Probab. Th. Rel. Fields 88, 107–120 (1991)

Probability Theory and Related Fields

Asymptotic minimax estimation in semiparametric models

Byeong Uk Park *
University of North Carolina at Chapel Hill, N.C., USA

Received December 19, 1989; in revised form March 1, 1990

Summary. We give several conditions on the estimator of efficient score function for estimating the parametric component of semiparametric models. A semiparametric version of the one-step MLE using an estimator of efficient score function which fulfills the conditions is shown to converge to the normal distribution with minimum variance locally uniformly over a fairly large neighborhood around the assumed semiparametric model. Consequently, it is shown to be asymptotically minimax with bounded subconvex loss functions. A few examples are also considered.

1. Introduction

As powerful criteria for identifying good statistical methods, asymptotic minimaxity and efficiency have been used widely in large sample theory. To avoid the confusion between these two notions in further discussion, let us clarify the distinction here. An estimator is called *asymptotically minimax* if its maximum risk taken over all probability measures in some neighborhood of the assumed model is asymptotically best possible. On the other hand, an estimator is called *asymptotically efficient* if the distribution of the estimator converges weakly to the normal distribution with minimum variance along either fixed or contiguously contaminated underlying probability measures.

In parametric problems the representation theorem of Hájek (1970), and the asymptotic minimax theorems of Hájek (1972) and Le Cam (1972), provide a rather complete description of asymptotic efficient and asymptotic minimax estimation. 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 in various particular semiparametric models. Those include Weiss and Wolfowitz (1970), Wol-

* *Permanent address*: Department of Computer Science and Statistics, Seoul National University, Seoul, Korea

fowitz (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 treated parametric models and the third dealt Hilbert space parameters.

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 ($n^{-1/2}$-rate) around $f^{1/2}(\cdot, \theta, g)$, namely, $\{q: n^{1/2}\|q^{1/2}-f^{1/2}\| \leqq c\}$ ($\|\cdot\|$ is the usual $L^2(\cdot)$ norm) which is essentially larger than the one considered in Begun et al. (1983). A special attention should be given here to the fact that our neighborhood is no longer contiguous. We show that a semiparametric version of the one-step MLE of Le Cam (1956, 1969), which uses an estimator of efficient score function satisfying a set of given conditions, converges to the normal distribution with minimum variance locally uniformly over this enlarged neighborhood. Consequently, the estimator is seen to be asymptotically minimax with bounded subconvex loss functions.

Our main result is in essence a semiparametric extension of the development in Beran (1981) for ordinary parametric models. Beran (1981) showed that adaptively modified one-step MLEs are asymptotically minimax. Some of the conditions imposed on the estimators of score function in this paper correspond to some of the properties possessed by the constructed score function in Beran (1981). However, Beran (1981) considered neighborhoods which are full Hellinger balls of probabilities, while we treat Hellinger balls of densities defined with respect to a fixed sigma-finite measure, which are subsets of those considered in Beran (1981).

Fabian and Hannan (1982) dealt the same problem in different mathematical formulation but they considered finite dimensional neighborhoods and treated only the cases in which adaptation is possible.

This paper is organized as follows. In Sect. 2 we present an appropriate asymptotic minimax theorem. In Sect. 3 we give several conditions on the estimator of efficient score function [see Begun et al. (1983) or Bickel et al. (1989) for definition] for asymptotic minimaxity of our estimator. A few examples are considered in Sect. 4 where we see how the conditions in Sect. 3 are satisfied and consequently establish asymptotically minimax estimators for those examples. The proof of our main theorem is given in Sect. 5.

2. Asymptotic minimax bound

Suppose that $X_1, \ldots, X_n$ are i.i.d. $\mathscr{X}$-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)]. Here we restate their conditions using the same notations.

