Asymptotic Minimax Estimation

in Semiparametric Models

By: Byeong Uk Park
University of North Carolina at Chapel Hill

Feb. 1988

Abstract

We give several conditions on the estimator of efficient score function for estimating the parametric component of semiparametric models. We show that a semiparametric version of the one–step MLE using the estimator of efficient score function which fulfills the conditions is asymptotically minimax. A few examples are also considered.

AMS (1980) Subject Classification: 62G05, 62G20.

key words and phrases: Asymptotic minimax bounds, efficient score function, asymptotically minimax estimators, semiparametric models.

1. Introduction

As powerful criteria for identifying good statistical methods, asymptotic minimaxity and efficiency have been used widely in large sample theory. The most outstanding works related to these notions are done by Hájek (1970, 1972) and Le Cam (1972) where representation theorems and asymptotic minimax theorems are presented. Hájek's works are for parametric model but Le Cam's are quite general. More recently representation theorems and asymptotic minimax theorems have been established for a variety of statistical problems including estimation of parametric or nonparametric components in semiparametric models (Begun et al. (1983)) and estimation of a distribution function (Beran (1977), Koshevnik and Levit (1976), Millar (1979)).

Many authors have discussed asymptotic efficient estimation (here "asymptotic efficiency" means that the limiting distribution of an estimator is the normal distribution in representation theorem) in various particular semiparametric models. Those include Weiss and Wolfowitz (1970), Wolfowitz (1974), van Eden (1970), Beran (1974, 1977, 1978), Stone (1975), Efron (1977), Tsiatis (1981), Bickel and Ritov (1987) and Park (1987). The construction of asymptotically efficient estimators in general semiparametric models have been discussed by Bickel (1982), Schick (1986, 1987) and Klaassen (1987). But notwithstanding the importance of the notion of asymptotic minimaxity, the construction of asymptotically minimax estimators has been rather neglected, especially in semiparametric models. A few works include Beran (1981), Fabian and Hannan (1982) and Millar (1984) among which the first and the third treated only parametric models.

Our present theme is asymptotic minimax estimation of parametric components in semiparametric models. Suppose $f(\cdot,\theta,g)$ is our semiparametric density model where θ and g are the parametric and nonparametric component respectively. We take a shrinking neighborhood ($n^{-\frac{1}{2}}$ – rate) around $f^{\frac{1}{2}}(\cdot,\theta,g)$, namely, $\{q: n^{\frac{1}{2}}\|q^{\frac{1}{2}}-f^{\frac{1}{2}}\| \le c\}$ ($\|\cdot\|$ is the usual $L^2(\cdot)$ norm) instead of taking it around θ and g separately as in Begun et al. (1983) because the former is more convenient for our purpose. We present an appropriate asymptotic minimax theorem considering this neighborhood in section 2 and asymptotically minimax estimator over this neighborhood in section 3. Fabian and Hannan (1982) dealt this problem in different mathematical formulation but the most significant difference between this paper and theirs is in that the neighborhood of $f^{\frac{1}{2}}(\cdot,\theta,g)$ they considered is of finite dimension and they treated only the cases in which adaptation is possible.

Our estimator is, in essence, a semiparametric version of the one–step MLE of Le Cam (1956, 1969). In section 3 we give several conditions on the estimator of efficient score function (see Begun et al. (1983) or Bickel et al. (1987) for definition) for asymptotic minimaxity of our estimator. A few examples are considered in section 4 where we see how the conditions in section 3 are satisfied and consequently establish asymptotically minimax estimators for those examples. The proof of our main theorem is given in section 5.

2. Asymptotic Minimax Bound

Suppose that $X_1,\ldots,X_n$ are iid $\mathscr{X}$-valued random variables with density function $f = f(\cdot,\theta,g)$ with respect to a σ–finite measure μ on the measurable space $(\mathscr{X},\mathscr{C})$ where $\theta \in \Theta \subset R^k$ and $g \in \mathscr{G} \subset$ the collection of all densities with respect to a σ–finite measure ν on some measurable space $(\mathscr{Y},\mathscr{D})$. We will find the asymptotic minimax bound for estimating the parametric component θ in the presence of the unknown nuisance parameter g. Our result in this section is a variation of Theorem 3.2 in Begun et al. (1983), which is a special case of the general Hájek–Le Cam–Millar asymptotic minimax theorem (see proposition 2.1 of Millar (1979)).

