LECTURE 6

Probability Density and Regression Estimation in the Case of Short-Range Dependence

A result of Bradley (1986) adapted for kernel density estimates assumes that (X_k) is a strictly stationary sequence with absolutely continuous marginal distribution. The marginal density is assumed continuous and positive. The joint probability distribution of (X_0, X_k) is also assumed absolutely continuous with a continuous joint density. The weight function of a kernel estimate ω is assumed to be nonnegative Borel and to satisfy (i) $\int \omega(u)\,du = 1$; (ii) $\int \omega^{2+\delta}(u)\,du < \infty$ for some $\delta > 0$; (iii) ω has bounded support. If the process (X_k) has asymptotic correlation zero property and $\sum \rho(2^n) < \infty$, where $\rho(j)$ is the correlation of X_0 and X_j, the probability density function estimates

$$f_n(x_s) = (nb_n)^{-1} \sum_{k=1}^{n} \omega\left(\frac{x_s - X_k}{b_n}\right),$$

$s = 1, 2, \ldots, L$, are such that $(nb_n)^{1/2}(f_n(x_s) - Ef_n(x_s))$ if $n \to \infty$, $b_n \to 0$, $nb_n \to \infty$ are asymptotically jointly normal and independent with means zero and variances

$$f(x_s)\int \omega^2(u)\,du.$$

The argument for this result is based on the extension of a central limit theorem for stationary sequences with asymptotic correlation zero condition to triangular arrays.

Here we will see to what extent results of a local character for probability density and regression estimates hold just as in the independent case when one has appropriate conditions, one of which is some form of short-range dependence. In this section a version of results on density and regression estimation in the case of short-range dependence is presented that corrects statements and derivations given in Rosenblatt [(1985), Chapter VII]. Let $\{X_j\}$ be a strictly stationary sequence that is strongly mixing. Assume that the joint distributions up to fourth order of the variables X_j are absolutely continuous with continuous bounded probability density functions. We shall also assume something like a set of cumulant-like conditions for joint probability densities. Let the joint densities of X_0, X_j be ${}_j f(x_1, x_2)$, those of X_0, X_j, X_k be ${}_{jk} f(x_1, x_2, x_3)$ and those of X_0, X_j, X_k, X_l be ${}_{jkl} f(x_1, x_2, x_3, x_4)$. The cumulant-like conditions for second and fourth order densities are

$$\sum_j{}' \left| {}_j f(x_0, x_1) - f(x_0) f(x_1) \right| < \infty,$$

$$\begin{aligned}\sum_{j_1, j_2, j_3}{}' \Big| & f_{j_1, j_2, j_3}(x_0, x_1, x_2, x_3) - f_{j_1}(x_0, x_1) f_{j_3 - j_2}(x_2, x_3) \\ & - f_{j_2}(x_0, x_2) f_{j_3 - j_1}(x_1, x_3) - f_{j_3}(x_0, x_3) f_{j_2 - j_1}(x_1, x_2) \\ & - f(x_0) f_{j_2 - j_1, j_3 - j_1}(x_1, x_2, x_3) - f(x_1) f_{j_2, j_3}(x_0, x_2, x_3) \\ & - f(x_2) f_{j_1, j_3}(x_0, x_1, x_3) - f(x_3) f_{j_1, j_2}(x_0, x_1, x_2) \\ & + 2 f(x_0) f(x_1) f_{j_3 - j_2}(x_2, x_3) + 2 f(x_0) f(x_2) f_{j_3 - j_1}(x_1, x_3) \\ & + 2 f(x_0) f(x_3) f_{j_2 - j_1}(x_1, x_2) + 2 f(x_1) f(x_2) f_{j_3}(x_0, x_3) \\ & + 2 f(x_1) f(x_3) f_{j_2}(x_0, x_2) + 2 f(x_2) f(x_3) f_{j_1}(x_0, x_1) \\ & - 6 f(x_0) f(x_1) f(x_2) f(x_3) \Big| < \infty, \end{aligned} \tag{6.1}$$

