COLLOQUIA MATHEMATICA SOCIETATIS JÁNOS BOLYAI
32. NONPARAMETRIC STATISTICAL INFERENCE,
BUDAPEST (HUNGARY), 1980.

AN APPLICATION OF THE METHOD OF SIEVES: FUNCTIONAL ESTIMATOR FOR THE DRIFT OF A DIFFUSION

S. GEMAN*

1. INTRODUCTION

From an observation of a sample path of a diffusion process one can construct consistent estimators for the diffusion drift. If the form of the drift function is known up to a finite collection of parameters then it is possible to use maximum likelihood, and obtain consistent and asymptotically normal estimators (see Brown and Hewitt [3], Feigin [4], Lee and Kozin [8], and Lipster and Shiryayev [9]). Even when no parametric form is known, concistent (and in some cases asymptotically normal) estimators for the value of the drift at a fixed argument have been developed (Banon [1], Banon and Nguyen [2], and Nguyen and Pham [10]).

This paper is also about nonparametric estimation of the drift function. But the estimator developed here is distinguished from the nonparametric estimators cited above by:

*Supported by the Department of the Army under grant DAAG 29-80-K-0006

(1) being a *functional* estimator,

(2) being based on the principle of maximum likelihood.

By a "functional estimator" I mean that at each t the estimation procedure produces a function defined on a prescribed interval, and, as $t \to \infty$, this estimator converges (a.s.) to the drift in the sense of a function space norm. To be more precise about (2), let us look at a diffusion equation and an associated likelihood function:

(1.1) $$dx_t = g(x_t)dt + \sigma dw_t \qquad x_0 = x_o.$$

w_t is a standard (one-dimensional) Brownian motion and x_0 is a constant. g and σ are assumed to be unknown; we wish to estimate g from an observation of a sample path of x_t. It is well known that the distribution of x_s, $s \in [0, t]$ is absolutely continuous with respect to the distribution of σw_s, $s \in [0, t]$ (assuming some mild regularity condition on g). A likelihood function for the process x_s, $s \in [0, t]$ is the Radon – Nikodym derivative:

(1.2) $$\exp\left\{\int_0^t g(x_s)\,dx_s - \frac{1}{2}\int_0^t g(x_s)^2\,ds\right\}.$$

The "natural" estimator for g would maximize (1.2) over a suitable parameter space, most appropriately the space of uniformly Lipschitz continuous functions. But the maximum of the likelihood is not attained, either in this or in any other of the usual function spaces. Some difficulty is not unexpected: maximum likelihood typically fails in nonparametric settings. To preserve the principle of maximum likelihood in nonparametric problems, Grenander [7] suggests a "Method of Sieves": maximize the likelihood over a subset of the parameter space, allowing the subset to grow as the observations increase. The method of sieves produces consistent estimators for a wide variety of nonparametric problems. This paper presents an application of the method to nonparametric estimation of the drift of a diffusion.

Consistency of the constrained maximum likelihood estimator requires an appropriate rate of growth for the subsets from which the estimator is taken. In particular, the "sieve" must not grow too rapidly. Very

The observation here is the sample path of a diffusion, and can not be viewed as a sequence of ~~iid obs's~~ independent identically distributed random variables. The results in [6] do

general sufficient conditions on sieve growth can be formulated when the observations are independent and identically distributed (possibly infinite dimensional) random variables. General conditions of this type are established in Geman and Hwang [6], where, among other applications, there is presented a consistent functional estimator for the drift of a diffusion. The estimator to be discussed here is an improvement, mainly in the sense that it uses essentially all of the information available in the sample path observation. The previous estimator ignores a portion of this information in order to define observations which are independent and identically distributed, and to thereby utilize the above mentioned consistency results (see [6] for details). The improvement preclused an application of these results; it is no longer possible to view the observations as i.i.d. random variables.

Pretty much the minimal conditions for consistent estimation of g are:

(1) conditions for the existence and uniqueness of the solution to (1.1),

(2) that the process x_t be recurrent (i.e. for every level λ there exists $t_1, t_2, \ldots$ increasing to infinity such that $x_{t_i} = \lambda$ for every i).*

It is under these conditions that the consistency of the estimator will be demonstrated.

2. STATEMENT OF MAIN RESULT

(1.1):

Start with the usual assumption for existence and uniqueness in (1.):

A1. For some constant L

$$|g(x) - g(y)| \leq L\,|x - y|$$

for all $x, y \in R$.