Let $N_n(f.c)$ be the set of all density functions $g \in \mathscr{G}$ with respect to μ such that $\|n^{\frac{1}{2}}(q^{\frac{1}{2}}-f^{\frac{1}{2}})\| \leq c$ where $\|\cdot\|$ is the usual norm of $L^2(\mu)$. Let

$$\theta(q) = \theta + 4\, I_*^{-1} \int \alpha^*(q^{\frac{1}{2}}-f^{\frac{1}{2}})d\mu \tag{2.1}$$

where $\alpha^* = \rho_\theta - A\beta^*$ and ρ_θ, A, β^* and I_* are the same as defined in Begun et al. (1983). We assume that Assumption S and the conclusion of proposition 2.1 in Begun et al. (1983) hold throughout this paper. Then by the second assumption

$$\theta(f_n) = \theta_n + o(n^{-\frac{1}{2}}) \tag{2.2}$$

where $f_n = f(\cdot,\theta_n,g_n)$ and θ_n and g_n are the same as in proposition 2.1 of Begun et al. (1983). The equation (2.2) tells us that $\theta(q)$ plays a part in identifying the parametric component of q. Hence if q is the underlying density, the loss should be a function of $\hat{\theta}_n-\theta(q)$ where $\hat{\theta}_n$ is an estimate of θ. An interesting motivation of $\theta(q)$ is illustrated in Beran (1981).

Let ℓ be a subconvex loss function (see Begun et al. (1983) for definition). The following theorem gives an asymptotic minimax bound for estimating θ.

Theorem 2.1. Under the assumptions described above,

$$\lim_{c\to\infty} \lim_n \inf \inf_{\hat{\theta}_n} \sup_{q\in N_n(f,c)} E_Q\, \ell(n^{\frac{1}{2}}(\hat{\theta}_n - \theta(q))) \geq E\, \ell(Z_*)$$

where $Z_* \sim N(0, I_*^{-1})$ and Q is a probability measure having density g.

Proof. From Theorem 3.2 of Begun et al. (1983)

$$\lim_{\substack{c_1\to\infty \\ c_2\to\infty}} \lim_n \inf \inf_{\hat{\theta}_n} \sup_{\substack{|h|\leq c_1 \\ \|\beta\|\leq c_2}} E_{P_n}\, \ell(n^{\frac{1}{2}}(\hat{\theta}_n - \theta_n)) \geq E\, \ell(Z_*)$$

where θ_n, h and β are the same as in the paper and P_n is the corresponding probability measure for $f_n = f(\cdot, \theta_n, g_n)$. From (2.2) we can replace $\ell(n^{\frac{1}{2}}(\hat{\theta}_n - \theta_n))$ by $\ell(n^{\frac{1}{2}}(\hat{\theta}_n - \theta(f_n)))$. Now observe that if $|h| \leq c_1$ and $\|\beta\| \leq c_2$, then $\|n^{\frac{1}{2}}(f_n^{\frac{1}{2}} - f^{\frac{1}{2}})\| \leq c$ for some $c > 0$. Hence we can replace f_n with $|h| \leq c_1$ and $\|\beta\| \leq c_2$ by q in $N_n(f,c)$. The theorem follows.

■

In view of Theorem 2.1, our aim in this paper is constructing $\hat{\theta}_n$ which satisfies

$$\lim_{c\to\infty} \lim_n \sup_{q\in N_n(f,c)} E_Q\, \ell(n^{\frac{1}{2}}(\hat{\theta}_n - \theta(q))) = E\, \ell(Z_*). \tag{2.3}$$

If we restrict ℓ to a bounded loss function, it suffices to find $\hat{\theta}_n$ such that $n^{\frac{1}{2}}(\hat{\theta}_n - \theta(q_n))$ converges weakly, under Q_n (probability measure having density q_n), to $N(0, I_*^{-1})$ for any sequence $q_n \in N_n(f,c)$ as discussed in (5.17) of Millar (1984) and at the end of the proof of proposition 1 in Beran (1981). In the next section, we will see how this goal can be achieved.

3. Asymptotic minimax estimation

3.1. Preliminary estimator

First we need an initial estimator which is $\sqrt{n}$–consistent in a slightly different sense than usual, namely, we need an estimator $\tilde{\theta}_n$ of θ such that $n^{\frac{1}{2}}(\tilde{\theta}_n-\theta)$ is tight under any Q_n where $q_n \in N_n(f,c)$. M–estimators and minimum distance type estimators have been used widely for this purpose, but noticing that Q_n^n is not contiguous to P^n, where Q_n^n and P^n are product measures having densities q_n and f respectively, the latter seems preferable. The following examples illustrate this point.