assumed bounded functions. The primed summations denote sums over different subscripts that are all nonzero. It is also assumed that corresponding conditions hold for all expressions obtained from the fourth order cumulant-like restraints by identifying some of the x's. The condition obtained, for example, by identifying x_0 with x_2 and x_1 with x_3 is

$$\begin{aligned}\sum_j{}' \Big| & f_j(x_0, x_1) - f(x_0) f(x_1) - 2 f_j(x_0, x_1)^2 \\ & - 2 f(x_0) f_j(x_0, x_1) - 2 f(x_1) f_j(x_0, x_1) \\ & + 2 f(x_0)^2 f(x_1) + 8 f(x_0) f(x_1) f_j(x_0, x_1) \\ & + 2 f(x_1)^2 f(x_0) - 6 f(x_0)^2 f(x_1)^2 \Big| < \infty. \end{aligned} \tag{6.2}$$

It is understood that $f_n(x)$ is a kernel density estimate

$$f_n(x) = \frac{1}{nb(n)} \sum_{j=1}^{n} \omega\left(\frac{x - X_j}{b(n)} \right)$$

with ω a bounded piecewise continuous integrable weight function such that

$$\int \omega(x)\, dx = 1.$$

Under these conditions the normed and centered statistics

$$\{nb(n)\}^{1/2}\left[f_n(x^{(i)}) - Ef_n(x^{(i)})\right],$$

$i = 1, \ldots, m$, are asymptotically independent and jointly normal as $n \to \infty$ with variances

$$f(x^{(i)}) \int \omega^2(v)\, dv,$$

$i = 1, \ldots, m$, if $b(n) \to 0$, $nb(n) \to \infty$. Just as in the independent case,

$$E\omega\big(b(n)^{-1}(x - X)\big) = b(n) \int \omega(v) f(x - b(n)v)\, dv,$$

$$E\omega^2\big(b(n)^{-1}(x - X)\big) = b(n) \int \omega^2(v) f(x - b(n)v)\, dv,$$

$$E\Big\{\omega\big(b(n)^{-1}(x - X)\big)\omega\big(b(n)^{-1}(y - X)\big)\Big\}$$
$$= b(n) \int \omega\big(b(n)^{-1}(y - x) + v\big)\omega(v) f(x - b(n)v)\, dv.$$

Further

$$\operatorname{cov}\Big\{\omega\big(b(n)^{-1}(x - X_0)\big), \omega\big(b(n)^{-1}(y - X_j)\big)\Big\}$$
$$= b(n)^2 \int \omega(v)\omega(u)\big\{{}_j f(x - b(n)u, y - b(n)v) \tag{6.3}$$
$$-f(x - b(n)u) f(y - b(n)v)\big\}\, du\, dv,$$

$j = 1, 2, \ldots$. These equalities imply that

$$\operatorname{var}\left[\sum_{j=1}^{m} \omega\big(b(n)^{-1}(x - X_j)\big)\right]$$
$$= \sum_{u=-m}^{m} (m - |u|)\operatorname{cov}\Big\{\omega\big(b(n)^{-1}(x - X_0)\big), \omega\big(b(n)^{-1}(x - X_u)\big)\Big\} \tag{6.4}$$
$$= mb(n) \int f(x - b(n)v)\omega^2(v)\, dv + O\big(mb(n)^2\big).$$

Suppose we consider estimating the fourth central moment

$$E\left|\sum_{j=1}^{m} g_j\right|^4 = \sigma^4\left(\sum_{j=1}^{m} g_j\right) + \operatorname*{cum}_4\left(\sum_{j=1}^{m} g_j\right),$$

where

$$g_j = w\left(b(n)^{-1}(x - X_j)\right) - Ew\left(b(n)^{-1}(x - X_j)\right).$$

The fourth cumulant

$$\begin{aligned}\operatorname*{cum}_4\left(\sum_{j=1}^{m} g_j\right) &= \sum_{j=1}^{m} \operatorname*{cum}_4(g_j) + 3 \sum_{j_1, j_2 = 1}^{m}{}'' \operatorname*{cum}_4\left(g_{j_1}^2, g_{j_2}^2\right) \\ &+ 6 \sum_{j_1, j_2, j_3 = 1}^{m}{}'' \operatorname*{cum}_4\left(g_{j_1}^2, g_{j_2}, g_{j_3}\right) \\ &+ \sum_{j_1, j_2, j_3, j_4 = 1}^{m}{}'' \operatorname*{cum}_4\left(g_{j_1}, g_{j_2}, g_{j_3}, g_{j_4}\right).\end{aligned} \tag{6.5}$$

