SEMIPARAMETRIC ESTIMATION FOR
STATIONARY MARKOV CHAINS WITH
CONTINUOUS STATE SPACE

SWAMINATHAN V.
Department of Statistics, North Maharashtra University,
Jalgaon - 425001, INDIA

1 Introduction

Minimum Hellinger distance estimation method has been widely used especially when one is concerned with achieving maximum efficiency as well as robustness of the estimator. In his seminal paper in 1977, Beran developed the theory of minimum Hellinger distance method for the i.i.d. setup. Also, Beran (1977) and Stather (1981) proved that these estimators are both asymptotically efficient and robust to small perturbations. Further, Tamura and Boos (1986) considered minimum Hellinger distance estimation for multivariate location and covariance and they proved that these estimators are affine invariant, consistent and asymptotically normal. In fact, their Monte Carlo results suggest that the MHDE's are an attractive robust alternative to the usual sample mean and covariance matrix. Simpson (1987) used the above method for the analysis of count data and he extended the result that MHDE is asymptotically equivalent to the maximum likelihood estimator when the model assumed is correct, to models with countable support. In a subsequent paper (Simpson (1989)) he proved that, if the model is correct, the Hellinger distance analogues of likelihood ratio tests are asymptotically equivalent to the latter.

All the above works have essentially emphasized the efficiency and robustness of the minimum Hellinger distance approach to estimation and testing problems. In what follows we modify the notion of minimum Hellinger distance and apply this modified minimum Hellinger distance approach to stationary Markov chains with continuous state space.

The required setup is explained in detail in Section 2. Also, this section describes the proposed estimation method for estimating the parameters of interest. The asymptotic properties of the estimators are also established.
2 Modified Minimum Hellinger Distance Method

In this section, we modify and extend the minimum Hellinger distance method of Beran (1977) from the case of independent and identically distributed random variables to stationary Markov chains and prove the asymptotic properties of the modified minimum Hellinger distance estimator (MMHDE). Now we present the preliminaries we require.

**Definition 2.1**: A Markov chain \( \{X_n, n = 0, \pm 1, \ldots \} \) is said to be geometrically ergodic if
\[
\int \| P^n(x, \cdot) - \pi(\cdot) \| \, dx = O(\gamma^n), \ n \geq 1
\]
for some \( \gamma \) such that \( 0 < \gamma < 1 \), where \( P^n(x, \cdot) \) is the \( n \)-step transition probability kernel starting from \( x \) and \( \pi \) is the invariant measure associated with \( \{X_n\} \).

**Definition 2.2**: Let \( \{X_n, n = 0, \pm 1, \ldots \} \) be a strictly stationary process. Let
\[
\Delta_n = \sup_{f \in S_n} [E(f \mid X_k, k < 1) - E(f)]
\]
where \( S_n \) is the set of random variables measurable with respect to the \( \sigma \)-field generated by \( X_k, k \geq n \) and bounded by unity in absolute value. Then \( \{X_n\} \) is said to be absolutely regular if \( E\Delta_n \to 0 \) as \( n \to \infty \) (cf. Pham (1986)).

**Lemma 2.1**: If \( \{X_n, n = 0, \pm 1, \ldots \} \) is a Markov chain with the invariant measure \( \pi \), then geometric ergodicity is equivalent to absolute regularity for this chain and absolute regularity of \( \{X_n\} \) implies that the sequence is also strong mixing with geometric mixing coefficients.

**Proof**: Refer Section 1, p. 291 of Pham (1986) and Section 1, p. 298 of Pham and Tran (1985).

If \( \{X_n\} \) is a Markov chain with the invariant measure \( \pi \), it is known that geometric ergodicity of this chain implies that it is absolutely regular; also if \( \{\alpha(n)\} \) is the sequence of mixing coefficients of a strongly mixing sequence \( \{X_n\} \), then \( \alpha(n) \leq 4E | \Delta_n | \) where \( \Delta_n \) are as defined earlier (cf. Pham and Tran (1985)). Hence absolute regularity implies strong mixing of \( \{X_n\} \).

Henceforth we shall assume that the Markov chain \( \{X_n, n \geq 1\} \) is strictly stationary with the invariant measure \( \pi \) and that this chain is geometrically ergodic with geometric mixing coefficients \( \alpha_n \).

Let the true transition density of \( \{X_n\} \) be \( f \) and the true stationary density be \( \pi \). Further, let us assume that \( f \) belongs to \( \{f_\theta, \ \theta \in \Omega\} \), a specified family of transition densities, where \( \theta \) is a real parameter. Then, as mentioned in Beran (1977), lack of information or data contamination or other factors beyond our control make it possible that the above family is not strictly correct. However the family may be close, in some sense, to the true transition density \( f \). On the other hand, \( f \) may be one of the members of the above family. So in either case, an estimator of \( \theta \) can be looked upon as the value...
\( \theta_0 \in \Omega \) such that the distance between \( f_{\theta_0} \) and \( f \) is the minimum among the distances between \( f_{\theta} \) and \( f \) for \( \theta \in \Omega \).

Now, unlike the usual Hellinger metric where the distance between two density functions is the value of the integral with respect to the Lebesgue measure, here we face a different situation: we assume a model \( \{ f_\theta, \theta \in \Omega \} \) for the Markov chain \( \{ X_n \} \), and for which the stationary density may not be in a tractable form, as illustrated in Swaminathan and Nimbalkar (1997) p.316 (this example deals with the intractability of the form of the stationary distribution of a random coefficient autoregressive model). Hence we have to work with the transition densities \( f_\theta \), and the definition of Hellinger metric has to be suitably modified to define the distance between two transition densities. The modified definition is given below.

Let \( \mathcal{F}_1 \) be the collection of all one-step transition densities on \( \mathcal{R} \) and \( \mathcal{F}_2 \) be the collection of all densities on \( \mathcal{R} \). Let \( \mathcal{F} = \mathcal{F}_1 \times \mathcal{F}_2 \).