Example 1. (One sample location model). Suppose $f = f(\cdot,\theta,g) = g(\cdot-\theta)$ where g is symmetric, $\theta \in R^1$ and $\int(\dot{g}^2/g)d\mu < \infty$ (μ is Lebesgue measure). Let $\tilde{\theta}_n$ be chosen so as to minimize

$$h(\theta) = \max_x |F_n(x) + F_n((2\theta-x)^-)-1|$$

where F_n is the usual empirical distribution function on a random sample $x_1,\ldots,x_n$ from Q_n with density q_n. The consistency and $\sqrt{n}$–consistency of this estimator have been shown by Schuster and Narvarte (1973) and Rao et al. (1975) under the model density f. But observing that

$$\|F_{2\tilde{\theta}_n-\theta} - F_\theta\|_s \le 4\|F_n-F_\theta\|_s \le 4(\|F_n-Q_n\|_s + \|Q_n-F_\theta\|_s) \tag{3.1}$$

where $\|\cdot\|_s$ is the sup–norm over R, consistency under Q_n follows directly from the continuity of the map $\theta \to F_\theta$ and the identifiability of θ, namely, $F_{\theta_1} = F_{\theta_2}$ implies $\theta_1 = \theta_2$. Similar argument on pp. 106–107 of Le Cam (1969) and the second inequality of (3.1) ensure that $\sqrt{n}(\tilde{\theta}_n-\theta)$ is tight under Q_n.

Example 2. (Two sample shift model). Suppose that $(X_1,Y_1),\ldots,(X_n,Y_n)$ are i.i.d. $g(\cdot)g(\cdot-\theta)$ where $\theta \in R^1$ and g is any unknown density function defined on R^1 with

$\int(\dot{g}^2/g)d\mu < \infty$. If we define $\tilde{\theta}_n$ to minimize

$$h(\theta) = \|F_n(\cdot - \theta) - G_n\|_s$$

where F_n and G_n are the usual empirical distribution functions of X_i's and Y_i's respectively, it can be shown that $\sqrt{n}(\tilde{\theta}_n - \theta)$ is tight under Q_n using similar arguments as in Example 1.

3.2. Asymptotic minimax estimator

Let

$$\hat{\theta}_n = \bar{\theta}_n + n^{-1}\,\hat{I}_*^{-1}\,\Sigma_{j=1}^{n}\ \hat{\ell}^*(X_j, \bar{\theta}_n) \tag{3.2}$$

where $\bar{\theta}_n$ is a discretized version of $\tilde{\theta}_n$, $\hat{\ell}^*(x,\theta)$ is a good estimator of $\ell^*(x,\theta) = 2\alpha^* f^{-\frac{1}{2}}(x,\theta,g)$ and $\hat{I}_* = n^{-1}\,\Sigma_{j=1}^{n}\ \hat{\ell}^*\hat{\ell}^{*T}(X_j,\bar{\theta}_n)$, an estimator of I_*. The asymptotic behavior of $\hat{\theta}_n$ depend heavily on that of $\hat{\ell}^*$. We state several conditions which $\hat{\ell}^*$ should satisfy in order that $\hat{\theta}_n$ defined in (3.1) satisfy (2.3), namely, be an asymptotically minimax estimator. Let $\{\theta_n: n \geq 1\}$ be any sequence such that $|\theta_n - \theta| = O(n^{-\frac{1}{2}})$.

(C1) $n^{\frac{1}{2}} \int \hat{\ell}^*(x,\theta_n)\, f(x,\theta_n,g)d\mu = o_{Q_n}(1)$

(C2) $E_{Q_n} \int |\hat{\ell}^*(x,\theta_n) - \ell^*(x,\theta_n)|^2\, f(x,\theta_n,g)d\mu = o(1)$

Let $\hat{\ell}_j^*$ be a cross validated estimator of ℓ^*, i.e., $\hat{\ell}_j^* = \hat{\ell}^*$ computed from $X_1,\ldots,X_{j-1}$, $X_{j+1},\ldots,X_n$.

(C3) $|\hat{\ell}_j^*(x,\theta) - \hat{\ell}^*(x,\theta)| \leq M_n = o(n^{-\frac{1}{2}})$, $j=1,\ldots,n$ where M_n is a constant

Although we can find $\hat{\ell}^*$ for which (C3) is satisfied in most cases, particularly in our examples considered, the proof of the asymptotic minimaxity of $\hat{\theta}_n$ relies on the following set of weaker conditions than (C3).