The double primed summations indicate sums over different subscripts. The first summation on the right of (6.5) is of order of magnitude $m(b(n))$. The second, third and fourth terms on the right of (6.5) are of the order of $m(b(n))^2$, $m(b(n))^3$ and $m(b(n))^4$, respectively, by the fourth order cumulant-like conditions and its contraction obtained by identifying some of the x's. Let $b(n) \to 0$ and $nb(n) \to \infty$ as $n \to \infty$. Choosing the sequence $m(n) = o(n)$, but such that $mb(n) \to \infty$ as $n \to \infty$, implies that the Liapounov condition of the central limit theorem of Lecture 5 will be satisfied with $\delta = 2$. The theorem implies asymptotic normality of

$$\{nb(n)\}^{1/2}[f_n(x) - Ef_n(x)]$$

with limiting variance $f(x)\int\omega^2(v)\,dv$. The equation (6.3) implies that $f_n(x)$, $f_n(y)$ are asymptotically uncorrelated if $f(x), f(y) > 0$ and $x \neq y$. Using the same argument for any linear combination of $f_n(x^{(i)})$ at distinct values $x^{(i)}$ implies joint asymptotic normality and independence of the density estimates with variances (6.3). An early discussion of density estimation in the context of Markov processes assuming a Doeblin condition can be found in Roussas (1969).

Let us consider the case of a regression function

$$r(x) = E(X_{n+1} | X_n = x)$$

with (X_n) as before a stationary sequence. A plausible estimate would be given by

$$\begin{aligned} r_n(x) &= \{nb(n)\}^{-1} \sum_{j=1}^{n} X_{j+1}\omega\left(b(n)^{-1}(x - X_j)\right) \\ &\quad \times \left[(nb(n))^{-1} \sum_{j=1}^{n} \omega\left(b(n)^{-1}(x - X_j)\right)\right]^{-1} \\ &= a_n(x)/f_n(x). \end{aligned}$$

We introduce the following family of functions and assume they are well defined and uniformly bounded:

$$E(X_1^a;\, X_0 = x) = h^{(a)}(x), \qquad a = 1,2,3,4,$$

$$E\left(X_1^a X_{j+1}^b;\, X_0 = x_1,\, X_j = x_2\right) = {}_j h^{(a,b)}(x_1, x_2), \qquad a, b = 1,2,3,$$

$$E\left(X_1^a X_{j+1}^b X_{k+1}^c;\, X_0 = x_1,\, X_j = x_2,\, X_k = x_3\right)$$
$$= {}_{j,k} h^{(a,b,c)}(x_1, x_2, x_3), \qquad a, b, c = 1,2,$$

$$E(X_1 X_{j+1} X_{k+1} X_{l+1};\, X_0 = x_1,\, X_j = x_2,\, X_k = x_3,\, X_l = x_4)$$
$$= {}_{j,k,l} h(x_1, x_2, x_3, x_4).$$

It is assumed also that

$$\sum_j{}' \left| {}_j h^{(a,b)}(x_1, x_2) - h^{(a)}(x_1) h^{(b)}(x_2) \right| < \infty,$$