Let

$$\mathscr{B}=\{\beta\in L^2(v): \|n^{1/2}(g_n^{1/2}-g^{1/2})-\beta\|_v\to 0 \quad \text{as } n\to\infty$$
$$\text{for some sequence } g_n \text{ with all } g_n\in\mathscr{G}\}.$$

Throughout this paper we assume that $\mathscr{B}$ is a subspace of $L^2(v)$. Furthermore, we assume that for every sequence (θ_n, g_n) such that $|n^{1/2}(\theta_n-\theta)-h|\to 0$ and $\|n^{1/2}(g_n^{1/2}-g^{1/2})-\beta\|_v\to 0$ as $n\to\infty$, there exist a function $\rho_\theta\in L^2(\mu)$ and a bounded linear operator A: $L^2(v)\to L^2(\mu)$ such that with $f_n\equiv f(\cdot, \theta_n, g_n)$

$$\|n^{1/2}(f_n^{1/2}-f^{1/2})-\alpha\|_\mu\to 0 \quad \text{as } n\to\infty \tag{2.1}$$

where $\alpha=h\rho_\theta+A\beta$.

Let $N_n(f, c)$ be the set of all density functions q with respect to μ such that $\|n^{1/2}(q^{1/2}-f^{1/2})\|_\mu\leqq c$. Let

$$\theta(q)=\theta+4I_*^{-1}\int\alpha^*(q^{1/2}-f^{1/2})\,d\mu \tag{2.2}$$

where $I_*=4\|\alpha^*\|_\mu^2$ and α^* is the orthogonal component of ρ_θ to the nuisance parameter space $\{A\beta: \beta\in\mathscr{B}\}$. Then by (2.1)

$$\theta(f_n)=\theta_n+o(n^{-1/2}) \tag{2.3}$$

where f_n is the same as in (2.1). The Eq. (2.3) 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 assumption described above,*

$$\lim_{c\to\infty}\liminf_{n}\inf_{\hat{\theta}_n}\sup_{q\in N_n(f,c)} E_Q\, l(n^{1/2}(\hat{\theta}_n-\theta(q)))\geqq E\,l(Z_*)$$

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

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

$$\lim_{\substack{c_1\to\infty\\ c_2\to\infty}}\liminf_{n}\inf_{\hat{\theta}_n}\sup_{\substack{|h|\leqq c_1\\ \|\beta\|_v\leqq c_2}} E_{P_n}\, l(n^{1/2}(\hat{\theta}_n-\theta_n))\geqq E\,l(Z_*)$$

where θ_n, h and β are the same as those in the condition described through (2.1) and P_n is the corresponding probability measure for $f_n=f(\cdot, \theta_n, g_n)$. From (2.3) we can replace $l(n^{1/2}(\hat{\theta}_n-\theta_n))$ by $l(n^{1/2}(\hat{\theta}_n-\theta(f_n)))$. Now observe that if $|h|\leqq c_1$ and $\|\beta\|_v\leqq c_2$, then $\|n^{1/2}(f_n^{1/2}-f^{1/2})\|_\mu<c$ for some $c>0$. Hence we can replace f_n with $|h|\leqq c_1$ and $\|\beta\|_v\leqq c_2$ 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

$$\lim_{c\to\infty} \lim_{n} \sup_{q\in N_n(f,c)} E_Q l(n^{1/2}(\hat{\theta}_n - \theta(q))) \geqq El(Z_*). \tag{2.4}$$

If we restrict l to be a bounded loss function, it suffices to find $\hat{\theta}_n$ such that

$$n^{1/2}(\hat{\theta}_n - \theta(q_n)) \Rightarrow N(0, I_*^{-1}) \tag{2.5}$$

under the probability measure Q_n having density q_n 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.

Remark. For asymptotic minimaxity in Begun et al.'s sense, we need to construct $\hat{\theta}_n$ such that

$$\lim_{\substack{c_1\to\infty \\ c_2\to\infty}} \lim_{n} \sup_{\substack{|h|\leqq c_1 \\ \|\beta\|\leqq c_2}} E_{P_n} l(n^{1/2}(\hat{\theta}_n - \theta_n)) = El(Z_*). \tag{2.6}$$

However, to show (2.6) we might still need to verify (2.5) since the maximizing density can only be said to belong to $N_n(f, c)$.

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 $n^{1/2}(\tilde{\theta}_n - \theta)$ is tight under any Q_n where $q_n \in N_n(f, c)$. Throughout this paper it is assume that such an estimator exists. M-estimation and minimum distance methods can be used to select an initial estimator, but considering Q_n^n is not contiguous to P^n, it should be chosen more carefully than in the usual cases of contiguous contaminating density. The following examples illustrate how to construct an initial estimator.

