( u \) need not be an element of \( \mathcal{U}_m \); yet \( u \) is of the form \( A(\omega) u \), where \( u \in \mathcal{U}_m \) and \( A(\omega) \) is a matrix with each element \( \lambda_{ij} \) being not necessarily free from \( \eta \).

An estimating function \( u \in \mathcal{U}_m \) is said to be optimal, if \( M(u) = M(u^*) \) is ndd for every \( u \in \mathcal{U}_m \), and for every \( \omega \in \Omega \). Godambe (1976), for real \( \theta \) and Ferreira (1982) and Chandrasekar (1983), for vector valued \( \theta \) showed how to obtain optimal estimating function for \( \theta \) ignoring \( \eta \). In particular Chandrasekar and Kale (1984) obtained a Cramer-Rao type lower bound for \( M(u) \), \( u \in \mathcal{U}_m \), and showed that \( M(u) - (I_{ee} - I_{\theta\eta} \eta^{-1} I_{\eta\eta} \eta^{-1}) \) is ndd.

Here the Fisher information \( I(\omega) \) is partitioned as

\[
I(\omega) = \begin{bmatrix}
I_{ee} & I_{\theta\eta} \\
I_{\eta\theta} & I_{\eta\eta}
\end{bmatrix},
\]

where \( I_{ee} = E \left\{ -\frac{\partial^2 \log p}{\partial \theta \partial \theta'} \right\} \) and \( I_{\eta\eta} = E \left\{ -\frac{\partial^2 \log p}{\partial \eta \partial \eta'} \right\} \) and

\[
I_{\theta\eta} = I_{\eta\theta} = \left\{ -\frac{\partial^2 \log p}{\partial \theta \partial \eta'} \right\}.
\]

Following Liang (1983), Godambe (1984) and Bhapkar (1990) we have the following definition.
Definition 2.2.2

The Fisher information in $X$ about the parameter of interest $e$, ignoring the nuisance parameter $\eta$ is defined as

$$I^{(\eta)}(e) = I_{ee} - I_{e\eta}^{-1} I_{\eta\eta} I_{e\eta}.$$

(2.2.1)

We note that $I^{(\eta)}(e)$ is the inverse of the lower bound given by Chandrasekar and Kale (1984) for the variance covariance matrix of the standardized estimating function for $e$ ignoring $\eta$.

2.3 LOSS OF INFORMATION

In this section we consider the loss of information about the structural parameter $e$ ignoring $\eta$, due to the unobservability of the latent variable $Y$. Let $x = (x_1, \ldots, x_n)$ be a random sample on $X$ with pdf given by the convolution formulae (1.1.1). In the theory of estimating functions, the likelihood equations provide the optimal estimating equations for the simultaneous estimation of $e$ and $\eta$ in (1.1.1). Under the regularity conditions for $f(x, \omega)$ assumed in Kale (1962b), the following theorem gives the likelihood equation for the parameter $\omega = (e, \eta)$.

Theorem 2.3.1

For the simultaneous estimation of the parameter $\omega = (e, \eta)$ based on the data $x = (x_1, \ldots, x_n)$ the likelihood equations are given by
Proof

Log-likelihood of the sample is given by

\[
\log p(x, \omega) = \sum_{i=1}^{n} \log f(x_i, \omega)
\]

\[
= \sum_{i=1}^{n} \log \left( \int g(y, \theta) h(x_i - y, \eta) \, dy \right)
\]

\[
\frac{\partial \log p(x, \omega)}{\partial \theta} = \sum_{i=1}^{n} \int \frac{\partial \log g(y, \theta)}{\partial \theta} \frac{g(y, \theta) h(x_i - y, \eta)}{f(x_i, \omega)} \, dy
\]

\[
= \sum_{i=1}^{n} \int \frac{\partial \log g(y, \theta)}{\partial \theta} \, k_{X_i}(y, \omega) \, dy,
\]

where \(k_{X_i}(y, \omega)\) denotes the conditional pdf of \(Y\) given \(X = x_i\).

Therefore

\[
\frac{\partial \log p(x, \omega)}{\partial \theta} = \sum_{i=1}^{n} \mathbb{E} \left\{ \frac{\partial \log g(Y, \theta)}{\partial \theta} \mid X = x_i \right\}
\]

Similarly, \(\log p(x, \omega) = \sum_{i=1}^{n} \log \left( \int g(x_i - z, \theta) h(z, \eta) \, dz \right)\)

and thus we have
\[
\frac{\partial \log p(x, \omega)}{\partial \eta} = \sum_{i=1}^{n} E \left\{ \frac{\partial \log h(Z, \eta)}{\partial \eta} \bigg| X = x_i \right\}
\]

**Remark 2.3.1**

Note that, when \(Y\) has a pdf belonging to \(m\) parameter exponential family with pdf

\[
g(y, \theta) = \beta(\theta) \exp \left\{ \sum_{j=1}^{E} \theta_j u_j(y) \right\} v(y), \tag{2.3.2}
\]

then the likelihood equation for \(\theta\) is given by

\[
\sum_{i=1}^{n} E \left\{ u(Y) \bigg| X = x_i \right\} + n \frac{\partial \log \beta(\theta)}{\partial \theta} = 0. \tag{2.3.3}
\]

where \(u(Y) = (u_1(Y), ..., u_m(Y))'\). This fact was noted by Kale (1962a), when \(Z \sim N(0, \eta)\), with \(\eta\) known. We observe that the equation (2.3.3) is true even when \(\eta\) is unknown and \(Z\) is non-normal.