**Definition 2.3**: For any \( f_1 \) and \( f_2 \) in \( \mathcal{F}_1 \) and \( \pi \) in \( \mathcal{F}_2 \), we define the modified Hellinger distance between \( f_1 \) and \( f_2 \) with respect to the reference density \( \pi \) as

\[
\| f_1^{1/2} - f_2^{1/2} \|_\pi = \int \| f_1^{1/2}(\cdot | y) - f_2^{1/2}(\cdot | y) \|^2 \pi(y) dy,
\]

where the norm in the right side of the expression is the \( L_2(\mathcal{R}) \)-norm with respect to the Lebesgue measure.

**Remark**: Note, in the above definition, the natural measure with respect to which the second integral (involving the argument \( y \)) is to be computed is the true stationary measure \( \pi \) as any other measure might either increase or decrease the true contribution of a state \( y \). Also, the integral may not be finite if we use a sigma-finite measure.

We shall drop the subscript \( \pi \) on the left henceforth when no confusion would arise as the reference density in all that follows will be the true stationary density \( \pi \).

Let \( T \) be a functional on \( \mathcal{F} \) taking values in \( \Omega \).

**Definition 2.4**: If \( T \) satisfies

\[
\| f_{T(f, \pi)}^{1/2} - f^{1/2} \| = \inf_{\theta \in \Omega} \int \| f_\theta^{1/2}(\cdot | y) - f^{1/2}(\cdot | y) \|^2 \pi(y) dy,
\]

then \( T \) is the modified minimum Hellinger distance estimator (MMHDE) of \( \theta \) when \( f \) and \( \pi \) are the true transition and stationary densities.

The setup for this section is as follows:

\( \Omega \) is compact in \( \mathcal{R}^k \).

**Theorem 2.1**: Let \( \Omega \) be a compact subset of \( \mathcal{R}^k \). Suppose that \( f_{\theta_1}(x | y) \neq f_{\theta_2}(x | y) \) for \( \theta_1 \neq \theta_2 \) on a set in \( \mathcal{R}^2 \) of non-vanishing Lebesgue measure. Also assume that \( f_\theta(\cdot | \cdot) \) is continuous in \( \theta \) almost everywhere in \( \mathcal{R}^2 \). Then

(a) for every \( (f, \pi) \in \mathcal{F} \), there exists \( T(f, \pi) \in \Omega \) satisfying (1):
(b) if $T(f, \pi)$ is unique, then $T$ is continuous at $(f, \pi)$ in the modified Hellinger topology; that is if $f_n^{1/2}(\cdot | y) \to f^{1/2}(\cdot | y)$ in $L_2$ for all $y$ and $\pi_n \to \pi$ in $L_1$, then $T(f_n, \pi_n) \to T(f, \pi)$.

**Proof:** (a) Define $m(\theta) = \int \| f^{1/2}_{\theta_1}(\cdot | y) - f^{1/2}_{\theta_2}(\cdot | y) \|^2 \pi(y) dy$. Let $\{\theta_n\} \subset \Omega$ be such that $\theta_n \to \theta$. Then

$$| m(\theta_n) - m(\theta) | \leq 4 \int \| f^{1/2}_{\theta_n}(\cdot | y) - f^{1/2}_{\theta}(\cdot | y) \|^2 \pi(y) dy$$

$$= 8 (1 - \int \int f^{1/2}_{\theta_n}(x | y) f^{1/2}_{\theta}(x | y) \pi(y) dx dy).$$

Now using the fact that $f_{\theta}(x | y)$ is continuous in $\theta$ and Fatou's lemma, we see that the right side of the above inequality goes to zero as $n \to \infty$. Hence $m$ is continuous on $\Omega$ and achieves a minimum in $\Omega$.

(b) Let $\{\{f_n, \pi_n\}\}$ be a sequence in $F$ such that $f_n^{1/2}(\cdot | y) \to f^{1/2}(\cdot | y)$ in $L_2$ for all $y$ and $\pi_n \to \pi$ in $L_1$. Let

$$m_n(\theta) = \int \| f^{1/2}_{\theta}(\cdot | y) - f^{1/2}_{\theta_n}(\cdot | y) \|^2 \pi(y) dy,$$

$$T(f_n, \pi_n) = \theta_n \text{ and } T(f, \pi) = \theta_0.$$

Now, for $\theta \in \Omega$,

$$| m_n(\theta) - m(\theta) |$$

$$\leq \int \| f^{1/2}_{\theta}(\cdot | y) - f^{1/2}_{\theta_n}(\cdot | y) \|^2 - \| f^{1/2}_{\theta_n}(\cdot | y) - f^{1/2}_{\theta}(\cdot | y) \|^2 \pi(y) dy$$

$$+ \int \| f^{1/2}_{\theta_n}(\cdot | y) - f^{1/2}_{\theta_n}(\cdot | y) \|^2 \pi_n(y) - \pi(y) dy$$

$$\leq \frac{1}{4} \left( \int \| f^{1/2}_{\theta_n}(\cdot | y) - f^{1/2}_{\theta_n}(\cdot | y) \|^2 \pi(y) dy + \int | \pi_n(y) - \pi(y) | dy \right).$$

Using Fatou's lemma and Scheffe's theorem we see that the right side goes to zero uniformly in $\theta$ as $n \to \infty$. That is,

$$\sup_{\theta \in \Omega} | m_n(\theta) - m(\theta) | \to 0, \text{ as } n \to \infty.$$  

Rest of the proof follows from Theorem 1, p. 447 of Beran (1977).