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

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

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.

Example 3 (linear regression with symmetric errors). Suppose that $(X_1, Y_1), \ldots, (X_n, Y_n)$ are i.i.d. $f(\cdot, \cdot, \theta)$ where

$$f(x, y, \theta) = g(y - \theta^T w(x))\, h(x),$$

$\theta \in R^p$, h is known with $\int |w(x)|^2 h(x)\, dx < \infty$, g is a unknown symmetric density function with $\int (\dot{g}^2/g)\, d\mu < \infty$. Let ψ be a bounded and antisymmetric function with a bounded positive first derivative. Let $\tilde{\theta}_n$ be the unique solution of

$$\sum_{j=1}^{n} w_i(X_j)\, \psi(Y_j - \tilde{\theta}_n^T w(X_j)) = 0 \qquad 1 \leqq i \leqq p,$$

where $w_i(\cdot)$ is the i^{th} component of $w(x)$. Then it is not so difficult to show that $\sqrt{n}(\tilde{\theta}_n - \theta)$ is tight under Q_n, following Huber's arguments [pp. 805–806, Huber (1973)].

3.2. *Asymptotic minimax estimator*

Let

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

where $\bar{\theta}_n$ is a discretized version of $\tilde{\theta}_n$, $\hat{l}^*(x, \theta)$ is a good estimator of $l^*(x, \theta) = 2\alpha^* f^{-1/2}(x, \theta, g)$ and $\hat{I}_* = n^{-1} \sum_{j=1}^{n} \hat{l}^* \hat{l}^{*T}(X_j, \bar{\theta}_n)$, and estimator of I_*. The asymptotic behavior of $\hat{\theta}_n$ depend heavily on that of $\hat{l}^*$. We state several conditions which $\hat{l}^*$ should satisfy in order that $\hat{\theta}_n$ defined in (3.2) satisfies (2.4), namely, be an asymptotically minimax estimator. Let $\{\theta_n : n \geqq 1\}$ be any sequence such that $|\theta_n - \theta| = O(n^{-1/2})$.

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

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

Let $\hat{l}_j^*$ be a cross-validated estimator of l^*, i.e., $\hat{l}_j^* = \hat{l}^*$ computed from $X_1, \ldots, X_{j-1}, X_{j+1}, \ldots, X_n$.
(C3) $|\hat{l}_j^*(x, \theta) - \hat{l}^*(x, \theta)| \leqq M_n = o(n^{-1/2}), j = 1, \ldots, n$ where M_n is a constant.

Although we can find $\hat{l}^*$ 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) $\sum_{j=1}^{n} E_{Q_n} \int |\hat{l}^*(x, \theta_n) - \hat{l}_j^*(x, \theta_n)|^2 q_n \, d\mu = o(1)$

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

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

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

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 $\hat{l}^*$.

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

(C5) $n^{-1/4} |\hat{l}^*(x, \theta)| \leqq L_n \to 0$

where N_n and L_n are constants.

Remark. The conditions (C1), (C2) and (C5) correspond to the properties (3.14), (3.15) and (3.16) of the constructed score function in Beran (1981).

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* $\theta \to l^*(x, \theta)$ *is continuous for each* $x \in \mathscr{X}$, *then*

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

under Q_n.

Proof. See Sect. 5.

As we discussed it earlier in Sect. 2, the conclusion of Theorem 3.1 implies that $\hat{\theta}_n$ with $\hat{l}^*$ satisfying (C1)–(C5) is asymptotically minimax if the loss function is bounded and subconvex.

As will become clearer in the proof of Theorem 3.1 [see (5.5) for example], the condition (C5) is essentially used for eliminating the effect due to model contamination, which can be done by adjusting the estimator according to the amount of contamination present. One unpleasant aspect of the condition (C5) is that it requires the estimator to be adjusted by a nonstochastic term (L_n) regardless of the actual amount of contamination. This means that the estimator breaks down for contaminated balls whose shrinking rates are slower than $n^{-1/2}$. Furthermore, in small sample cases, the estimator may fail to be adjusted appropriately since the nonstochastic term confuses sample size with the amount of contamination present.