**Theorem 2.3.2**

Let the pdf \(f(x, \omega)\) given by (1.1.1) satisfy the regularity conditions and suppose that \(\theta\) is real. Then we have the inequality \(I_X^{(\eta)}(\theta) \leq I_Y(\theta)\), where \(I_X^{(\eta)}(\theta)\) is the Fisher information in \(f(x, \omega)\) about \(\theta\) ignoring \(\eta\) and \(I_Y(\theta)\) is the Fisher information about \(\theta\) in \(g(y, \theta)\).
Proof

By definition, we have

\[ I_\gamma(\theta) = \int -\frac{\partial^2 \log \mathcal{g}(y,\theta)}{\partial \theta^2} \mathcal{g}(y,\theta) \, dy \]

and

\[ \frac{\partial \log \mathcal{f}(x,\theta)}{\partial \theta} = \int \frac{\partial \log \mathcal{g}(y,\theta)}{\partial \theta} k_\chi(y,\omega)dy, \text{ (by (2.3.1)).} \]

Therefore

\[ \frac{\partial^2 \log \mathcal{f}(x,\omega)}{\partial \theta^2} = \int \frac{\partial \log \mathcal{g}(y,\theta)}{\partial \theta} \frac{\partial k_\chi(y,\omega)}{\partial \theta} dy \]

\[ + \int \frac{\partial^2 \log \mathcal{g}(y,\theta)}{\partial \theta^2} k_\chi(y,\omega) dy. \]

Note that

\[ \frac{\partial k_\chi(y,\omega)}{\partial \theta} = \frac{\partial \log \mathcal{g}(y,\theta)}{\partial \theta} k_\chi(y,\omega) \]

\[ - \frac{\partial \log \mathcal{f}(x,\omega)}{\partial \theta} k_\chi(y,\omega). \]

Thus, we have

\[ \frac{\partial^2 \log \mathcal{f}(x,\omega)}{\partial \theta^2} = \int Q k_\chi(y,\omega)dy \]

\[ = E \{ Q \mid X = x \} \]

where

\[ Q = \left\{ \frac{\partial^2 \log \mathcal{g}(Y,\theta)}{\partial \theta^2} + \left( \frac{\partial \log \mathcal{g}(Y,\theta)}{\partial \theta} \right)^2 - \frac{\partial \log \mathcal{f}(X,\omega)}{\partial \theta} \frac{\partial \log \mathcal{g}(Y,\theta)}{\partial \theta} \right\} \]

22
As pointed out by one of the examiners, Theorem 2.3.2. is true only when $\theta$ and $\eta$ are not functionally related and $\omega$ is an interior point of $\Omega$, which we assume to be the case.
Taking expectations on both sides and noting the fact that

\[
E \left\{ \frac{\partial^2 \log g(Y, \theta)}{\partial \theta^2} + \left( \frac{\partial \log g(Y, \theta)}{\partial \theta} \right)^2 \right\} = 0, \text{ we have}
\]

\[
I_{ee} = E \left\{ \frac{\partial \log f(X, \omega)}{\partial \theta} \frac{\partial \log g(Y, \theta)}{\partial \theta} \right\}
\]

\[
= \rho \sqrt{I_{ee} I_Y(\theta)}, \quad (2.3.4)
\]

where \( \rho = \text{Corr} \left\{ \frac{\partial \log f(X, \omega)}{\partial \theta}, \frac{\partial \log g(Y, \theta)}{\partial \theta} \right\} \).

We note that \( \rho \) is always nonnegative. From (2.3.4) we have \( I_{ee} = \rho^2 I_Y(\theta) \). Therefore

\[
I_Y(\theta) = I_X^{(\eta)}(\theta) = I_Y(\theta) - I_{ee} + \frac{I^2_{\theta \eta}}{I_{\eta \eta}}
\]

\[
= (1 - \rho^2) I_Y(\theta) + \left[ \frac{I^2_{\theta \eta}}{I_{\eta \eta}} \right] \geq 0.
\]

We, now consider the case in which the parameters \( \theta \) and \( \eta \) are vector valued.

**Theorem 2.3.3**

Let \( \theta = (\theta_1, ..., \theta_m)' \) and \( \eta = (\eta_1, ..., \eta_r)' \) and the pdf \( f(x, \omega) \) satisfy the regularity conditions, then \( I_Y(\theta) - I_X^{(\eta)}(\theta) \) is nnd.
Proof

We have

\[ I_{Y}(\theta) - I_{X}(\theta) = I_{Y}(\theta) - I_{\theta \epsilon} + I_{\theta \eta} \quad \eta \quad \eta \theta \]

By the convolution formulae, we have

\[ \frac{\partial \log f(x,\omega)}{\partial \theta_j} = \int \frac{\partial \log g(y,\theta)}{\partial \theta_j} k(x,y) \, dy, \quad j = 1, \ldots, m. \]