(C3.1) $\Sigma_{j=1}^{n} E_{Q_n} \int |\hat{\ell}^*(x,\theta_n) - \hat{\ell}_j^*(x,\theta_n)|^2 q_n \, d\mu = o(1)$

(C3.2) $n^{-1} \Sigma_{j=1}^{n} |\hat{\ell}^*(X_j,\theta_n) - \hat{\ell}_j^*(X_j,\theta_n)|^2 = o_{Q_n}(1)$

(C3.3) $n^{-\frac{1}{2}} \Sigma_{j=1}^{n} \int (\hat{\ell}^*(x,\theta_n) - \hat{\ell}_j^*(x,\theta_n)) q_n \, d\mu = o_{Q_n}(1)$

(C3.4) $n^{-\frac{1}{2}} \Sigma_{j=1}^{n} (\hat{\ell}^*(X_j,\theta_n) - \hat{\ell}_j^*(X_j,\theta_n)) = o_{Q_n}(1)$.

Some of the conditions described above are motivated from an interesting paper by Schick (1987) and they are similar to the conditions in his lemma 3.1 but Schick's lemma is for constructing asymptotically linear estimators and is useful when the underlying probability measure Q_n^n is contiguous to P^n. For more details see Schick (1987). We add two more conditions on $\hat{\ell}^*$.

(C4) $n^{-\frac{1}{2}} \left|\frac{\partial}{\partial\theta} \hat{\ell}^*(x,\theta)\right| \leq N_n \to 0$

(C5) $n^{-\frac{1}{4}} |\hat{\ell}^*(x,\theta)| \leq L_n \to 0$

where N_n and L_n are constants.

Here is our main theorem.

Theorem 3.1. If the conditions (C1)–(C5) are satisfied and $I_* = I_*(\theta)$ is continuous as a function of θ and the map $\theta \to \ell^*(x,\theta)$ is continuous for each $x \in \mathcal{X}$, then

$$n^{\frac{1}{2}}(\hat{\theta}_n - \theta(q_n)) \Rightarrow N(0, I_*^{-1})$$

under Q_n.

Proof. See section 5.

As we discussed it earlier in section 2, the conclusion of Theorem 3.1 implies that $\hat{\theta}_n$

with $\hat{\ell}^*$ satisfying (C1)–(C5) is asymptotically minimax if the loss function is bounded and subconvex.

4. Examples (continued)

In this section we will see how $\hat{\ell}^*(\cdot,\cdot)$ can be constructed to satisfy the conditions (C1)–(C5) in the examples considered in section 3.

4.1 Example 1 (continued)

Note that $\ell^*(x,\theta) = -\dot{g}/g(x-\theta)$ and $I_* = \int(\dot{g}^2/g)d\mu$. Define

$$\hat{g}(x,\theta) = b_n + n^{-1}\, b_n^{-1}\, \Sigma_{j=1}^{n}\, K(b_n^{-1}(x-X_j + \theta))$$

$$\hat{\ell}^*(x,\theta) = -\tfrac{1}{2}\,(\dot{\hat{g}}/\hat{g}(x-\theta,\theta) - \dot{\hat{g}}/\hat{g}(\theta-x,\theta))$$

where $b_n \to 0$, $nb_n^{10} \to \infty$ and K is logistic density. Then

$$\hat{\ell}_j^*(x,\theta) = -\tfrac{1}{2}(\dot{\hat{g}}_j/\hat{g}_j(x-\theta,\theta) - \dot{\hat{g}}_j/\hat{g}_j(\theta-x,\theta))$$

where $\hat{g}_j(x,\theta) = \hat{g}(x,\theta) - n^{-1}\, b_n^{-1}\, K(b_n^{-1}(x-X_j+\theta))$. Now (C1) is obviously satisfied and (C3) can be easily verified (see Schick (1987)). If we note that $|\hat{\ell}^*(x,\theta)| \le b_n^{-1}$ and $|\frac{\partial}{\partial\theta}\hat{\ell}^*(x,\theta)| \le C\, b_n^{-3}$ for some C, (C4) and (C5) are obvious. Let $p_n(x,\theta) = E_{Q_n}\, \hat{g}(x,\theta) - b_n$ and $g_n(x,\theta) = E_P\, \hat{g}(x,\theta) - b_n$. By the same method as in Schick (1987) we can show

$$E_{Q_n} |\dot{\hat{g}}/\hat{g}(x,\theta) - \dot{p}_n(x,\theta)/(p_n(x,\theta)+b_n)|^2 \le 4\, n^{-1}\, b_n^{-6} \to 0$$

and by Bickel (1982) and Schick (1987) it has been shown that

$$\int(\dot{g}_n(x,\theta)/(g_n(x,\theta)+b_n) - \dot{g}/g(x))^2\, g(x)dx \to 0.$$

Hence (C2) is satisfied if we have

(4.2) $$\int(\dot{p}_n(x,\theta)/(p_n(x,\theta)+b_n) - \dot{g}_n(x,\theta)/(g_n(x,\theta)+b_n))^2\, g(x)dx \to 0.$$

but the square term in (4.2) is bounded by $B\, n^{-\frac{1}{2}}\, b_n^{-5}$ for some constant $B > 0$. Hence the conditions (C1)–(C5) are satisfied with $\hat{\ell}^*$ defined in (4.1).