Beran (1981), p. 99, also gives discussion of the above difficulty and suggests to use goodness-of-fit test statistics (instead of nonstochastic terms) to take the actual contamination into account for the estimator. But in our case, the use of the nonstochastic term appears to be rather crucial for the proof of Theorem 3.1 in technical aspects. At the present time, the author does not know whether this is due to the nature of the estimator defined in (3.2) or inherent in the problem. Further research needs to be done on this.

4. Examples (continued)

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

4.1. *Example 1* (continued)

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

(4.1)
$$\hat{g}(x,\theta)=b_n+n^{-1}b_n^{-1}\sum_{j=1}^{n}K(b_n^{-1}(x-X_j+\theta))$$

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

where $b_n\to 0$, $nb_n^6\to\infty$ and K is logistic density. Then

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

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 $|\hat{l}^*(x,\theta)|\leqq b_n^{-1}$ and $\left|\frac{\partial}{\partial\theta}\hat{l}^*(x,\theta)\right|\leqq Cb_n^{-3}$ 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

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

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 $Bn^{-1}b_n^{-6}$ for some constant $B>0$. Hence the conditions (C1)–(C5) are satisfied with $\hat{l}^*$ defined in (4.1).

4.2. *Example 2* (continued)

Observe that $l^*(x, y, \theta) = -\frac{1}{2}(\dot{g}/g(y-\theta) - \dot{g}/g(x)\}$ 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

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

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

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

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

With $\hat{l}_j^*$ 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.

4.3. *Example 3* (continued)

The efficient score function for θ is given by $l^*(x, y, \theta) = -w(x)\dot{g}/g(y - \theta^T w(x))$. Define

$$\hat{g}(y, \theta) = b_n + n^{-1} b_n^{-1} \sum_{j=1}^{n} K(b_n^{-1}(y - Y_j + \theta^T w(X_j)))$$

and

$$\hat{l}^*(x, y, \theta) = -\tfrac{1}{2} w(x)(\dot{\hat{g}}/\hat{g}(y - \theta^T w(x), \theta) - \dot{\hat{g}}/\hat{g}(\theta^T w(x) - y, \theta))\, I(|w(x)| \leqq c_n)$$

where $b_n \to 0$, $c_n \to \infty$, $n b_n^6 c_n^{-4} \to \infty$ and K is logistic density. Define $\hat{g}_j$ and $\hat{l}_j^*$ in the same fashion as in Examples 1 and 2. Now (C1) is obviously satisfied.

Noting that

$$|\hat{l}_j^*(x, y, \theta) - \hat{l}^*(x, y, \theta)| \leqq A n^{-1} b_n^{-3} c_n$$

$$|\hat{l}^*(x, y, \theta)| \leqq B b_n^{-1} c_n$$

$$\left|\frac{\partial}{\partial \theta} \hat{l}^*(x, y, \theta)\right| \leqq C b_n^{-2} c_n^2$$

for some A, B, $C > 0$, (C3)–(C5) can be seen to be satisfied. Now for (C2) it suffices to show that

$$\text{(4.3)}\quad E_{Q_n} \iint |w(x)|^2 [\dot{\hat{g}}/\hat{g}(y, \theta_n) I(|w(x)| \leqq c_n) - \dot{g}/g(y)]^2\, g(y)\, h(x)\, dy\, dx = o(1).$$

Since $\int |w(x)|^2 h(x)\,dx < \infty$ and $\int (\dot{g}^2/g)\,d\mu < \infty$, (4.3) is satisfied if

$$E_{Q_n} \int [\dot{\hat{g}}/\hat{g}(y, \theta_n) - \dot{g}/g(y)]^2 g(y)\,dy \to 0. \tag{4.4}$$

But (4.4) can be shown by the same way as in Example 1.