$$\begin{aligned}
\sum_{j_1, j_2, j_3}{}' \Big| & {}_{j_1, j_2, j_3} h(x_0, x_1, x_2, x_3) - {}_{j_1} h(x_0, x_1)_{j_3 - j_2} h(x_2, x_3) \\
& - {}_{j_2} h(x_0, x_2)_{j_3 - j_1} h(x_1, x_3) - {}_{j_3} h(x_0, x_3)_{j_2 - j_1} h(x_1, x_2) \\
& - h(x_0)_{j_2 - j_1, j_3 - j_1} h(x_1, x_2, x_3) - h(x_1)_{j_2, j_3} h(x_0, x_2, x_3) \\
& - h(x_2)_{j_1, j_3} h(x_0, x_1, x_3) - h(x_3)_{j_1, j_2} h(x_0, x_1, x_2) \\
& + 2h(x_0) h(x_1)_{j_3 - j_2} h(x_2, x_3) + 2h(x_0) h(x_2)_{j_3 - j_1} h(x_1, x_3) \\
& + 2h(x_0) h(x_3)_{j_2 - j_1} h(x_1, x_2) + 2h(x_1) h(x_2)_{j_3} h(x_0, x_3) \\
& + 2h(x_1) h(x_3)_{j_2} h(x_0, x_2) + 2h(x_2) h(x_3)_{j_1} h(x_0, x_1) \\
& - 6h(x_0) h(x_1) h(x_2) h(x_3) \Big| < \infty
\end{aligned} \tag{6.6}$$

are bounded functions. It is also assumed that corresponding conditions hold for all expressions obtained from the fourth order cumulant-like restraints (6.6) by identifying some of the x's. The condition obtained, for example, by identifying x_0 with x_2 and x_1 with x_3 is

$$\begin{aligned}
\sum_j{}' \Big| & {}_j h^{(2,2)}(x_0, x_1) - h^{(2)}(x_0) h^{(2)}(x_1) \\
& - 2{}_j h(x_0, x_1)^2 - 2h(x_0)_j h^{(1,2)}(x_0, x_1) \\
& - 2h(x_1)_j h^{(2,1)}(x_0, x_1) + 2h(x_0)^2 h^{(2)}(x_1) \\
& + 8h(x_0) h(x_1)_j h(x_0, x_1) + 2h(x_1)^2 h^{(2)}(x_0) \\
& - 6h(x_0)^2 h(x_1)^2 \Big| < \infty.
\end{aligned}$$

If these conditions as well as the conditions for the density function estimate are satisfied, one will have

$$(6.7) \qquad (nb(n))^{1/2}\{r_n(x^{(i)}) - Ea_n(x^{(i)})/Ef_n(x^{(i)})\},$$

$j = 1, \dots, m$, jointly asymptotically normal and independent as $n \to \infty$ with variances

$$\left[E\left(X_{j+1}^2 | X_j = x^{(i)}\right) - \left\{E\left(X_{j+1} | X_j = x^{(i)}\right)\right\}^2\right] f(x^{(i)})^{-1} \int \omega^2(v)\, dv,$$

$i = 1, \dots, m$, if $b(n) \to 0$, $nb(n) \to \infty$ and $f(x^{(i)}) > 0$. We can write

$$\begin{aligned}
r_n(x) &= \frac{a_n(x)}{f_n(x)} = \frac{a_n(x) - Ea_n(x) + Ea_n(x)}{f_n(x) - Ef_n(x) + Ef_n(x)} \\
&= \{a_n(x) - Ea_n(x) + Ea_n(x)\} \\
&\quad \times (Ef_n(x))^{-1}\left\{1 - \frac{f_n(x) - Ef_n(x)}{Ef_n(x)} + O(f_n(x) - Ef_n(x))^2\right\} \\
&= \frac{Ea_n(x)}{Ef_n(x)} + (Ef_n(x))^{-1}(a_n(x) - Ea_n(x)) \\
&\quad - \frac{Ea_n(x)}{(Ef_n(x))^2}(f_n(x) - Ef_n(x)) \\
&\quad + O(f_n(x) - Ef_n(x))^2 + O(a_n(x) - Ea_n(x))^2.
\end{aligned}$$

It is clear that

$$(6.8) \quad E\left\{X_{j+1}\omega\left(b(n)^{-1}(x - X_j)\right)\right\} = b(n)\int \omega(\alpha) h(x - b(n)\alpha)\, d\alpha,$$

$$(6.9) \quad E\left\{X_{j+1}^2\omega^2\left(b(n)^{-1}(x - X_j)\right)\right\} = b(n)\int \omega^2(\alpha) h^{(2)}(x - b(n)\alpha)\, d\alpha.$$