4.2. Example 2 (continued)

Observe that $\ell^*(x,y,\theta) = -\frac{1}{2}\{\dot{g}/g(y-\theta) - \dot{g}/g(x)\}$ and $I_* = \frac{1}{2}\int(\dot{g}^2/g)d\mu$. Now define

$$\hat{g}(x,\theta) = \tfrac{1}{2}(\hat{g}_1(x,\theta) + \hat{g}_2(x,\theta))$$

where

$$\hat{g}_1(x,\theta) = b_n + n^{-1}b_n^{-1}\, \Sigma_{j=1}^{n}\, K(b_n^{-1}(x-X_j)),$$

$$\hat{g}_2(x,\theta) = b_n + n^{-1}\, b_n^{-1}\, \Sigma_{j=1}^{n}\, K(b_n^{-1}(x-Y_j+\theta)),$$

K and b_n are the same as in Example 1. Define

$$\hat{\ell}^*(x,y,\theta) = -\tfrac{1}{2}\{\dot{\hat{g}}/\hat{g}(y-\theta) - \dot{\hat{g}}/\hat{g}(x)\}.$$

with $\hat{\ell}_j^*$ and $\hat{g}_j$, defined in the same way as in Example 1. The conditions (C1)–(C5) can be verified. We omit the proofs since they are essentially the same as those in Example 1.

5. Proof of Theorem 3.1

Note that with $\ell^* = \ell^*(\cdot,\theta)$

$$n^{\frac{1}{2}}(\hat{\theta}_n - \theta(q_n)) = n^{\frac{1}{2}}(\bar{\theta}_n + n^{-1}\, \hat{I}_*^{-1}\, \Sigma_{j=1}^{n}\, \hat{\ell}^*(X_j,\bar{\theta}_n) - \theta - 2I_*^{-1}\int \ell^* f^{\frac{1}{2}}(q_n^{\frac{1}{2}} - f^{\frac{1}{2}})d\mu)$$

$$= n^{\frac{1}{2}}\{n^{-1}\, \hat{I}_*^{-1}\, \Sigma_{j=1}^{n}\, \hat{\ell}^*(X_j,\bar{\theta}_n) - \hat{I}_*^{-1}\int \hat{\ell}^*(x,\bar{\theta}_n)q_n(x)d\mu\}$$

$$+ n^{\frac{1}{2}}\, \hat{I}_*^{-1}\{\int \hat{\ell}^*(x,\bar{\theta}_n)q_n(x)d\mu - \int \hat{\ell}^*(x,\bar{\theta}_n)f(x)d\mu - 2\int \ell^*\, f^{\frac{1}{2}}(q_n^{\frac{1}{2}} - f^{\frac{1}{2}})d\mu\}$$

$$+ n^{\frac{1}{2}}\{\hat{I}_*^{-1} \int \hat{\ell}^*(x,\overline{\theta}_n)f(x)d\mu + (\overline{\theta}_n-\theta)\}$$

$$+ 2\, n^{\frac{1}{2}}(\hat{I}_*^{-1} - I_*^{-1}) \int \ell^* f^{\frac{1}{2}}(q_n^{\frac{1}{2}} - f^{\frac{1}{2}})d\mu$$

$$= A_n + B_n + C_n + D_n.$$

We will show D_n, C_n, $B_n \to 0$ in Q_n^n–probability and $A_n \Rightarrow N(0,I_*^{-1})$ under Q_n in Lemma 5.1–5.4 respectively. Throughout this section we assume the continuity of $I_*(\theta)$ and $\ell^*(x,\theta)$ for each $x \in \mathscr{X}$ as functions of θ.

Lemma 5.1 Under the conditions (C2), (C3.1), (C3.2) and (C5), $D_n \to 0$ in Q_n^n–probability.