Remark. In Example 3, c_n is used for technical merit. In practice, for fixed sample size, one may take c_n large enough to include every data point. A much more important issue here is the choice of bandwidth b_n, since it is well-known that it has great influence on the performance of probability density function estimators and other related ones. For probability density function estimators, several data-driven bandwidth selection methods have been proposed. See Park and Marron (1990) or Hall et al. (1989), for example. However, for our examples discussed above those bandwidth selectors should not be used since the amounts of smoothing for different problems are different. Instead, data-based bootstrap methods as discussed in Park (1990) are recommended.

5. Proof of Theorem 3.1

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

$$\begin{aligned}
n^{1/2}(\hat{\theta}_n - \theta(q_n)) &= n^{1/2}\Big\{\bar{\theta}_n + n^{-1}\hat{I}_n^{-1}\sum_{j=1}^{n}\hat{l}^*(X_j, \bar{\theta}_n) - \theta \\
&\qquad - 2I_*^{-1}\int l^* f^{1/2}(q_n^{1/2} - f^{1/2})\,d\mu\Big\} \\
&= n^{1/2}\Big\{n^{-1}\hat{I}_*^{-1}\sum_{j=1}^{n}\hat{l}^*(X_j, \bar{\theta}_n) - \hat{I}_*^{-1}\int \hat{l}^*(x, \bar{\theta}_n)\,q_n(x)\,d\mu\Big\} \\
&\quad + n^{1/2}\hat{I}_*^{-1}\{\int \hat{l}^*(x, \bar{\theta}_n)\,q_n(x)\,d\mu - \int \hat{l}^*(x, \bar{\theta}_n) f(x)\,d\mu \\
&\quad - 2\int l^* f^{1/2}(q_n^{1/2} - f^{1/2})\,d\mu\} \\
&\quad + n^{1/2}\{\hat{I}_*^{-1}\int \hat{l}^*(x, \bar{\theta}_n) f(x)\,d\mu + (\bar{\theta}_n - \theta)\} \\
&\quad + 2n^{1/2}(\hat{I}_*^{-1} - I_*^{-1})\int l^* f^{1/2}(q_n^{1/2} - f^{1/2})\,d\mu \\
&= A_n + B_n + C_n + D_n.
\end{aligned}$$

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 Lemmas 5.1–5.4, respectively.

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

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

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

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

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

in Q_n^n-probability. And also

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

by the dominated convergence theorem, the continuity of $I_* = I_*(\theta)$ and the map $\theta \to l^*(x, \theta)$ for each x and the following

$$\begin{aligned} & \left| \int_{S_n(\theta_n)} l^* l^{*T}(x, \theta_n)(q_n(x) - f(x, \theta_n, g)) \, d\mu \right| \\ & = \left| \int_{S_n(\theta_n)} l^* l^{*T}(x, \theta_n)(q_n^{1/2}(x) + f^{1/2}(x, \theta_n, g))(q_n^{1/2}(x) - f^{1/2}(x, \theta_n, g)) \, d\mu \right| \\ & \leqq O(n^{-1/2}) \, a_n \to 0 \end{aligned}$$

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

Furthermore,

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

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

$$\begin{aligned} (5.5) \quad n^{1/2} L_n^2 Q_n(|l^*(X, \theta_n)|^2 > a_n) & \leqq n^{1/2} L_n^2 (P_{\theta_n}(|l^*(X, \theta_n)|^2 > a_n) + O(n^{-1/2})) \\ & \leqq n^{1/2} L_n^2 a_n^{-1} O(1) + L_n^2 O(1) \to 0 \end{aligned}$$

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

$$(5.6) \qquad n^{-1} \sum_{j=1}^{n} (\hat{l}^* \hat{l}^{*T}(X_j, \theta_n) I_{S_n(\theta_n)}(X_j) - l^* l^{*T}(X_j, \theta_n) I_{S_n(\theta_n)}(X_j)) \to 0, \qquad (5.6)$$

in Q_n^n-probability. Hence (5.1) follows from (5.2)–(5.4) and (5.6).