*The necessity of (2) derives from the fact that the measures on $C[0,t]$ induced by the solution to (1.1) and indexed by α, form an absolutely continuous family for any finite t (under some well-known regularity conditions on g). The situation for σ is quite different. σ can in principle be determined from any (arbitrarily small) interval of observation of x_t.

And, an assumption which is equivalent to recurrence (see Friedman [5], Chapter 9):

A2. If

$$\theta(x) = \int_0^x \exp\left\{-\frac{2}{\sigma^2}\int_0^z g(u)\,du\right\} dz$$

then $\theta(+\infty) = +\infty$ and $\theta(-\infty) = -\infty$.

The estimator will approximate g on a fixed, but arbitrary, interval $[\lambda_1, \lambda_2]$. For this purpose the likelihood function, (1.2), will be replaced by a function which depends only on the behavior of x_t during the time spent in the interval $[\lambda_1, \lambda_2]$. Specifically, we will seek to maximize

$$\begin{aligned} & f_t(x_.;\alpha) \\ & \equiv \exp\left\{\int_0^t I_{[\lambda_1,\lambda_2]}(x_s)\alpha(x_s)\,dx_s - \frac{1}{2}\int_0^t I_{[\lambda_1,\lambda_2]}(x_s)\alpha(x_s)^2\,ds\right\} \\ & = \exp\left\{\int_0^t I_{[\lambda_1,\lambda_2]}(x_s)\left(\alpha(x_s)g(x_s) - \frac{1}{2}\alpha(x_s)^2\right) ds \right. \\ & \left. + \sigma\int_0^t I_{[\lambda_1,\lambda_2]}(x_s)\alpha(x_s)\,dw_s\right\}. \end{aligned} \tag{2.1}$$

As with the full likelihood function, (1.2), nothing useful comes from an unconstrained maximization of (2.1). One remedy is to introduce a sieve, S_t, parameterized by the length of the interval of observation $[0, t]$:

$$S_t = \left\{\sum_{j=1}^{m_t} a_j\psi_j(x) \colon \sum_{j=1}^{m_t} |a_j| \leqslant k_1 (\log m_t)^{k_2}\right\}$$

where

1. m_t is a nondecreasing sequence of integers governing the size of the sieve at time t, and k_1 and k_2 are arbitrary positive constants,

2. $\{\psi_j(x)\}_{j=1}^{\infty}$ is any sequence of measurable functions satisfying

   a. $|\psi_j(x)| \leqslant 1 \quad x \in [\lambda_1, \lambda_2]$ and all j

b. Every continuous function f on $[\lambda_1, \lambda_2]$, satisfying $f(\lambda_1) = f(\lambda_2)$, can be uniformly approximated by a linear combination of $\psi_1(x), \psi_2(x), \ldots$.

For example, $\{\psi_j(x)\}_{j=1}^{\infty}$ may be the trigonometric polynomials $\{\exp[2\pi i l(\frac{x-\lambda_1}{\lambda_2-\lambda_1})]\}_{l=-\infty}^{\infty}$ (with a change of indices), or the polynomials $\{(\frac{x}{|\lambda_1|+|\lambda_2|})^l\}_{l=0}^{\infty}$. Note that an implication of b is that the span of $\{\psi_j(x)\}_{j=1}^{\infty}$ is dense in $L_2([\lambda_1, \lambda_2], B, dm)$, for any finite measure dm (B = Borel sets in $[\lambda_1, \lambda_2[$).

Define M_t to be the set of functions in S_t which maximize the "likelihood" (2.1):

$$M_t = \{\alpha \in S_t : f_t(x_.; \alpha) = \sup_{\beta \in S_t} f_t(x_.; \beta)\}.$$

(If we write $\alpha = \sum_{j=1}^{m} a_j \psi_j$, then $f_t(x_.; \alpha)$ is continuous in $a_1, a_2, \ldots, a_m$ and it follows that M_t is not empty.) How fast should the sieve grow in order that M_t converge, in some sense, to g? (i.e. how rapidly can the sequence m_t increase to ∞?) Clearly, the growth of the sieve should be governed not directly by t, but by the amount of time that the process has spent in the interval $[\lambda_1, \lambda_2]$ up to time t. Define

$$\mathcal{J}(t) = \int_0^t I_{[\lambda_1, \lambda_2]}(x_s)\, ds.$$

The Theorem says that if m_t grows sufficiently slowly with respect to $\mathcal{J}(t)$, then

$$\sup_{\alpha \in M_t} \|\alpha - g\| \to 0 \quad \text{almost surely,}$$

where the norm is L_2 with respect to "the natural" measure for this problem:

Proposition. *Asssume* A1 *and* A2. *Then for every Borel set* A *in* $[\lambda_1, \lambda_2]$

$$\text{(2.2)} \qquad \lim_{t\to\infty} \frac{\int_0^t I_A(x_s)\,ds}{\int_0^t I_{[\lambda_1,\lambda_2]}(x_s)\,ds}$$

exists and is constant almost surely, and the set function defined by this limit is a probability measure.