Proof. Let $S_n(\theta) = \{x : |\ell^*(x,\theta)|^2 \leq a_n\}$ where $a_n \to \infty$, $n^{-\frac{1}{2}} a_n \to 0$ and $n^{-\frac{1}{2}} a_n L_n^{-2} \to \infty$. By the argument in Le Cam (1960,1969), it suffices to show that

(5.1) $$n^{-1} \Sigma_{j=1}^n \hat{\ell}^* \hat{\ell}^{*T}(X_j,\theta_n) \to I_*$$

in Q_n^n–probability where $\{\theta_n : n \geq 1\}$ is any deterministic sequence such that $n^{\frac{1}{2}}|\theta_n-\theta| = O(1)$. Using Chebyshev's inequality and $n^{-\frac{1}{2}} a_n \to 0$, we can show

(5.2) $$n^{-1} \Sigma_{j=1}^n \ell^*\ell^{*T}(X_j,\theta_n)I_{S_n(\theta_n)}(X_j) - \int_{S_n(\theta_n)} \ell^*\ell^{*T}(x,\theta_n)q_n(x)d\mu \to 0$$

in Q_n^n–probability. And also

(5.3) $$\int_{S_n(\theta_n)}\ell^*\ell^{*T}(x,\theta_n)q_n(x)d\mu \to I_*$$

by the Dominated Convergence Theorem, the continuity of $I_* = I_*(\theta)$ and the map $\theta \to \ell^*(x,\theta)$ for each x and the following

$$\left|\int_{S_n(\theta_n)} \ell^*\ell^{*T}(x,\theta_n)(q_n(x) - f(x,\theta_n,g))d\mu\right|$$

$$= \left|\int_{S_n(\theta_n)} \ell^*\ell^{*T}(x,\theta_n)(q_n^{\frac{1}{2}}(x) + f^{\frac{1}{2}}(x,\theta_n,g))(q_n^{\frac{1}{2}}(x) - f^{\frac{1}{2}}(x,\theta_n,g))d\mu\right|$$

$$\leq O(n^{-\frac{1}{2}})a_n \to 0$$

where $|A| = (\sum_{i,j} a_{ij}^2)^{\frac{1}{2}}$ if $A = (A_{ij})_{k\times k}$.

Furthermore,

(5.4) $$n^{-1} \Sigma_{j=1}^{n} \hat{\ell}^* \hat{\ell}^{*T}(X_j,\theta_n)\, I_{S_n^c(\theta_n)}(X_j) \to 0$$

in Q_n^n–probability since the expectation taken under Q_n of the absolute value of each component in the left hand side of (5.4) is bounded by

$$n^{\frac{1}{2}} L_n^2\, Q_n(|\ell^*(X,\theta_n)|^2 > a_n)$$

$$\leq n^{\frac{1}{2}} L_n^2(P_{\theta_n}(|\ell^*(X,\theta_n)|^2 > a_n) + O(n^{-\frac{1}{2}}))$$

$$\leq n^{\frac{1}{2}} L_n^2\, a_n^{-1}\, O(1) + L_n^2\, O(1) \to 0$$

where P_{θ_n} is the probability measure associated with the density $f(\cdot,\theta_n,g)$. Now using (C2), (C3.1), (C3.2) and (5.3), it is straightforward to arrive at

(5.5) $$n^{-1} \Sigma_{j=1}^{n} (\hat{\ell}^* \hat{\ell}^{*T}(X_j,\theta_n) I_{S_n(\theta_n)}(X_j) - \ell^* \ell^{*T}(X_j,\theta_n) I_{S_n(\theta_n)}(X_j)) \to 0$$

in Q_n^n–probability. Hence (5.1) follows from (5.2)–(5.5). ■

<u>Lemma 5.2</u>. Under the conditions (C1), (C2), (C3.1), (C3.2), (C4) and (C5), $C_n \to 0$ in Q_n^n–probability.

<u>Proof</u>. By the same argument as in the proof of Lemma 5.1, it suffices to show that

(5.6) $$n^{\frac{1}{2}}\{\int \hat{\ell}^*(x,\theta_n) f(x) d\mu + \hat{I}_*(\theta_n - \theta)\} \to 0$$

in Q_n^n–probability for any sequence such that $|\theta_n-\theta| = O(n^{-\frac{1}{2}})$. By (C1), (5.6) is equivalent to

$$(5.7) \qquad n^{\frac{1}{2}}\{\int \hat{\ell}^*(x,\theta_n)f(x)d\mu - \int \hat{\ell}^*(x,\theta_n)f(x,\theta_n,g)d\mu + \hat{I}_*(\theta_n-\theta)\} \to 0$$

in Q_n^n–probability. But the left hand side of (5.7) is equal to

$$(5.8) \qquad n^{\frac{1}{2}}(\hat{I}_*(\theta_n-\theta) - 2\int \hat{\ell}^*(x,\theta_n)f^{\frac{1}{2}}(x)(f^{\frac{1}{2}}(x,\theta_n,g) - f^{\frac{1}{2}}(x))d\mu) + o_{Q_n}(1)$$