Thus,

\[ \frac{\partial^2 \log f(x,\omega)}{\partial \theta_j \partial \theta_k} = E \left\{ \begin{array}{c} Q_{jk} | \ X = x \end{array} \right\} \]

where

\[ Q_{jk} = \]

\[ \frac{\partial^2 \log g(Y,\theta)}{\partial \theta_j \partial \theta_k} + \frac{\partial \log g(Y,\theta)}{\partial \theta_j} \cdot \frac{\partial \log g(Y,\theta)}{\partial \theta_k} - \frac{\partial \log f(X,\omega)}{\partial \theta_j} \cdot \frac{\partial \log g(Y,\theta)}{\partial \theta_k} \]

\[ j, k = 1, \ldots, m. \]

Taking expectations on both sides we have

\[ I_{\theta \epsilon} = \text{Cov} \left[ \frac{\partial \log f(X,\omega)}{\partial \theta}, \frac{\partial \log g(Y,\theta)}{\partial \theta} \right]. \]

Now by Cauchy - Schwartz's inequality, for any two vectors \( a, b \in \mathbb{R}^m \), we have

\[ \text{Cov}^2 \left\{ a' \frac{\partial \log f(X,\omega)}{\partial \theta}, b' \frac{\partial \log g(Y,\theta)}{\partial \theta} \right\} \]
That is,

\[(a' \circ \theta \circ b)^2 \leq a' \circ \theta \circ a \cdot b' \circ \theta \circ b, \text{ for all } a, b \in \mathbb{R}_m.\]

Putting \(a = b\), we have

\[(a' \circ \theta \circ a)^2 \leq a' \circ \theta \circ a \cdot a' \circ \theta \circ \psi_a a\]

i.e., \(a' [\circ \theta \circ \psi_a] \circ \theta \circ a \geq 0\), for all \(a \in \mathbb{R}_m\).

Therefore, \(\psi_a - \circ \theta \circ \psi_a\) is nnd. Also note that \(\circ \theta \circ \psi_a - \circ \theta \circ \psi_a\) is always nnd. Hence the required result.

Remark 2.3.1

It can easily be shown that the proofs of Theorem 2.3.1 and Theorem 2.3.3 hold good for the general latent variables models also in which \(X = \psi(Y, Z)\) under mild regularity conditions on \(\psi\). Suppose that \(Y = \psi_1(X, Z)\) and \(Z = \psi_2(X, Y)\) and \(Y\) and \(Z\) are continuous random variables. The pdf of \(X\) is given by

\[
f(x, \omega) = \int g(y, \omega) h(\psi_2(x, y), \eta) \xi_1(x, y) dy \tag{2.3.5}
\]

\[
= \int g(\psi_1(x, z), \omega) h(z, \eta) \xi_2(x, z) dz
\]

where \(\xi_1(x, y)\) and \(\xi_2(x, z)\) are the Jacobian of transformations corresponding to the transformations \((X, Z) \rightarrow (X, Y)\) and \((Y, Z) \leftarrow (X, Z)\) respectively which we assume to be one to one so that
*Remark 2.3.2*

As pointed out by one of the examiners, Hajek and Sidek (1967) prove a result similar to Theorem 2.3.2 when \( \theta \) is a location parameter.
\[ \frac{\partial \log f(x, \omega)}{\partial \theta} = \int \frac{\partial \log g(y, \theta)}{\partial \theta} \frac{g(y, \theta) h(y, h_x(x, y), \eta)}{f(x, \omega)} dy \]

\[ = E \left\{ \frac{\partial \log g(Y, \theta)}{\partial \theta} \mid X = x \right\} \]

\[ \frac{\partial \log f(x, \omega)}{\partial \eta} = E \left\{ \frac{\partial \log h(Z, \eta)}{\partial \eta} \mid X = x \right\} \]

Also \( I_Y(\theta) - I_X(\eta, \theta) \) is nnd. The proof is same as in Theorem 2.3.3.

**2.4 APPROXIMATE OPTIMAL ESTIMATING FUNCTIONS FOR STRUCTURAL PARAMETERS**

In this section we deal with the question of obtaining optimal estimating equation for the parameter \( \theta \), treating \( \eta \) as a nuisance parameter. We assume that \( Y \) and \( Z \) are scalar continuous r.v.s and the pdf \( h(z, \eta) \) is such that \( E(Z) = 0, \text{Var}(Z) = \eta \) and \( Z \) is symmetric around zero with \( \gamma \eta \) as a scale parameter. Because of the complicated nature of the pdf \( f(x, \omega) \) given by (1.1.1), it is difficult to obtain exact optimal estimating function for \( \theta \) ignoring \( \eta \). Even in the simple case in which \( Z \) is N(0, \( \eta \)), \( \eta \) known and \( g(y, \theta) \) belongs to one parameter exponential family, \( f(x, \omega) \) has rather complicated expressions except in the case in which \( Y \) is normal. We, therefore, obtain approximation to \( f(x, \omega) \) assuming \( \eta \) to be small following Chesher (1991) and then obtain approximate optimal estimating function for \( \theta \) based on
2.4.1 SMALL ERROR VARIANCE APPROXIMATION TO THE PROBABILITY DENSITY FUNCTION OF THE OBSERVABLE VARIABLE