Let $X_1, X_2, \ldots, X_{n+1}$ be the observations from the Markov chain $\{X_n\}$. Since the true transition and stationary densities are unknown, we estimate them using kernels. Under appropriate conditions on the densities to be estimated and the kernels to be used, we show the consistency of the MMHDE of the transition density $f$. In order to achieve this, we introduce an estimator of the conditional cdf $F(x | y)$ associated with the transition density $f(x | y)$. The details are given below.

Let $K : \mathcal{R} \to \mathcal{R}^+ \cup \{0\}$ be a function such that
C1: $\int K(z)dz = 1$.

C2: $K$ is bounded with compact support, say the closed interval $[-a, a]$, and $K$ is symmetric about zero.

C3: $\{h_n, n \geq 1\}$ is a sequence of positive reals such that $h_n \downarrow 0$ as $n \to \infty$.

A real-valued function $g(z)$ on $\mathbb{R}^p$ is said to be of class $\mathcal{D}_p(y; \lambda)$, where $y \in \mathbb{R}^p$ and $\lambda$ is a positive constant, if and only if for some $\delta > 0$ there exists $N < \infty$ such that for $\|z\| < \delta$ (here $\|\cdot\|$ is the euclidean norm),

$$|g(y-z) - P_0 - P_1 - \ldots - P_r| \leq N \|z\|^\lambda,$$

where $P_0 = g(y)$ and for $j \geq 1$, $P_j$ is a homogeneous polynomial of degree $j$ in $y, y_1, \ldots, y_p$ (the components of $y$) and $r$ is the largest integer less than $\lambda$. Note that $N$ might depend on $y$. This condition essentially assures the smoothness of the function $g$ at $y$.

We say $g(z) \in \mathcal{D}_p(y; \lambda)$ if for some $\delta > 0$, $\sup_{\|z\| < \delta} |g(z)| < \infty$ (cf. Robinson (1983)).

C4: $\pi(z) \in \mathcal{D}_1(y; \lambda)$.

It can be seen that if the kernel chosen itself is a probability density then $r$ is at most 1 and $\lambda \leq 2$ (cf. Robinson (1983), p.191).

C5: $nh_n^{1+2\lambda} \to 0$ as $n \to \infty$ for $1 < \lambda \leq 2$ and $nh_n^{1+2\lambda} \to \infty$ as $n \to \infty$ for $0 \leq \lambda \leq 1$.

C6: For all $m \geq 1$ the pdf of $(X_n, X_{k+m})$ exists and is of class $\mathcal{D}_2(y, x)$ and the bound on the pdf in some neighborhood of $(y, x)$ is uniform in $m$.

The estimator of $F(x | y)$ is given by

$$\hat{F}_n(x | y) = \frac{\sum_{i=1}^n I(X_{i+1} \leq x)K(\frac{x-X_{i+1}}{h_n})}{\sum_{i=1}^n K(\frac{x-X_{i+1}}{h_n})}.$$ 

Further, assume that

C7: $\pi(y) > 0$;

C8:

(a) $F(x | z)$, as a function in $z$, belongs to $\mathcal{D}_1(y; \lambda)$ for all $x$;

(b) for fixed $y$, the corresponding set of constants $\{N(x)\}$ is bounded;

C9: $F(x | z)$, as a function of $z$, is continuous at $z = y$ for all $x$.

**Theorem 2.2**: Let $\{X_n, n \geq 1\}$ be a stationary and geometrically ergodic Markov chain. Let the conditions C1 - C9 hold and $y$ be fixed. Further, assume that $F(. | y)$ is uniformly continuous. Let

$$S_n(x | y) = (nh_n)^{1/2}[\hat{F}_n(x | y) - F(x | y)]$$
and let \( S_n \) be the stochastic process \( \{S_n(x | y), x \in \mathcal{R}\} \). Then the above sequence of processes \( \{S_n\} \) converges weakly to a Gaussian process \( \{S(x | y), x \in \mathcal{R}\} \) where

\[
E(S(x | y)) = 0 \quad \forall x
\]

and

\[
\text{Cov}(S(x_1 | y), S(x_2 | y)) = \frac{F(x_1 \land x_2 | y) - \tilde{F}(x_1 | y) \tilde{F}(x_2 | y)}{\pi(y)} \int_{-\infty}^{\infty} K^2(z) \, dz.
\]

**Proof:** The result that, for every \( x \), \( S_n(x | y) \xrightarrow{D} S(x | y) \) follows from Theorem 5.1 of Robinson (1983) by choosing the function \( g \) given in that theorem as

\[
g_x(X) = I(X \leq x),
\]

and using \( C1 \cdot C8(a) \) and \( C9 \).

Similarly, using Cramer-Wold device, the finite-dimensional convergence in distribution, with the limiting covariance structure as given in the above statement of theorem, too follows immediately.

Now it remains to show that the sequence of probability measures associated with \( \{S_n\} \) is tight. In view of the fact that \( S_n \) is a random function in the function space \( D = D(-\infty, \infty) \), we shall use an extension of Theorem 3' of Lindvall (1973).

Let \( (\Omega, \mathcal{F}, P) \) be a probability space and let \( X, \{X_n\} \) be random functions in \( D \) defined on \( \Omega \). Define

\[
T_X = \{ t : P(X(t) = X(t-)) = 1 \}.
\]

For \( \beta > 0 \), let \( r_\beta : D \rightarrow D[-\beta, \beta] \) be defined by \( (r_\beta x)(t) = x(t), -\beta \leq t \leq \beta \) where \( x \in D \). Then the extension of Theorem 3' of Lindvall (1973) reads as follows:

\( X_n \) converges weakly to \( X \) if and only if \( r_\beta X_n \) converges weakly to \( r_\beta X \) for every \( \beta \in T_X \).

So, for our purposes, it would suffice to consider the weak convergence over the compact sets \( [-\beta, \beta] \) where \( \beta \in T_S \).