by (C5) and the fact that $f(\cdot,\theta_n,g) \in N_n(f,c)$. Now by (C4) we can replace θ_n in $\hat{\ell}^*(\cdot,\theta_n)$ by θ and then by (C2) we can replace $\hat{\ell}^*(\cdot,\theta)$ by $\ell^*(\cdot,\theta)$ in the expression (5.8). Now (5.6) is obvious if we observe that

$$|\theta_n-\theta|^{-1}\|f^{\frac{1}{2}}(\cdot,\theta_n,g) - f^{\frac{1}{2}} - (\theta_n-\theta)^T\rho_\theta\| \to 0$$

and $2 < \ell^*(\cdot,\theta)f^{\frac{1}{2}}, \rho_\theta^T > = I_*$ and use Lemma 5.1. ∎

Lemma 5.3. Under the conditions (C2), (C3.1), (C3.2), (C4) and (C5), $B_n \to 0$ in Q_n^n–probability.

Proof. Again thanks to Le Cam (1960, 1969), it suffices to show that

$$(5.9) \qquad n^{\frac{1}{2}}(\int \hat{\ell}^*(x,\theta_n)q_n(x)d\mu - \int \hat{\ell}^*(x,\theta_n)f(x)d\mu - 2\int \ell^*(x,\theta)\, f^{\frac{1}{2}}(q_n^{\frac{1}{2}}-f^{\frac{1}{2}})d\mu) \to 0$$

in Q_n^n–probability. But by (C5) and the fact that $q_n \in N_n(f,c)$ we only need to show that

$$(5.10) \qquad n^{\frac{1}{2}}(\int \hat{\ell}^*(x,\theta_n)f^{\frac{1}{2}}(q_n^{\frac{1}{2}}-f^{\frac{1}{2}})d\mu - \int \ell^*(x,\theta)f^{\frac{1}{2}}(q_n^{\frac{1}{2}}-f^{\frac{1}{2}})d\mu) \to 0$$

in Q_n^n–probability. Now the euclidean norm of the left hand side of (5.10) is bounded by

$$\{\int |\hat{\ell}^*(x,\theta_n) - \ell^*(x,\theta)|^2 f(x)d\mu\}^{\frac{1}{2}}\, n^{\frac{1}{2}}\|q_n^{\frac{1}{2}}-f^{\frac{1}{2}}\|$$

which goes to zero in Q_n^n–probability by (C2) and (C4). ∎

Lemma 5.4. Under the conditions (C2)–(C5). $A_n \Rightarrow N(0,I_*^{-1})$ under Q_n.

Proof. Let $T_n = \{x : |\ell^*(x,\theta)|^2 \leq b_n\}$ where $b_n \to \infty$ and $n^{-\frac{1}{2}} b_n \to 0$. Note that

$$n^{\frac{1}{2}}(n^{-1} \Sigma_{j=1}^n \ell^*(X_j,\theta) I_{T_n}(X_j) - \int_{T_n} \ell^*(x,\theta) q_n(x) d\mu) \Rightarrow N(0,I_*)$$

under Q_n. Hence it suffices to show that

$$(5.11)\qquad n^{\frac{1}{2}}(n^{-1} \Sigma_{j=1}^n \hat{\ell}^*(X_j,\theta_n) - n^{-1} \Sigma_{j=1}^n \ell^*(X_j,\theta) I_{T_n}(X_j) - \int \hat{\ell}^*(x,\theta_n) q_n(x) d\mu + \int_{T_n} \ell^*(x,\theta) q_n(x) d\mu\} \to 0$$

in Q_n^n–probability. We will use Schick's approach to show (5.11). First of all, by (C3.3) and (C3.4), (5.11) is equivalent to

$$(5.12)\qquad n^{-\frac{1}{2}} \Sigma_{j=1}^n (\hat{\ell}_j^*(X_j,\theta_n) - \ell^*(X_j,\theta) I_{T_n}(X_j) - \int \hat{\ell}_j^*(x,\theta_n) q_n(x) d\mu + \int_{T_n} \ell^*(x,\theta) q_n(x) d\mu) \to 0$$

in Q_n^n–probability. Now the above term can be decomposed into two terms, namely, $n^{-\frac{1}{2}} \Sigma_{j=1}^n E_{nj}$ and $n^{-\frac{1}{2}} \Sigma_{j=1}^n F_{nj}$ where

$$E_{nj} = \hat{\ell}_j^*(X_j,\theta_n) I_{T_n}(X_j) - \ell^*(X_j,\theta) I_{T_n}(X_j) - \int_{T_n} \hat{\ell}_j^*(x,\theta_n) q_n(x) d\mu + \int_{T_n} \ell^*(x,\theta) q_n(x) d\mu$$

and

$$F_{nj} = \hat{\ell}_j^*(X_j,\theta_n) I_{T_n^c}(X_j) - \int_{T_n^c} \hat{\ell}_j^*(x,\theta_n) q_n(x) d\mu .$$