We assume that the observable variable \( X = Y + Z \), where \( Y \) and \( Z \) are independent r.v.'s with absolutely continuous distributions over \( \mathbb{R} \). Let the pdf \( h(z, \eta) \) be such that \( Z \) is symmetric around zero and \( V(Z) = \eta \), which is assumed to be small. This assumption is equivalent to the assumption that the measurement error \( Z \) is small with very high probability. By the inversion formula of the characteristic function, the pdf of \( X \) is given by

\[
f(x, \omega) = \frac{1}{2\pi} \int_{-\infty}^{\infty} e^{-itx} \xi_Y(t) \xi_Z(t) \, dt,
\]

where \( \xi_Y(t) \) and \( \xi_Z(t) \) are the characteristic functions of \( Y \) and \( Z \) respectively. Expanding \( \xi_Z(t) \) in the above equation we have

\[
f(x, \omega) = \frac{1}{2\pi} \int_{-\infty}^{\infty} e^{-itx} \xi_Y(t) \left\{ 1 + \frac{\eta(it)^2}{2} + o(\eta) \right\} \, dt
\]

\[= g(x, \theta) + \frac{\eta}{2} g^{(2)}(x, \theta) + o(\eta)\]

\[= g(x, \theta) \left\{ 1 + \frac{\eta}{2} \frac{g^{(z)}(x, \theta)}{g(x, \theta)} \right\} + o(\eta) \quad (2.4.1)\]

where \( g^{(2)}(x, \theta) = \frac{\delta^2 g(x, \theta)}{\delta x^2} \), which we assume to exist for every \( x \in \mathbb{R}_+ \). Ignoring terms of \( o(\eta) \) from (2.4.1), we have an
approximation to \(f(x, \omega)\) as

\[
f(x, \omega) \cong g(x, \theta) \left[ 1 + \frac{\eta}{2} \frac{g^{(2)}(x, \theta)}{g(x, \theta)} \right]. \tag{2.4.2}
\]

In regions where \(g(x, \theta)\) is concave, the above approximation can be negative. (See Chesher (1991)). Further the approximation need not integrate out to unity. Thus (2.4.2) can not be a proper pdf in most of the cases. So we use the following approximation due to Chesher (1991), given by

\[
f(x, \omega) \cong g(x, \theta) \exp \left\{ \frac{\eta}{2} \frac{g^{(2)}(x, \theta)}{g(x, \theta)} - k(\theta, \eta) \right\}. \tag{2.4.3}
\]

where

\[
\exp(k(\theta, \eta)) = \int g(x, \theta) \exp \left\{ \frac{\eta}{2} \frac{g^{(2)}(x, \theta)}{g(x, \theta)} \right\} \, dx
\]

\[
= E \left\{ \exp \left[ \frac{\eta}{2} \frac{g^{(2)}(Y, \theta)}{g(Y, \theta)} \right] \right\},
\]

which is the moment generating function of the r.v. \(\frac{g^{(2)}(Y, \theta)}{g(Y, \theta)}\) evaluated at \(\eta/2\), which we assume to exist. We note that the approximation given by (2.4.3) is a proper pdf and it differs from \(f(x, \omega)\) by \(o(\eta)\) terms. We use this approximation to obtain an approximate optimal estimating function for \(\theta\) when \(g(y, \theta)\) belongs to an exponential family.
2.4.2 APPROXIMATE OPTIMAL ESTIMATING FUNCTIONS WHEN THE pdf OF Y BELONGS TO AN EXPONENTIAL FAMILY.

We first consider the case in which \( g(y, \theta) \) belongs to one parameter exponential family given by

\[
g(y, \theta) = \beta(\theta) \exp \left( \theta u(y) \right) v(y). \tag{2.4.4}
\]

Assuming \( \eta \) is small and using (2.4.3), an approximation to the pdf of \( X \) is given by

\[
f(x, \omega) = \beta(\theta) e^{-k(\theta, \eta)} v(x) \exp \left\{ \theta u(x) + \frac{\eta^2}{2} [u'(x)]^2 + \frac{\eta}{2} \left[ \log v(x) + \log u'(x) \right] \right\}, \tag{2.4.5}
\]

where

\[
e^{-k(\theta, \eta)} = \int \beta(\theta) \exp \left( \theta u(x) \right) v(x) \exp \left\{ \frac{\eta}{2} \left[ \theta^2 (u'(x))^2 + \theta (u''(x)) + 2 \log v(x) u'(x) \right] + \log v(x) \right\} dx
\]

Approximate likelihood of the data \( x = (x_1, \ldots, x_n) \) is given by

\[
p(x, \theta, \eta) = A(\theta, \eta) \prod_{i=1}^{n} v(x_i) \exp \left\{ \theta T_1(x) + \frac{\eta^2}{2} T_2(x) + \frac{\eta}{2} T_3(x) + \frac{\eta}{2} T_4(x) \right\}. \tag{2.4.6}
\]
where $A(0, \eta) = \beta^0(0) \exp \left\{ - nk(0, \eta) \right\}$.

$$T_1(x) = \sum_{i=1}^{n} u(x_i), \quad T_2(x) = \sum_{i=1}^{n} [u^{(1)}(x_i)]^2$$

$$T_3(x) = \sum_{i=1}^{n} \left\{ u^{(2)}(x_i) + 2 \log^{(1)} v(x_i) u^{(1)}(x_i) \right\}$$

and $T_4(x) = \sum_{i=1}^{n} \left\{ \log^{(2)} v(x_i) + [\log^{(1)} v(x_i)]^2 \right\}$.