Now, by equation (9.3) of Robinson (1983), we have, for \( x < \gamma \).

\[
S_n(z | y) - S_n(x | y) = \sqrt{nh_n} \{ U_n^{-1}(y)((nh_n)^{-1}\sum_{i=1}^{n} I(x < X_{i+1} \leq z)K(\frac{y-X_i}{nh_n})
- (F(z | y) - F(x | y))\pi(y))\}
- \sqrt{nh_n} \{ V_n(z, x, y)U_n^{-2}(y)((nh_n)^{-1}\sum_{i=1}^{n} K((\frac{y-X_i}{nh_n}) - \pi(y))) \}
\]

(2)
where \( U_n(y) \) and \( V_n(z, x, y) \) are random variables such that

\[
| U_n(y) - \pi(y) | \leq (nh_n)^{-1} \sum_{i=1}^{n} K(\frac{y - X_i}{h_n}) - \pi(y) \tag{3}
\]

\[
| V_n(z, x, y) - (F(z | y) - F(x | y))\pi(y) | \leq (nh_n)^{-1} \sum_{i=1}^{n} I(x < X_i+1 \leq z)K(\frac{y - X_i}{h_n}) - (F(z | y) - F(x | y))\pi(y) \tag{4}
\]

and that \( U_n(y) \xrightarrow{p} \pi(y) \) and \( V_n(z, x, y) \xrightarrow{p} (F(z | y) - F(x | y))\pi(y) \).

Now, the tightness of the sequence of processes \( \{ (S_n(x | y))_{x \in \mathbb{R}} \} \) would follow if the terms on the right hand side of (2) are shown to be tight. We further note that, to prove the tightness of the second term on the RHS of (2), in view of (4), it is enough to prove the tightness of the first term on the RHS of that equation. We justify it as follows:

Let

\[
R_n(y) = \sqrt{nh_n((nh_n)^{-1} \sum_{i=1}^{n} K(\frac{y - X_i}{h_n}) - \pi(y))}. \tag{5}
\]

Firstly, for a given \( \epsilon > 0 \), we should find a \( \delta > 0 \) such that the left hand side of the following inequality is small:

\[
P[ \sup_{|x-z| < \delta} | V_n(z, x, y)U_n^{-2}(y)R_n(y) | > \epsilon ] \leq P[U_n^{-2}(y) | R_n(y) | \sup_{|x-z| < \delta} | V_n(z, x, y) - (F(z | y) - F(x | y))\pi(y) | > \frac{\epsilon}{2}] + P[U_n^{-2}(y) | R_n(y) | \pi(y) \sup_{|x-z| < \delta} | F(x | y) - F(z | y) | > \frac{\epsilon}{2}].
\]

It is known that \( R_n(y) \xrightarrow{D} R(y) \) where \( R(y) \) has a normal distribution with mean zero and variance \( \pi(y) \int_{-\infty}^{y} K^2(z)dz \) (cf. Robinson (1983), Theorem 4.1). Hence the sequences \( \{ U_n(y) \} \) and \( \{ R_n(y) \} \) are bounded in probability. Also, by the assumption that \( F(\cdot | y) \) is uniformly continuous, the tightness of the second term on the right hand side of (2) would follow if the first term is shown to be tight.

Now, let

\[
L_n(x, y) = (nh_n)^{-1} \sum_{i=1}^{n} I(X_i+1 \leq x)K(\frac{y - X_i}{h_n}).
\]

Then the first term on the RHS of (2) becomes

\[
(nh_n)^{1/2}U_n^{-1}(y)\{ L_n(x, y) - F(x | y)\pi(y) \} = (nh_n)^{1/2}U_n^{-1}(y)\{ L_n(x, y) - E[L_n(x, y)] \} + (nh_n)^{1/2}U_n^{-1}(y)\{ E[L_n(x, y)] - F(x | y)\pi(y) \}. \tag{5}
\]
Now, by Lemma 8.8 of Robinson (1983) and C8(b), we see that
\[
\sup_{x \in \mathbb{R}} \{ (nh_n)^{1/2} \epsilon_n^{-1}(y) (E(L_n(x,n) - F(x \mid y) \pi(y)) \} \rightarrow 0
\]
as \( n \to \infty \). Hence it is sufficient to verify the tightness of the first term on the RHS of (5).

Let \( \beta \in T_i \) and consider \( S_n \) over \( [-\beta, \beta] \). Let
\[
M_{ni}(x,y) = I(X_{i+1} \leq x) K(x) - E(I(X_{i+1} \leq x)) K(x),
\]
i = 1, 2, ..., n. Further, let \( -\beta \leq x < z \leq \beta \) and consider
\[
E(M_{ni}(x,y) - M_{ni}(z,y))^2
\]
where \( f(.,.) \) is the stationary density corresponding to \( \{(S_i, Y_{i+1})\} \), with cdf \( F(.,.) \), \( c \) is such that \( K(.) \leq c \), and \( G_n(.) = F(y + ah_n, .) - F(y - ah_n, .) \).

We note that \( G_n(z) \) is an increasing function in \( z \). Let
\[
s_n(x \mid y) = (nh_n)^{-1/2} \sum_{i=1}^{n} M_{ni}(x,y).
\]
Since \( \{X_n\} \) is a stationary mixing sequence with geometric mixing coefficients \( \{\alpha_n\} \), the sequence \( \{(X_n, Y_{n+1})\} \) is also mixing with new mixing coefficients \( \{\mu_n\} \) (in fact, \( \mu_n = \alpha_{n+1}, n \geq 1 \)). Hence the sequence \( \{M_{ni}(x,y), i = 1,2,...,n; n \geq 1 \} \) is mixing for all \( x \). Therefore, following Lemma 1 of Section 22 of Billingsley (1968) (since the results in Section 22 of Billingsley (1968) remain valid for strong mixing sequences also (cf. Deo (1973))), we get
\[
E(\sum_{i=1}^{n} M_{ni}(x,y))^4 \leq c_1 \{ n^2 E^2(M_{ni}^2(x,y)) + n E(M_{ni}^2(x,y)) \}
\]
where \( c_1 \) depends on \( \{\mu_n\} \) alone. Further, we note that the condition \( \sum_{n=1}^{\infty} n^2 \mu_{n+1/2-\tau} < \infty \) for some \( \tau \) with \( 0 < \tau < 1/2 \), is satisfied.