Instead of showing $n^{-\frac{1}{2}}\Sigma_{j=1}^{n} E_{nj} \to 0$, first we show $n^{-\frac{1}{2}}\Sigma_{j=1}^{n} G_{nj} \to 0$ where G_{nj} is defined as E_{nj} except that $\hat{\ell}_j^*(\cdot,\cdot)$ is replaced by $\bar{\ell}_j^*(\cdot,\cdot)$, the conditional expectation of $\hat{\ell}^*(\cdot,\cdot)$ given $X_1,\ldots,X_{j-1},X_{j+1},\ldots,X_n$. Then we will show that latter implies the former. Observe that

$$E_{Q_n}|n^{-\frac{1}{2}}\Sigma_{j=1}^{n} G_{nj}|^2 = n^{-1}\Sigma_{j=1}^{n} E_{Q_n}|G_{nj}|^2 + n^{-1}\Sigma_{j\neq k} E_{Q_n} G_{nj}^T G_{nk}\,.$$

Now

$$\begin{aligned} E_{Q_n}|G_{nj}|^2 \le\ & 2\,E_{Q_n}\int_{T_n}|\hat{\ell}^*(x,\theta_n) - \bar{\ell}_j^*(x,\theta_n)|^2\, q_n(x)d\mu \\ & + 4\,E_{Q_n}\int_{T_n}|\hat{\ell}^*(x,\theta_n) - \hat{\ell}^*(x,\theta)|^2\, q_n(x)d\mu \qquad (5.13)\\ & + 4\,E_{Q_n}\int_{T_n}|\hat{\ell}^*(x,\theta) - \ell^*(x,\theta)|^2\, q_n(x)d\mu\,. \end{aligned}$$

The first term of the right hand side of (5.13) is bounded by $2\,E_{Q_n}\int_{T_n}|\hat{\ell}^*(x,\theta_n) - \hat{\ell}_j^*(x,\theta_n)|^2\, q_n d\mu$ by the property of conditional variances and the third term is bounded by $4\,E_{Q_n}\int_{T_n}|\hat{\ell}^*(x,\theta) - \ell^*(x,\theta)|^2 f\, d\mu + O(n^{-\frac{1}{2}})E_{Q_n}[\sup_x|\hat{\ell}^*(x,\theta)|^2 + b_n]$. Hence by (C2), (C3.1), (C4) and (C5), $n^{-1}\Sigma_{j=1}^{n} E_{Q_n}|G_{nj}|^2 \to 0$ in Q_n^n–probability. Now using the argument in Schick (1987) we can show that

$$n^{-1}\Sigma_{j\neq k}|E_{Q_n} G_{nj}^T G_{nk}| \le \Sigma_{j=1}^{n} E_{Q_n}\int_{T_n}|\hat{\ell}^*(x,\theta_n) - \bar{\ell}_j^*(x,\theta_n)|^2 q_n\, d\mu\,. \qquad (5.14)$$

But the right hand side of (5.14) goes to zero by (C3.1) and the property of conditional variances. Hence we have shown that $n^{-\frac{1}{2}}\Sigma_{j=1}^{n} G_{nj} \to 0$ in Q_n^n–probability. Now it remains to show that the above implies $n^{-\frac{1}{2}}\Sigma_{j=1}^{n} E_{nj} \to 0$ in Q_n^n–probability. By (C3.1), Cauchy–Schwarz inequality and the fact that the right hand side of (5.14) goes to zero, we can see

$$n^{-\frac{1}{2}} \Sigma_{j=1}^{n} (\bar{\ell}_j^*(X_j,\theta_n) I_{T_n}(X_j) - \hat{\ell}_j^*(X_j,\theta_n) I_{T_n}(X_j)) \to 0$$

in Q_n^n–probability and

$$n^{-\frac{1}{2}} \Sigma_{j=1}^{n} (\int_{T_n} \bar{\ell}_j^*(x,\theta_n) q_n \, d\mu - \int_{T_n} \hat{\ell}_j^*(x,\theta_n) q_n \, d\mu) \to 0$$

in Q_n^n–probability, establishing $n^{-\frac{1}{2}} \Sigma_{j=1}^{n} E_{nj} \to 0$ in Q_n^n–probability. Similarly we can show $n^{-\frac{1}{2}} \Sigma_{j=1}^{n} F_{nj} \to 0$ in Q_n^n–probability. ■