**Theorem 2.4.2**

Let the pdf of $Y$ belong to the family given by (2.4.6) and let the approximation (2.4.5) be well defined. Further, we have the following assumptions:

(i) Parameters $e$ and $\eta$ are identifiable in $f(x, \omega)$.

(ii) $E(Z) = 0$ and $V(Z) = \eta$, which is assumed to be small so that the approximation (2.4.5) is valid.

(iii) The r.v.'s $Y$ and $Z$ are free of the parameters $(e, \eta)$.

(iv) $T_1(x) = u(x)$, $T_2(x) = [u^{(1)}(x)]^2$,

$$T_3(x) = \left\{ u^{(2)}(x) + 2 \log^{(1)} h(x) u^{(1)}(x) \right\}$$

and $T_4(x) = \left\{ \log^{(2)} h(x) + [\log^{(1)} h(x)]^2 \right\}$ are linearly independent.

An approximate optimal estimating function for $e$ ignoring $\eta$ is given by
\[ S(\kappa, \theta) = T_1(\kappa) - E \left\{ T_2(\kappa) \mid T_2(\kappa), T_3(\kappa), T_4(\kappa) \right\} \] (2.4.7)

where the conditional expectation is evaluated using the approximated likelihood (2.4.6).

**Proof**

By equation (2.4.6) and assumption (iv) of the theorem, the likelihood of the sample belongs to a curved exponential family and \( T = (T_1, T_2, T_3, T_4) \) is the minimal sufficient statistics for \((\kappa, \eta)\). The pdf of \( T \) is given by

\[
k(t, \kappa, \eta) = A(\kappa, \eta) H(t) \exp \left\{ \kappa t_1 + \frac{\eta^2}{2} t_2 + \frac{\kappa \eta}{2} t_3 + \frac{\eta}{2} t_4 \right\},
\]

where \( H(t) \) is a function of \( t \) independent of \((\kappa, \eta)\). The marginal pdf of \((T_2, T_3, T_4)\) is

\[
m(t_2, t_3, t_4; \kappa, \eta) = A(\kappa, \eta) \exp \left\{ \frac{\eta}{2} t_2 + \frac{\kappa \eta}{2} t_3 + \frac{\eta}{2} t_4 \right\} \]

\[
\nu(\kappa, t_2, t_3, t_4),
\]

where \( \nu(\kappa, t_2, t_3, t_4) = \int \exp(\kappa t_1) H(t) dt_1 \).

Hence the conditional pdf of \( x \) given \((T_2 = t_2, T_3 = t_3, T_4 = t_4)\) is given by

\[
w(x \mid t_2, t_3, t_4; \kappa) = \exp(\kappa t_1) \left[ \nu(\kappa, t_2, t_3, t_4) \right]^{n-1} \frac{\eta^2}{2} \nu(x_1).
\]
Also \( p(x, \theta, \eta) = \omega(x \mid t_2, t_3, t_4; \theta) m(t_2, t_3, t_4; \theta, \eta). \)

Moreover, \( m(t_2, t_3, t_4; \theta, \eta) \) is complete for every fixed \( \theta \in \Theta \) since, in this case the marginal density belongs to a one parameter exponential family and the parameter space contains a one dimensional open interval. Hence, following Godambe (1976) and Kale (1987) the optimal estimating function for \( \theta \) ignoring \( \eta \) is the conditional score function given by

\[
S(x, \theta) = t_1(x) - \mathbb{E}\left\{ T_2(x) \mid T_2(x) = t_2, T_3(x) = t_3, T_4(x) = t_4 \right\}
\]

**Remark 2.4.2**

If the function \( u(y) \) in the above theorem is identically equal to \( y \), then we can avoid condition (1) in the theorem. Note that in this case \( 1/\beta(\theta) \) is the Laplace transform of the function \( v(y) \). Also \( \mathbb{E}(Y \mid \theta) = \frac{-\partial \log \beta(\theta)}{\partial \theta} = \mathbb{E}(X \mid \theta, \eta) \) for all \( (\theta, \eta) \in \Omega \) and \( \text{Var}(X \mid \theta, \eta) = \frac{\partial^2 \log \beta(\theta)}{\partial \theta^2} + \eta \). Therefore \( \mathbb{E}(X \mid \theta_1, \eta_1) = \text{Var}(X \mid \theta_2, \eta_2) \Rightarrow \theta_1 = \theta_2 \) and \( \eta_1 = \eta_2 \), for all \( \omega_1, \omega_2 \in \Omega \), where \( \omega_1 = (\theta_1, \eta_1) \) and \( \omega_2 = (\theta_2, \eta_2) \). Hence \( \theta \) and \( \eta \) are identifiable in \( f(x, \omega) \).