Also

$$\begin{aligned}
&E\left\{X_1 X_{j+1}\omega\left(b(n)^{-1}(x - X_0)\right)\omega\left(b(n)^{-1}(x' - X_j)\right)\right\} \\
(6.10) \quad &= b(n)^2\int \omega(\alpha_0)\omega(\alpha_1)_j h(x - \alpha_0 b(n), x' - \alpha_1 b(n))\, d\alpha_0\, d\alpha_1,
\end{aligned}$$

$j = 1, 2 \dots$. The relations (6.8), (6.9) and (6.10) have as a consequence

$$\begin{aligned}
&\operatorname{var}\left[\sum_{j-1}^{m} X_{j+1}\omega\left(b(n)^{-1}(x - X_j)\right)\right] \\
&\quad = mb(n)\int \omega^2(v) h^{(2)}(x - b(n)v)\, dv + O\left(mb(n)^2\right).
\end{aligned}$$

Set

$$q_j = X_{j+1}\omega\left(b(n)^{-1}(x - X_j)\right) - E\left[X_{j+1}\omega\left(b(n)^{-1}(x - X_j)\right)\right].$$

The fourth central moment

$$E\left|\sum_{j=1}^{m} q_j\right|^4 = \sigma^4\left(\sum_{j=1}^{m} q_j\right) = \operatorname*{cum}_4\left(\sum_{j=1}^{m} q_j\right)$$

and

$$\begin{aligned}\operatorname*{cum}_4\left(\sum_{j=1}^{m} q_j\right) &= \sum_{j=1}^{m} \operatorname*{cum}_4(q_j) + 3\sum_{j_1,j_2=1}^{m}{}'' \operatorname*{cum}_4\left(q_{j_1}^2, q_{j_2}^2\right) \\ &\quad + 6\sum_{j_1,j_2,j_3=1}^{m}{}'' \operatorname*{cum}_4\left(q_{j_1}^2, q_{j_2}, q_{j_3}\right) \\ &\quad + \sum_{j_1,j_2,j_3,j_4=1}^{m}{}'' \operatorname*{cum}_4\left(q_{j_1}, q_{j_2}, q_{j_3}, q_{j_4}\right),\end{aligned} \tag{6.11}$$

where as before the double primed summation is a sum over distinct subscript values. The relations (6.6) as in the earlier corresponding discussion for density functions imply that the first sum on the right of (6.11) is of order $mb(n)$, the second of order $m(b(n))^2$, the third of order $m(b(n))^3$ and the last of order $mb(n)^4$. The completion of the argument for joint asymptotic normality and independence of

$$\{nb(n)\}^{1/2}\left[a_n(x_i) - Ea_n(x_i)\right], \qquad i = 1, \dots, m,$$

as $b_n \to 0$, $nb_n \to \infty$ with variances

$$h^{(2)}(x^{(i)})\int \omega^2(\alpha)\,d\alpha$$

follows just as in the case of density estimates. Also

$$\begin{aligned}&E\left[X_1\omega\left(b(n)^{-1}(x - X_0)\right)\omega\left(b(n)^{-1}(x' - X_0)\right)\right] \\ &\quad = b(n)\int h(x - b(n)\alpha)\omega(\alpha)\omega\left(b(n)^{-1}(x' - x) + \alpha\right)d\alpha\end{aligned}$$

so that

$$\lim_{n\to\infty} nb(n)\operatorname{cov}\left[a_n(x), f_n(x')\right] = \delta_{x,x'}h(x)\int \omega^2(\alpha)\,d\alpha.$$

One can apply the central limit theorem of Lecture 5 to linear combinations of the $a_n(x^{(i)})$ and $f_n(x^{(i)})$, $i = 1, \dots, m$, using the same types of estimates of moments as those employed earlier and thereby demonstrate joint asymptotic normality. The fact that $\lim_{n\to\infty} Ea_n(x) = h(x)$ and $\lim_{n\to\infty} f_n(x) = f(x)$ together with the observations just made imply that (6.7) holds. Results related to the questions dealt with in this lecture on curve estimation in the context of short-range dependence can be found in Masry (1989) as well as Györfi, Härdle, Sarda and Vieu (1989).