Lemma 5.2. *Under the conditions* (C1), (C2), (C3.1), (C3.2), (C4) *and* (C5), $C_n \to 0$ *in Q_n^n-probability.*

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

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

in Q_n^n-probability for any sequence such that $|\theta_n-\theta|=O(n^{-1/2})$. By (C1), (5.7) is equivalent to

$$(5.8)\qquad n^{1/2}\{\int \hat{l}^*(x,\theta_n)\,f(x)\,d\mu-\int \hat{l}^*(x,\theta_n)\,f(x,\theta_n,g)\,d\mu+\hat{I}_*(\theta_n-\theta)\}\to 0$$

in Q_n^n-probability. But the left hand side of (5.8) is equal to

$$(5.9)\quad n^{1/2}(\hat{I}_*(\theta_n-\theta)-2\int \hat{l}^*(x,\theta_n)\,f^{1/2}(x)(f^{1/2}(x,\theta_n,g)-f^{1/2}(x))\,d\mu)+o_{Q_n}(1)$$

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 $\hat{l}^*(\cdot,\theta_n)$ by θ and then by (C2) we can replace $\hat{l}^*(\cdot,\theta)$ by $l^*(\cdot,\theta)$ in the expression (5.9). Now (5.7) is obvious if we observe that

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

and $2\langle l^*(\cdot,\theta)\,f^{1/2},\rho_\theta^T\rangle=I_*$ 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

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

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

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

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

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

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 $T_n=\{x\colon |l^*(x,\theta)|^2\leqq b_n\}$ where $b_n\to\infty$ and $n^{-1/2}b_n\to 0$. Note that

$$n^{1/2}\left(n^{-1}\sum_{j=1}^{n} l^*(X_j,\theta)\,I_{T_n}(X_j)-\int_{T_n} l^*(x,\theta)\,q_n(x)\,d\mu\right)\Rightarrow N(0,I_*)$$

under Q_n. Hence it suffices to show that

$$(5.11)\qquad n^{1/2}\left\{n^{-1}\sum_{j=1}^{n}\hat{l}^*(X_j,\theta_n)-n^{-1}\sum_{j=1}^{n} l^*(X_j,\theta)\,I_{T_n}(X_j)\right.$$
$$\left.-\int \hat{l}^*(x,\theta_n)\,q_n(x)\,d\mu+\int_{T_n} l^*(x,\theta)\,q_n(x)\,d\mu\right\}\to 0$$

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

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

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

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

and

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

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

$$E_{Q_n} \left| n^{-1/2} \sum_{j=1}^{n} G_{nj} \right|^2 = n^{-1} \sum_{j=1}^{n} E_{Q_n} |G_{nj}|^2 + n^{-1} \sum_{j \neq k} E_{Q_n} G_{nj}^T G_{nk}.$$

Now

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

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

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

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 $n^{-1/2}\sum_{j=1}^{n} G_{nj}\to 0$ in Q_n^n-probability. Now it remains to show that the above implies $n^{-1/2}\sum_{j=1}^{n} E_{nj}\to 0$ 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

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

in Q_n^n-probability and

$$n^{-1/2}\sum_{j=1}^{n}\Big(\int_{T_n}\bar{l}_j^*(x,\theta_n)q_n\,d\mu-\int_{T_n}\hat{l}_j^*(x,\theta_n)q_n\,d\mu\Big)\to 0$$

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

Acknowledgements. The author wants to express his sincere gratitude to the referees for many valuable and helpful comments.

References

Begun, J.M., Hall, W.J., Huang, W., Wellner, J.A.: Information and asymptotic efficiency in parametric-nonparametric models. Ann. Stat. **11**, 432–452 (1983)

Beran, R.: Asymptotically efficient adaptive rank estimates in location model. Ann. Statist. **2**, 63–74 (1974)

Beran, R.: Estimating a distribution function. Ann. Stat. **5**, 400–404 (1977)

Beran, R.: An efficient and robust adaptive estimator of location. Ann. Stat. **6**, 292–313 (1978)

Beran, R.: Efficient robust estimates in parametric models. Z. Wahrscheinlichkeitstheor. Verw. Geb. **55**, 91–108 (1981)

Bickel, P.J.: On adaptive estimation. Ann. Stat. **10**, 647–671 (1982)

Bickel, P.J., Klassen, C.J., Ritov, Y., Wellner, J.A.: Efficient and adaptive inference in semiparametric models. Baltimore: Johns Hopkins University Press 1989