**Remark 2.4.3.**

We note that it is possible to obtain approximate optimal estimating function for \( \theta \) even when there is some linear relationship between \( T_2(x) \), \( T_3(x) \) and \( T_4(x) \). For example, suppose that \( T_2(x) \) and \( T_3(x) \) are linearly dependent, i.e., \( T_2(x) \)
\( \equiv c \, T_3(x) \) for some constant \( c \). In this case, if \( T_1(x), T_2(x) \) and \( T_4(x) \) are linearly independent, then the optimal estimating function reduces to

\[
S(x, \theta) = T_4(x) - E \left\{ T_1(x) \mid T_2 = t_2, T_4 = t_4 \right\}.
\]

**Example 2.4.1**

Let \( g(y, \theta) = \frac{\sin(\pi \theta)}{\pi \theta} e^{ey} \frac{e^y}{(1 + e^y)^2} \), \(-\infty < y < \infty\)

and \( 0 < \theta < 1 \). \hfill (2.4.8)

Trivially, \( g(y, \theta) \) belongs to one parameter exponential family

with \( \beta(\theta) = \frac{\sin(\pi \theta)}{\pi \theta} \), \( u(y) = y \) and \( v(y) = \frac{e^y}{(1 + e^y)^2} \). Assume that

the support of the error variable \( Z \) is free of \( \eta \) and \( \eta \) is small

so that condition (ii) and (iii) of the theorem hold good. By

the Remark 2.4.2, condition (i) holds. Now, consider the

condition (iv) in the theorem. Here, we have

\[
T_1(x) = x, \quad T_2(x) = 1, \quad T_3(x) = \frac{2(1 - e^x)}{1 + e^x}
\]

and

\[
T_4(x) = 1 - \frac{6e^x}{(1 + e^x)^2}
\]

The approximate pdf of \( X \) is given by

\[
f(x, \theta, \eta) \approx \frac{\sin \pi \theta}{\pi \theta} \exp \left\{ -k(\theta, \eta) + \frac{\theta^2 \eta}{2} + \frac{\eta}{2} \right\} \frac{e^x}{(1 + e^x)^2}
\]
\[
\exp \left\{ \exp \left( x + \eta \right) \frac{1-e^x}{1+e^x} - \eta \frac{3e^x}{(1+e^x)^2} \right\}
\]

Here,

\[
\exp(k(\theta, \eta)) = \exp \left\{ \frac{\eta}{2} (e^x + 1) \right\} \int \frac{\sin \pi \eta}{\pi \theta} \frac{e^x}{(1 + e^x)^2} \exp \left\{ \exp \left( x + \eta \right) \frac{1-e^x}{1+e^x} - \eta \frac{3e^x}{(1+e^x)^2} \right\} dx.
\]

Note that

\[
\exp(k(\theta, \eta)) \leq \exp \left\{ \frac{\eta}{2} (e^x + 1) \right\} \int \frac{\sin \pi \eta}{\pi \theta} \frac{e^x}{(1 + e^x)^2} \exp(\exp(x + \eta) dx
\]

\[
= \exp \left\{ \frac{\eta}{2} (e^x + 1)^2 \right\}
\]

Hence \(\exp(k(\theta, \eta))\) is well defined. Therefore, (2.4.2) is a proper pdf and can be considered as an approximation to the pdf of \(X\). By Theorem 2.4.2, an approximate optimal estimating function for \(\theta\) ignoring \(\eta\) is

\[
S(\theta, \eta) = T_1(\theta) - E \left\{ T_2(\theta) \middle| T_3(\theta), T_4(\theta) \right\},
\]

where \(T_j(\theta) = \sum_{i=1}^{n} T(x_i), j = 1, ..., 4.\)

**Example 2.4.2**

Suppose that \(Y\) is normal with unknown mean \(\theta (-\infty < \theta < \infty)\)
and known variance 1, without loss of generality. The pdf of Y belongs to one parameter exponential family with $\beta(\theta) = \exp(-\theta^2/2)$, $u(x) = x$ and $v(x) = \exp(-x^2/2)$. The first three conditions of the Theorem 2.4.2 hold good, but condition (iv) does not. Here we have $T_1(x) = x$, $T_2(x) = 1$, $T_3(x) = x$ and $T_4(x) = x^2$, so that they are not linearly independent. Hence, Theorem 2.4.2 is not applicable in this case. But, in small error variance approximation, we essentially assume that the error variable is approximately normal with zero mean and variance $\eta$ (see; Chesher (1991)). Thus, if $Y \sim N(\theta, \sigma^2)$, then $X \sim AN(\theta, \sigma^2 + \eta)$, where AN denotes approximately normal. Now it is well known that for $N(\theta, \tau)$, $\tau$ known or unknown, $\sum_{i=1}^{n} (x_i - \theta)$ is the optimal estimating function for $\theta$ ignoring $\tau$ and therefore $\sum_{i=1}^{n} (x_i - \theta)$ is, approximately, the optimum estimating function for $\theta$ ignoring $(\eta, \sigma^2)$, whether $\sigma^2$ is known or unknown.

We now extend Theorem 2.4.2 to the multi-parameter case in which the latent variable Y belongs to an m-parameter exponential family with pdf given by

$$g(y, \theta) = \beta(\theta) \exp \left\{ \sum_{i=1}^{m} \theta_i u_i(y) \right\} v(y) \quad (2.4.10)$$

where $\theta = (\theta_1, \ldots, \theta_m)' \in \Theta \subseteq \mathbb{R}^m$.