The proof is deferred (follows from Lemma 1 below). I will use $\mathcal{C}(A)$ to refer to this measure.

Theorem. *Assume* A1 *and* A2, *and that* m_t *is a nondecreasing sequence of integers, diverding to* $+\infty$, *and satisfying*

$$m_t \leqslant k_3 \mathcal{J}(t)^{1-\delta}$$

for some positive constant k_3 *and* δ. *Then, as* $t \to \infty$,

$$\sup_{\alpha \in M_t} \int_{\lambda_1}^{\lambda_2} |\alpha(x) - g(x)|^2 \mathcal{C}(dx) \to 0 \quad \textit{almost surely.}$$

3. PROOF OF THE THEOREM

It will be convenient to assume that $x_0 \notin [\lambda_1, \lambda_2]$ (the modification for $x_0 \in [\lambda_1, \lambda_2]$ is trivial). Define two sequences of stopping times as follows:

$$e_1 = \inf\{t \geqslant 0: x_t \in [\lambda_1, \lambda_2]\}$$

Given $e_1, \ldots, e_k$,

$$l_k = \begin{cases} e_k + 1 & \text{if} \quad x_{e_k+1} \notin [\lambda_1, \lambda_2], \\ \inf\{t \geqslant e_k + 1: x_t \notin [\lambda_1, \lambda_2]\} & \text{if} \quad x_{e_k+1} \in [\lambda_1, \lambda_2]. \end{cases}$$

Given $l_1, \ldots, l_k$,

$$e_{k+1} = \inf\{t \geqslant l_k: x_t \in [\lambda_1, \lambda_2]\}.$$

Because of assumption A2, x_t is recurrent, and therefore these stopping times are well-defined. Notice that $\bigcup_{k=1}^{\infty} [e_k, l_k]$ includes all of the time that x_t spends on the interval $[\lambda_1, \lambda_2]$.*

Because of the strong Markov property of x_t, the discrete time process $x_{e_1}, x_{e_2}, \ldots$ is Markov with state space $\{\lambda_1, \lambda_2\}$ and stationary transition probabilities

$$p_{ij} = \mathsf{P}(x_{e_2} = \lambda_j \mid x_{e_1} = \lambda_i) \qquad (i = 1, 2,\ j = 1, 2).$$

Obviously $p_{ij} > 0$ $\forall i$ and j, and therefore there exists a unique stationary distribution $\{\pi_1, \pi_2\}$ on the states $\{\lambda_1, \lambda_2\}$. For any bounded measurable function α define

1 OK just this

$$\widetilde{\mathsf{E}}[\alpha(x)] = \sum_{i=1}^{2} \sum_{j=1}^{2} \pi_i p_{ij} \mathsf{E}\Big[\int_{e_1}^{l_1} \alpha(x_s)\, ds \mid x_{e_1} = \lambda_i,\ x_{e_2} = \lambda_j\Big].$$

A routine argument establishes the existence of a $\rho < 1$ such that

$$(3.1) \qquad \mathsf{P}((l_1 - e_1) > p \mid x_{e_1} = \lambda_i,\ x_{e_2} = \lambda_j) \leqslant \rho^p$$

for all i and j, and all $p = 0, 1, \ldots$. Hence, the "expectation", $\widetilde{\mathsf{E}}[\alpha(x)]$, is finite for bounded α.

The key to the proof of the Theorem will be the following lemma:

Lemma 1. *Fix $\epsilon > 0$. There exists a constant k (which may depend on ϵ) such that for any constant $c \geqslant 1$, and any two measurable functions α and β uniformly bounded by c,*

$$\mathsf{P}\Big\{\Big|\frac{1}{n}\sum_{k=1}^{n}\Big\{\int_{e_k}^{l_k} \alpha(x_s)\, ds + \int_{e_k}^{l_k} \beta(x_s)\, dw_s\Big\} - \widetilde{\mathsf{E}}[\alpha(x)]\Big| > \epsilon\Big\}$$

$$\leqslant k\Big(1 - \frac{1}{kc^2} + \frac{k}{n}\Big)^{\frac{n}{k}} \quad \textit{for all} \quad n \geqslant 1.$$

*After examining the proof the reader may wonder why I do not simply define e_{k+1} to be the first entrance of x_t into $[\lambda_1, \lambda_2]$ (upcrossing of λ_1 or downcrossing of λ_2) after $e_k + 1$, and then take $l_k = e_{k+1}$. The difficulty here is that, without additional assumptions, it may happen that $\mathsf{E}(l_k - e_k) = \infty$ (as when $g = 0$). In this case the proof would no longer apply.