Hence, using (6) and (7), we have
\[
E(\sum_{i=1}^{n} (M_{ni}(z,y) - M_{ni}(x,y))^4
\]
\[
\leq c_1 \{ n^2 c^4(G_n(z) - G_n(x))^2 + n c^2(G_n(z) - G_n(x)) \}.
\]
Now we note that
\[ n(G_n(z) - G_n(x)) = 2\alpha h_n \left( \frac{G_n(z)}{2\alpha h_n} - \frac{G_n(x)}{2\alpha h_n} \right). \]

Since \( nh_n \to \infty \) and \( \frac{G_n(z)}{2\alpha h_n} \to F(. \mid y) \pi(y) \), we have that \( n(G_n(z) - G_n(x)) \to \infty \). Hence, given \( \epsilon > 0 \), there exists an \( n_0 \) such that, for all \( n \geq n_0 \)
\[ \frac{\epsilon}{n} < G_n(z) - G_n(x) \]
and hence (8) implies
\[
E(s_n(z) - s_n(x))^4 \\
\leq \frac{c_1}{n^2 h_n^2} \{ n^2 c^4 (G_n(z) - G_n(x))^2 + nc^4 (G_n(z) - G_n(x)) \} \\
\leq \frac{c_1}{\epsilon} (G_n(z) - G_n(x))^2 \left( \frac{c^4 + c^2}{\epsilon} \right) \\
= \frac{c_2}{\epsilon} \left( \frac{G_n(z)}{h_n} - \frac{G_n(x)}{h_n} \right)^2 \tag{9}
\]

Let \( \delta > 0 \) such that \( x \leq z \leq x + \delta \). Then it is known that (cf. Lemma 1 on page 50 of Shorack and Wellner (1986) )
\[
\sup_{x \leq z \leq x + \delta} | s_n(z \mid y) - s_n(x \mid y) | \\
\leq | s_n(x + \delta \mid y) - s_n(x \mid y) | \\
+ \sup_{x \leq z \leq x + \delta} | (s_n(z \mid y) - s_n(x \mid y)) \wedge (s_n(x + \delta \mid y) - s_n(z \mid y)) | \\
= | s_n(x + \delta \mid y) - s_n(x \mid y) | + L \text{ say.}
\]

Then
\[
P \left( \sup_{x \leq z \leq x + \delta} | s_n(z \mid y) - s_n(x \mid y) | \geq 2\epsilon \right) \\
\leq P \left( | s_n(x + \delta \mid y) - s_n(x \mid y) | \geq \epsilon \right) + P(L \geq \epsilon) \\
\leq \frac{E | s_n(x + \delta \mid y) - s_n(x \mid y) |^4}{\epsilon^4} + P(L \geq \epsilon) \\
\leq \frac{c_2}{\epsilon^5} \left( \frac{G_n(x + \delta)}{h_n} - \frac{G_n(x)}{h_n} \right)^2 + P(L \geq \epsilon) \tag{10}
\]

Now
\[
P \left( | (s_n(z \mid y) - s_n(x \mid y)) \wedge (s_n(x + \delta \mid y) - s_n(z \mid y)) | \geq \epsilon \right)
\]
\[ \frac{E[(s_n(z \mid y) - s_n(x \mid y))^2(s_n(x + \delta \mid y) - s_n(z \mid y))^2]}{\epsilon^4} \]
\[ \leq \frac{E^{1/2}[(s_n(z \mid y) - s_n(x \mid y))^4]E^{1/2}[(s_n(x + \delta \mid y) - s_n(z \mid y))^4]}{\epsilon^4} \]
\[ \leq \frac{c_2}{\epsilon^5} \left( \frac{G_n(x + \delta)}{h_n} - \frac{G_n(x)}{h_n} \right)^2 \]

The reference for the first inequality above is Shorack and Wellner (1986), page 50.

Hence, by fluctuation inequality on page 17, Theorem 6.2 of Billingsley (1971), we have

\[ P(L \geq \epsilon) \leq \frac{c_3}{\epsilon^5} \left( \frac{G_n(x + \delta)}{h_n} - \frac{G_n(x)}{h_n} \right)^2 \]  
(11)

where \( c_3 \) is a universal constant.

Using (11) in (10), we get

\[ P \left( \sup_{x \leq x \leq x + \delta} |s_n(z \mid y) - s_n(x \mid y)| \geq 2\epsilon \right) \leq \frac{c_4c_2}{\epsilon^5} \left( \frac{G_n(x + \delta)}{h_n} - \frac{G_n(x)}{h_n} \right)^2 \]

where \( c_4 = 1 + c_3 \), proving the required tightness (cf. Theorem 6 on page 51 of Shorack and Wellner (1986)) and hence the theorem.

Now define the estimator of \( f(x \mid y) \) as

\[ \hat{f}_n(x \mid y) = (h_n)^{-1} \sum_{i=1}^{n} K \left( \frac{x - X_i}{h_n} \right) \]

and let

\[ \hat{f}_n(x \mid y) = (h_n)^{-1} \int K \left( \frac{x - u}{h_n} \right) dF(u \mid y). \]

**Theorem 2.3**: Let the conditions of Theorem 2.2 hold. Further assume that

**C10**: \( K \) is absolutely continuous and \( K' \) is bounded;

**C11**: \( \pi \) is uniformly continuous and for all \( y \) such that \( \pi(y) > 0 \), \( f(. \mid y) \) is uniformly continuous.