Note that
\[
\frac{\mathcal{g}^{(2)}(y, \theta)}{\mathcal{g}(y, \theta)} = \frac{\partial^2 \log \mathcal{g}(y, \theta)}{\partial y^2} + \left\{ \frac{\partial \log \mathcal{g}(y, \theta)}{\partial y} \right\}^2
\]

\[
= \sum_{j=1}^{m} \theta_j u_j^{(2)}(y) + \log^{(2)} v(y) + \left\{ \sum_{j=1}^{m} \theta_j u_j^{(1)}(y) + \log^{(1)} v(y) \right\}^2
\]

\[
+ \sum_{j=1}^{m} \left\{ \sum_{j=1}^{m} \theta_j u_j^{(1)}(y) + \log^{(1)} v(y) \right\}^2
\]

Hence, small error variance approximation to the pdf of \( X \) is

\[
f(x, \theta, \eta) \cong \beta(\theta) e^{-k(\theta, \eta)} v(x) \exp \left\{ \sum_{i=1}^{k} \theta_j u_j(x) \right\}
\]

\[
+ \sum_{i=1}^{k} \theta_j \eta \left[ u_j^{(2)}(x) + 2u_j^{(1)}(x) \log^{(1)} v(x) \right] + \sum_{i=1}^{k} \frac{\theta_j \eta^2}{2} \left[ u_j^{(2)}(x) \right]^2
\]

\[
+ \sum_{j=1}^{k} \eta \theta_j u_j^{(1)}(x) \left[ u_j^{(1)}(x) \right] + \frac{\eta}{2} \left[ \log^{(2)} v(x) + \log^{(1)} v(x) \right] \}
\]

(2.4.11)

where \( e^{k(\theta, \eta)} = \int \mathcal{g}(x, \theta) \exp \left\{ \frac{\eta}{2} \frac{\mathcal{g}^{(2)}(x, \theta)}{\mathcal{g}(x, \theta)} \right\} dx \)

Hence approximate likelihood of the sample \( x = (x_1, \ldots, x_n) \) is

\[
p(x, \theta, \eta) = \beta^n(\theta) e^{-n k(\theta, \eta)} \prod_{i=1}^{n} v(x_i) \cdot \exp \left\{ \sum_{j=1}^{k} \theta_j T_j(x) \right\}
\]

\[
+ \frac{\theta_j \eta}{2} \left[ T_j^{(1)}(x) \right] + \sum_{j=1}^{k} \theta_j \eta \left[ T_j^{(1)}(x) \right] + \frac{\eta}{2} \left[ T_j^{(1)}(x) \right] \}
\]

(2.4.12)
where
\[ T_{1j}(x) = \sum_{i=1}^{n} u_{ij}(x) \]

\[ T_{2j}(x) = \sum_{i=1}^{n} \left\{ u_{ij}^{(2)}(x) + 2u_{ij}^{(4)}(x) \log^{(4)} v(x) \right\} \]

\[ T_{3j}(x) = \sum_{i=1}^{n} \left[ u_{ij}^{(1)}(x) \right]^2, \quad j = 1, \ldots, k \]

\[ T_{4j}(x) = \sum_{i=1}^{n} \left\{ u_{ij}^{(4)}(x) u_{ij}^{(1)}(x) \right\}, \quad j < \ell, j, \ell = 1, 2, \ldots, k \]

\[ T_{5}(x) = \sum_{i=1}^{n} \left\{ \log^{(2)} v(x) + \log^{(4)} h(x) \right\}^2 \].

**Theorem 2.4.3.**

Let the pdf of \( Y \) belong to the \( k \)-parameter exponential family (2.4.9) and the approximation (2.4.10) be well defined. Further, we have the following assumptions,

(i) Parameters \( \theta \) and \( \eta \) are identifiable in the pdf of \( X \).
(iii) \( \eta \) is small so that approximation (2.4.10) is valid.
(iv) The r.v.s \( Y \) and \( Z \) are free of \( \theta \) and \( \eta \).

Then an approximate optimal estimating equation for \( \theta \) ignoring \( \eta \) is given by

\[ S(x, \theta) = T(x) - E \left\{ T(x) \mid T_{1}(x), T_{2}(x), T_{4}(x), T_{5}(x) \right\}, \]

where \( T(x) = [T_{1}(x), T_{2}(x), \ldots, T_{r}(x)], r = 1, 2, 3 \) and \( 5 \)
Proof

Proof of the theorem is analogous to that of Theorem 2.4.2.

2.5 ESTIMATING EQUATIONS FOR SEMI-PARAMETRIC LATENT VARIABLES MODELS

In this section we assume that the first four moments (cumulants) of the latent variable $Y$ are known functions of $\theta$ and $E(Z) = 0$, $\text{Var}(Z) = \eta$ and $\eta$ is small. The cumulant generating function of $Z$ can be written as

$$K_Z(t, \eta) = -\frac{t^2 \eta}{2} + o(\eta).$$

Hence, when $\eta$ is small, $N(0, \eta)$ can be considered as an approximate distribution of $Z$ even when $Z$ is non-normal and we therefore, assume that $Z \sim N(0, \eta)$.

The cumulant generating function of $X$ can be written as

$$K_X(t; \theta, \eta) = K_Y(t; \theta) - \frac{t^2 \eta}{2},$$

where $K_Y(t; \theta)$ is the cumulant generating function of $Y$. Hence, cumulants of $X$ are $k_r(X) = k_r(\theta)$, $r = 1, 3, 4, \ldots$ and $k_2(X) = k_2(\theta) + \eta$ where $k_r(\theta)$ denotes the $r$-th cumulant of $Y$. The following estimating equations provide moment estimators of $\theta$ and $\eta$. 

and $T \{X\} = \{T \{X\}, T \{X\}, \ldots, T \{X\}\}$. 