Bickel, P.J., Ritov, Y.: Efficient estimation in the errors in variables model. Ann. Stat. **15**, 513–540 (1987)

Efron, B.: The efficiency of Cox's likelihood function for censored data. J. Am. Stat. Assoc. **72**, 557–565 (1977)

Fabian, V., Hannan, J.: On estimation and adaptive estimation for locally asymptotically normal families. Z. Wahrscheinlichkeitstheor. Verw. Geb. **59**, 459–578 (1982)

Hájek, J.: A characterization of limiting distributions of regular estimates. Z. Wahrscheinlichkeitstheor. Verw. Geb. **14**, 323–330 (1970)

Hájek, J.: Local asymptotic minimax and admissibility in estimation. Proc. Sixth Berkeley Symp. Math. Stat. Probab. **1**, 175–194 (1972)

Hall, P., Marron, J.S., Park, B.U.: Smoothed cross-validation. Manuscript (1989)

Huber, P.: Robust regression: asymptotics, conjectures and Monte Carlo. Ann. Stat. **1**, 799–821 (1973)

Klaassen, C.J.: Consistent estimation of the influence function of locally asymptotically linear estimators. Ann. Stat. **15**, 1548–1562 (1987)

Koshevnik, Y.A., Levit, B.Y.: On a nonparametric analogue of the information matrix. Theory Probab. Appl. **21**, 738–753 (1976)

Le Cam, L.: On the asymptotic theory of estimation and testing hypotheses. Proc. Third Berkeley Symp. Math. Stat. Probab. **1**, 129–156 (1956)

Le Cam, L.: Locally asymptotically normal families of distributions. Univ. of California Publ. Statist. **3**, 37–98 (1960)

Le Cam, L.: Théorie asymptotique de la décision statistique. Les Presses de l'Universite de Montreal 1972

Le Cam, L.: Limits of experiments. Proc. Sixth Berkeley Symp. Math. Stat. Probab. **1**, 245–261 (1972)

Millar, P.W.: Asymptotic minimax theorems for the sample distribution function. Z. Wahrscheinlichkeitstheor. Verw. Geb. **48**, 233–252 (1979)

Millar, P.W.: A general approach to the optimality of minimum distance estimators. Trans. Am. Math. Soc. **286**, 377–418 (1984)

Park, B.U.: Efficient estimation in the two-sample semiparametric location-scale model and the orientation shift model. Ph.D. Thesis, University of California, Berkeley 1987

Park, B.U.: Efficient estimation in the two-sample semiparametric location-scale model. Probab. Th. Rel. Fields **86**, 21–39 (1990)

Park, B.U., Marron, J.S.: Comparison of data-driven bandwidth selectors. J. Am. Stat. Assoc. **85**, 66–72 (1990)

Rao, P., Schuster, E., Littell, R.: Estimation of shift and center of symmetry based on Kolmogorov-Smirnov statistics. Ann. Stat. **3**, 862–873 (1975)

Schick, A.: On asymptotically efficient estimation in semiparametric models. Ann. Stat. **14**, 1139–1151 (1986)

Schick, A.: A note on the construction of asymptotically linear estimators. J. Stat. Plann. Inference **16**, 89–105 (1987)

Schuster, E., Narvarte, J.: A new nonparametric estimator of the center of a symmetric distribution. Ann. Stat. **1**, 1096–1104 (1973)

Stone, C.J.: Adaptive maximum likelihood estimators of a location parameter. Ann. Stat. **3**, 267–284 (1975)

Tsiatis, A.: A large sample study of Cox's regression model. Ann. Stat. **9**, 93–108 (1981)

van Eden, C.: Efficieny-robust estimation of location. Ann. Math. Stat. **41**, 172–181 (1970)

Weiss, L., Wolfowitz, J.: Asymptotically nonparametric estimators of location and scale parameters. Z. Wahrscheinlichkeitstheor. Verw. Geb. **16**, 134–150 (1970)

Wolfowitz, J.: Asymptotically efficient nonparametric estimators of location and scale parameters, II. Z. Wahrscheinlichkeitstheor. Verw. Geb. **30**, 117–128 (1974)