Then for all \( y \) such that \( \pi(y) > 0 \),

\[ \| \hat{f}_n^{1/2}(. \mid y) - f^{1/2}(.) \mid y \| \overrightarrow{P} 0. \]

If the argmin functional \( T \) is continuous at \( (f, \pi) \), then \( T(\hat{f}_n, \hat{\pi}_n) \overrightarrow{P} T(f, \pi) \) where

\[ \hat{\pi}_n(y) = (nh_n)^{-1} \sum_{i=1}^{n} K \left( \frac{y - X_i}{h_n} \right). \]
Proof: The fact that \( \hat{\pi}_n \) converges to \( \pi \) in \( L_1 \) norm can be proved on the lines of Theorem 3 of Beran (1977) and using the weak convergence of the empirical process \( \hat{F}_n \) (corresponding to \( \pi \)) (cf. Billingsley (1968), p. 197) of a stationary mixing sequence of variables.

Consider

\[
| \hat{f}_n(x \mid y) - \hat{f}_n(x \mid y) | = (h_n)^{-1} \left| \int K \left( \frac{x - u}{h_n} \right) d(\hat{F}_n(x \mid y) - F(x \mid y)) \right|
\]

Now \( S_n(x \mid y) = (nh_n)^{1/2} (\hat{F}_n(x \mid y) - F(x \mid y)) \). Therefore,

\[
| \hat{f}_n(x \mid y) - \hat{f}_n(x \mid y) | = (n^{1/2}(h_n)^{3/2})^{-1} \left| \int K \left( \frac{x - u}{h_n} \right) dS_n(u \mid y) \right|
\]

Now integration by parts yields

\[
\int K \left( \frac{x - u}{h_n} \right) dS_n(u \mid y) = (h_n)^{-1} \int S_n(u \mid y) K' \left( \frac{x - u}{h_n} \right) du.
\]

Hence

\[
| \hat{f}_n(x \mid y) - \hat{f}_n(x \mid y) | \leq (n^{1/2}(h_n)^{3/2})^{-1} \sup_u | S_n(u \mid y) | \int_a^b | K'(x) | dx.
\]

Similarly,

\[
| \hat{f}_n(x \mid y) - f(x \mid y) | = (h_n)^{-1} \left| \int K \left( \frac{x - u}{h_n} \right) f(u \mid y) du - f(x \mid y) \right|
\]

\[
\leq \sup_{-a \leq x \leq a} \left| f(x - h_n z \mid y) - f(x \mid y) \right|.
\]

Using C6, C11, C12 and Theorem 2.2 we see that

\[
| \hat{f}_n(x \mid y) | \leq | \hat{f}_n(x \mid y) | + | \hat{f}_n(x \mid y) - f(x \mid y) | \xrightarrow{P} 0 \text{ as } n \to \infty.
\]

The rest of the proof follows from Beran (1977), by identifying versions of \( \hat{f}_n \) defined on a suitable probability space such that

\[
\sup_{a \leq x \leq b} \left| \hat{f}_n(x \mid y) - f(x \mid y) \right| \to 0 \text{ a.s.}
\]

for all \( y \) such that \( \pi(y) > 0 \). Hence

\[
\lim_{n \to \infty} \| \hat{f}_n(\cdot \mid y) - f(\cdot \mid y) \| = 0 \text{ a.s.}
\]

for these versions and the second part of the theorem follows immediately by appealing to Theorem 2.1.
To prove the asymptotic normality of MMHDE we need the differentiability of the argmin functional. For notational conveniences, let $g_\theta = f_\theta^{1/2}$. We will assume

**D**: for a given $\theta \in \Omega$ and $y$, for which $\pi(y) > 0$ there exists a $k \times 1$ vector $g_\theta'(. \mid y)$ with elements in $L_2$ and a $k \times k$ matrix $g_\theta''(., y)$ with elements in $L_2$, such that for every $k \times 1$ real vector $e$ of unit Euclidean length and for every scalar $\alpha$ in a neighbourhood of zero,

$$
g_{\theta + \alpha e}(., y) = g_\theta(. \mid y) + \alpha e' g_\theta'(., y) + \alpha e' u_\alpha(., y)$$  \hspace{1cm} (13)

$$
g_{\theta + \alpha e}'(., y) = g_\theta'(., y) + \alpha g_\theta''(., y)e + \alpha v_\alpha(., y)e$$  \hspace{1cm} (14)

where $u_\alpha(., y)$ and $v_\alpha(., y)$ are respectively a $k \times 1$ vector and a $k \times k$ matrix such that their elements individually converge to zero in $L_2$ as $\alpha \rightarrow 0$. The superscript $t$ stands for transposition of a vector.

**Theorem 2.4**: Suppose that the condition D holds for all interior points $\theta \in \Omega$. Also suppose that the argmin functional $T$ for a given $f$ and $\pi$ exists uniquely and lies in $\text{int}(\Omega)$ and that

$$
\int g_{T(f, \pi)}''(x \mid y) = f^{1/2}(x \mid y)\pi(y) dxdy
$$

is a nonsingular matrix. Then for any sequence $\{(f_n, \pi_n)\}$ converging to $(f, \pi)$ in the modified Hellinger topology,

$$
T(f_n, \pi_n) = T(f, \pi) + \int \mu(f_n, \pi)(x \mid y) \left[ (f_n^{1/2}(x \mid y) - f^{1/2}(x \mid y)) \pi(y) \right] dxdy + a_n \int g_{T(f, \pi)}'(x \mid y) \left[ (f_n^{1/2}(x \mid y) - f^{1/2}(x \mid y)) \pi(y) \right] dxdy + f^{1/2}(x \mid y) \pi(y) dxdy
$$  \hspace{1cm} (15)

where

$$
\mu(f, \pi)(x \mid y) = -\left[ \int g_{T(f, \pi)}''(x \mid y) f^{1/2}(x \mid y) \pi(y) dxdy \right]^{-1} g_{T(f, \pi)}'(x \mid y)
$$

and $a_n$ is a $k \times k$ matrix of reals whose elements individually tend to zero as $n \rightarrow \infty$.