Proof

Proof of the theorem is analogous to that of Theorem 2.4.2.
\[
\bar{X} = n^{-1} \sum_{i=1}^{n} X_i = k_1(\theta) \\
S^2 = n^{-1} \sum_{i=1}^{n} (X_i - \bar{X})^2 = k_2(\theta) + \eta
\]

Moment estimators of \( \theta \) and \( \eta \) are given by

\[
\hat{\theta} = k_1^{-1}(\bar{X}) \quad \text{and} \quad \hat{\eta} = S^2 - k_2[k_1^{-1}(\bar{X})]
\]

proved \( k_1(\theta) \) is an invertable function of \( \theta \).

Now, assume that the first four cumulants of \( Y \) are specified functions of \( \theta \) and \( k_r(\theta) \) is free of \( \theta \), for \( r = 2,3,4 \). Assuming \( \eta \) is small, approximate optimal estimating function for \( \theta \) and \( \eta \) can be obtained. The skewness and kurtosis of \( X \) are given by

\[
\beta_1(X) = \frac{[E(X) - k_1(\theta)]^3}{(k_2 + \eta)^2} = \frac{k_3^2}{(k_2 + \eta)^2} + o(\eta) = \beta_1(y)
\]

\[
\beta_2(X) = \frac{E(X) - k_1(\theta)^4}{(k_2 + \eta)^2} = \frac{k_4}{(k_2 + \eta)^2} + 3 = \frac{k_4}{k_2^2} + 3 + o(\eta) \cong \beta_2(y)
\]
where we denote \( k_r(\theta) = k_r \), for \( r = 2, 3, 4 \). Thus, when \( \eta \) is small, skewness and kurtosis of \( X \) are approximately equal to those of \( Y \). Thus, we have a semi-parametric model for \( X \) with 

\[
E(X) = k_1(\theta), \quad V(X) = k_2 + \eta, \quad \beta_1(X) = \beta_1(Y), \quad \beta_2(X) = \beta_2(Y),
\]

where \( \beta_1(Y) \) and \( \beta_2(Y) \) are free of \( (\theta, \eta) \). Following Godambe and Thompson (1989), the optimum estimating function to estimate \( \theta \) and \( \eta \) for the above semi-parametric model are given by

\[
\xi_1(x, \theta, \eta) = \sum_{j=1}^{n} \xi_{1j}(x, \theta, \eta)
\]

and

\[
\xi_2(x, \theta, \eta) = \sum_{j=1}^{n} \xi_{2j}(x, \theta, \eta),
\]

where

\[
\xi_{1j} = \frac{-h_{1j} \frac{\partial}{\partial \theta} k_1(\theta)}{k_2 + \eta} + \frac{[h_{2j}(\eta)^{1/2} \gamma_1 \frac{\partial}{\partial \theta} k_1(\theta)] - \frac{\partial}{\partial \theta}(k_2 + \eta)}{(k_2 + \eta)^{1/2} (\gamma_2 + 2 - \gamma_1^2)}
\]

and

\[
\xi_{2j} = \frac{-h_{2j} \frac{\partial}{\partial \eta} (k\eta)}{(k_2 + \eta)^{1/2} (\gamma_2 + 2 - \gamma_1^2)}
\]

where

\[
h_{1j} = x_j - k_1(\theta),
\]

\[
h_{2j} = [x_j - k_1(\theta)]^2 - (k_2 + \eta) - \gamma_1 (k_2 + \eta)^{1/2} [x_j - k_1(\theta)],
\]

40
\[ y_1 = \sqrt{\beta_1} \quad \text{and} \quad y_2 = \beta_2 - 3. \]

On substitution we have

\[ g_1(x, \theta, \eta) = \frac{-k'_1(\theta)}{k_2 + \eta} \sum_{j=1}^{n} (x - k(\theta)) \]

\[ + \frac{[\gamma_1(k_2 + \eta)^{1/2} k'_1(\theta)] - k_2(\theta)}{(k_2 + \eta)^2 (\gamma_2 + 2 - \gamma_1^2)} \left\{ \frac{\sum_{j=1}^{n} (x - k(\theta))^2}{n(k_2 + \eta) - \gamma_1(k_2 + \eta)^{1/2} \sum_{j=1}^{n} (x - k(\theta))} \right\} \]

and \[ g_2(x, \theta, \eta) = \frac{-1}{(k_2 + \eta)^2 (\gamma_2 + 2 - \gamma_1^2)} \left\{ \sum_{j=1}^{n} (x - k(\theta))^2 - n(k_2 + \eta) - \gamma_1(k_2 + \eta)^{1/2} \sum_{j=1}^{n} (x - k(\theta)) \right\} \]

provided \( \gamma_2 + 2 - \gamma_1^2 \) is nonvanishing, which is true whenever \( X \) is not degenerate. The estimators of \( (\theta, \eta) \) can be obtained by solving equations

\[ g_1(x, \theta, \eta) = 0 \]

\[ g_2(x, \theta, \eta) = 0 \]

Now it is easy to show that the solutions of the above equations are the same as (2.5.2). Further, as a direct consequence of the results due to Godambe and Thompson (1978) for the above semi-parametric model for \( X \), the estimating function \( [\bar{X} - k_1(\theta)] \) will be an optimal estimating function for \( \theta \) ignoring \( \eta \).