**Proof**: Let $\theta_n = T(f_n, \pi_n)$ and $\theta_0 = T(f, \pi)$. Since from our assumption, $\theta_0 \in \text{int}(\Omega)$ maximizes $\int g_\theta(x \mid y) f^{1/2}(x \mid y) \pi(y) dxdy$ and since D holds

$$
\lim_{\alpha \rightarrow 0} \frac{1}{\alpha} \int g_{\theta + \alpha e}(x \mid y) - g_\theta(x \mid y) f^{1/2}(x \mid y) \pi(y) dxdy = e^t \int g_\theta'(x \mid y) f^{1/2}(x \mid y) \pi(y) dxdy,
$$

for all real vectors $e$ of unit length and $\theta \in \text{int}(\Omega)$, it follows that

$$
\int g_{\theta_0}'(x \mid y) f^{1/2}(x \mid y) \pi(y) dxdy = 0.
$$
Similarly, for all \( n \geq 1 \),
\[
\int \int g_{0n}^*(x \mid y) f_n^{1/2}(x \mid y) \pi_n(y) dx dy = 0.
\]

Further, from D,
\[
0 = \int \int g_{0n}^*(x \mid y) f_n^{1/2}(x \mid y) \pi_n(y) dx dy \\
= \int \int [g_{0n}^*(x \mid y) + g_{0n}^*(x \mid y) (\theta_n - \theta_0) \\
+ v_n(x \mid y) (\theta_n - \theta_0)] f_n^{1/2}(x \mid y) \pi_n(y) dx dy
\]

For \( n \) large enough, after some simplifications, we get
\[
0_n - 0_0 = -\left( \int \int g_{0n}''(x \mid y) f_n^{1/2}(x \mid y) \pi(y) dx dy \right)^{-1} \\
\times \int \int g_{0n}''(x \mid y) \{ (f_n^{1/2}(x \mid y) - f^{1/2}(x \mid y)) \pi_n(y) \}
\]
\[
+ f^{1/2}(x \mid y) (\pi_n(y) - \pi(y)) dx dy \\
+ a_n \int \int g_{0n}''(x \mid y) \{ (f_n^{1/2}(x \mid y) - f^{1/2}(x \mid y)) \pi_n(y) \}
\]
\[
+ f^{1/2}(x \mid y) (\pi_n(y) - \pi(y)) dx dy.
\]

where
\[
a_n = \left[ \int \int g_{0n}''(x \mid y) f_n^{1/2}(x \mid y) \pi(y) dx dy \right]^{-1} \\
- \left[ \int \int (g_{0n}''(x \mid y) + v_n(x \mid y)) f_n^{1/2}(x \mid y) \pi_n(y) dx dy \right]^{-1}
\]

which goes to zero as \( n \to \infty \), as desired.

Now we state and prove the central limit theorem for MMHDE.

**Theorem 2.5**: Let C1 - C12 and D hold. Further, assume that

C13: \( T \) satisfies (15) and \( \mu_{(f,\cdot)}(\cdot \mid \cdot) \) has compact rectangle support \( J \) in \( \mathbb{R}^2 \) and \( \mu_{(f,\cdot)} \) is continuous on \( J \);

C14: \( \pi(y) > 0 \) for all \( x \)-sections of \( J \) such that \( (x, y) \) is in \( J \); \( f(\cdot \mid \cdot) > 0 \) on \( J \) and \( f(\cdot \mid \cdot) \) is absolutely continuous; \( \sup_x |f(x \mid y)| \) is finite for all \( y \) and the second partial derivative of \( f(x \mid y) \) with respect to \( x \) is bounded.

Then
\[
(n h_n)^{1/2} \left( T(\hat{f}_n, \hat{\pi}_n) - T(f, \pi) \right) \xrightarrow{D} T_0
\]

where \( T_0 \) is \( k \times 1 \) normal vector with mean zero and variance-covariance matrix
\[
\frac{K_0}{4} \int \int F(x \mid y)(1 - F(x \mid y)) \pi(y) \mu_{(f,\cdot)}(x \mid y) \mu_{(f,\cdot)}(x \mid y) \frac{dx dy}{f(x \mid y)}
\]
and

\[ K_0 = \int_{-\infty}^{\infty} K^2(z) \, dz. \]

**Proof:** From (15) we get

\[
(n h_n)^{1/2} (T(\hat{f}_n, \hat{\pi}_n) - T(f, \pi)) = (n h_n)^{1/2} \int \int \mu_{(f, \pi)}(x \mid y) \{ (\hat{f}_n^{1/2}(x \mid y) - f^{1/2}(x \mid y))\hat{\pi}_n(y) \} \, dx \, dy
\]

\[
+ (n h_n)^{1/2} \int \int \mu_{(f, \pi)}(x \mid y) \{ f^{1/2}(x \mid y)(\hat{\pi}_n(y) - \pi(y)) \} \, dx \, dy
\]

\[
+ (n h_n)^{1/2} A_n \int \int g_{T(f, \pi)}(x \mid y) \{ (\hat{f}_n^{1/2}(x \mid y) - f^{1/2}(x \mid y))\hat{\pi}_n(y) \} \, dx \, dy
\]

\[
+ (n h_n)^{1/2} A_n \int \int g_{T(f, \pi)}(x \mid y) \{ f^{1/2}(x \mid y)(\hat{\pi}_n(y) - \pi(y)) \} \, dx \, dy
\]

(16)

where \( A_n \xrightarrow{P} 0 \).

From the proof of Theorem 4 of Beran (1977), with the above assumptions, we see that the second and the fourth terms in (16) go to zero in probability. If we prove that the first term converges in distribution to the normal variable as given in the statement, the proof would be complete in view of the fact that \( A_n \xrightarrow{P} 0 \).

Making use of the following algebraic identity, for \( b \geq 0 \) and \( a > 0 \),

\[
\sqrt{b} - \sqrt{a} = \frac{b - a}{2\sqrt{a}} - \frac{(b - a)^2}{2\sqrt{a}(\sqrt{b} + \sqrt{a})^2},
\]

we can write the first term on the right of equation (16) as

\[
(n h_n)^{1/2} \int \int \mu_{(f, \pi)}(x \mid y) \frac{\hat{\pi}_n(y)(\hat{f}_n(x \mid y) - f(x \mid y))}{2f^{1/2}(x \mid y)} \, dx \, dy + R_n \quad \text{(say)}. \tag{17}
\]

Let \( \delta = \inf_{(x, y) \in J} f(x \mid y) > 0 \). Then

\[
| R_n | \leq (n h_n)^{1/2} \frac{\delta^{-3/2}}{2} \int \int | \mu_{(f, \pi)}(x \mid y) | (\hat{f}_n(x \mid y) - f(x \mid y))^2 \hat{\pi}_n(y) \, dx \, dy
\]

\[
\leq \frac{\delta^{-3/2}}{2} (W_{1n} + W_{2n})
\]

where

\[
W_{1n} = (n h_n)^{1/2} \int \int | \mu_{(f, \pi)}(x \mid y) | (\hat{f}_n(x \mid y) - f(x \mid y))^2 \hat{\pi}_n(y) \, dx \, dy,
\]

\[
W_{2n} = (n h_n)^{1/2} \int \int | \mu_{(f, \pi)}(x \mid y) | (\hat{f}_n(x \mid y) - f(x \mid y))^2 \hat{\pi}_n(y) \, dx \, dy,
\]

where \( \hat{f}_n(\cdot \mid \cdot) \) is as given in Theorem 2.3.
Proceeding on the lines of the proof of Theorem 4 of Beran (1977), with the required modifications, we see that both $W_{1n}$ and $W_{2n}$ converge to zero in probability.

Now, the first term of (17) can be written as $U_{1n} + U_{2n}$ where

$$U_{1n} = (nh_n)^{1/2} \int \int \mu_{(f, \pi)}(x \mid y) \frac{\hat{f}_n(x \mid y) - \tilde{f}_n(x \mid y)}{2f^{1/2}(x \mid y)} \hat{\pi}_n(y) dx dy,$$

$$U_{2n} = (nh_n)^{1/2} \int \int \mu_{(f, \pi)}(x \mid y) \frac{f_n(x \mid y) - f(x \mid y)}{2f^{1/2}(x \mid y)} \hat{\pi}_n(y) dx dy.$$

As mentioned above, $U_{2n} \xrightarrow{P} 0$. Now letting

$$\lambda(., .) = \mu_{(f, \pi)}(., .) (2f^{1/2}(., .))^{-1},$$

$U_{1n}$ can be written as

$$\int \int \{ \int \lambda(z + wh_n \mid y) \hat{\pi}_n(y)K(w)dw \} dS_n(z \mid y) dy.$$

Arguing similarly as in Theorem 4 of Beran, we have that

$$U_{1n} - \int \int \lambda(z \mid y) \hat{\pi}_n(y) dS_n(z \mid y) dy \xrightarrow{P} 0,$$

and the limit of the second term above is

$$\int \int \mu_{(f, \pi)}(x \mid y) \mu_{(f, \pi)}(x \mid y) \frac{\pi(y)}{2f^{1/2}(x \mid y)} dS(x \mid y) dy$$

where $S(x \mid y)$ is a normal variable with mean zero and variance $\sigma^2(x, y)$ as given in Theorem 2.2 gives us the required result.

**Remarks:** We note here that the choice of the kernel $K$ plays role in the asymptotic variance of the MMHDE. Let $\mathcal{K}$ be the class of all kernels $K$ which satisfy the conditions given in this paper. Then the optimal MMHDE is the one corresponding to $K_1 \in \mathcal{K}$ such that $\inf_{K \in \mathcal{K}} \int K^2(z)dz = \int K_1^2(z)dz$. The kernel $K_1$ turns out to be the Epanechnikov kernel (cf. Subsection 3.3.2 of Silverman (1986)) defined by

$$K_1(t) = \frac{3}{4\sqrt{5}} \left(1 - \frac{t^2}{5}\right)$$

for all $-\sqrt{5} \leq t \leq \sqrt{5}$, and zero elsewhere.

**Acknowledgements:** The author wishes to thank his supervisor Dr. U. V. Naik-Nimbalkar for her constructive comments. Part of this work was supported by the University Grants Commission (India), through a research fellowship.
Bibliography


SUMMARY

In this paper, we consider the semiparametric estimation problem for stationary Markov chains with continuous state-space. This is a generalization of the work of Beran (1977) which deals with the i.i.d. setup. We assume a parametric model for the transition density of a stationary Markov chain and assume that the parameter space is compact. The proposed estimation procedure involves the modified minimum Hellinger distance between two transition densities. We assume that the Markov chain is geometrically ergodic which makes the chain strong mixing. The estimator gotten from the proposed method uses kernel estimators. Then, a weak convergence result is proved for the empirical process corresponding to the estimator of the conditional distribution function. Using this result we then prove the consistency and asymptotic normality of the proposed modified minimum Hellinger distance estimator (MMHDE). It is shown that among the class of the MMHDEs, the MMHDE corresponding to the Epanechnikov kernel is the optimal estimator in that, its asymptotic variance is the minimum.