If we assume Lemma 1 (it will be proved later) then we can establish the Proposition as follows: Take $\beta(x) = 0$ and $\alpha(x) = I_A(x)$, where A is any Borel set in $[\lambda_1, \lambda_2]$. By Lemma 1, and the Borel – Cantelli Lemma,

$$\frac{1}{n} \sum_{n=1}^{n} \int_{e_k}^{l_k} I_A(x_s)\, ds \to \tilde{\mathrm{E}}[I_A(x)] \quad \text{almost surely,}$$

as $n \to \infty$. If $n_t = \sup\{k\colon l_k \leqslant t\}$ then

$$\frac{1}{n_t} \int_0^t I_A(x_s)\, ds = \frac{1}{n_t} \sum_{k=1}^{n_t} \int_{e_k}^{l_k} I_A(x_s)\, ds + \frac{1}{n_t} \int_{l_{n_t}}^{t} I_A(x_s)\, ds.$$

Obviously,

$$\frac{1}{n_t} \int_{l_{n_t}}^{t} I_A(x_s)\, ds \to 0 \quad \text{almost surely,}$$

as $t \to \infty$. Hence, as $t \to \infty$,

$$(3.2) \qquad \frac{1}{n_t} \int_0^t I_A(x_s)\, ds \to \tilde{\mathrm{E}}[I_A(x)] \quad \text{almost surely}$$

and, therefore,

$$(3.3) \qquad \frac{\int_0^t I_A(x_s)\, ds}{\int_0^t I_{[\lambda_1, \lambda_2]}(x_s)\, ds} \to \frac{\tilde{\mathrm{E}}[I_A(x)]}{\tilde{\mathrm{E}}[I_{[\lambda_1, \lambda_2]}(x)]} \quad \text{almost surely.}$$

Observe that:

1. This proves the Proposition since the right hand side of (3.3) is a probability measure.

2. To prove the Theorem, it will be enough to show

$$\sup_{\alpha \in M_t} \tilde{\mathrm{E}}\,|I_{[\lambda_1, \lambda_2]}(x_s)(\alpha(x) - g(x))|^2 \to 0 \quad \text{almost surely}$$

as $t \to \infty$, and

3. (3.2) implies that

$$(3.4) \qquad \frac{\mathcal{I}(t)}{n_t} \to \widetilde{\mathrm{E}}[I_{[\lambda_1,\lambda_2]}(x)] \quad \text{almost surely}$$

as $t \to \infty$.

Another consequence of Lemma 1 is:

Lemma 2.

$$\overline{\lim_{t\to\infty}} \sup_{\alpha \in S_t} \Big| \frac{1}{n_t} \log f_t(x;\alpha) - \widetilde{\mathrm{E}}\big[I_{[\lambda_1,\lambda_2]}(x)\big(\alpha(x)g(x) - \tfrac{1}{2}\alpha(x)^2\big)\big]\Big| = 0 \quad \text{almost surely.}$$

Assume, for now, that this is proven too. Then, to prove the Theorem, choose for each $t \geqslant 0$ $\alpha_t \in S_t$ such that

$$\widetilde{\mathrm{E}}|I_{[\lambda_1,\lambda_2]}(x)(\alpha_t(x) - g(x))|^2 \to 0$$

as $t \to \infty$.

(The assumptions about $\{\psi_j(x)\}_{j=1}^{\infty}$ guarantee that this can be done.) And now reason that

$$\underline{\lim_{t\to\infty}} \inf_{\alpha\in M_t} \widetilde{\mathrm{E}}\big[I_{[\lambda_1,\lambda_2]}(x)\big(\alpha(x)g(x) - \tfrac{1}{2}\alpha(x)^2\big)\big]$$

$$= \underline{\lim_{t\to\infty}} \inf_{\alpha\in M_t} \frac{1}{n_t} \log f_t(x;\alpha) \quad \text{almost surely}$$

$$\geqslant \underline{\lim_{t\to\infty}} \frac{1}{n_t} \log f_t(x;\alpha_t) \quad \text{(from the definition of } M_t)$$

$$= \underline{\lim_{t\to\infty}} \widetilde{\mathrm{E}}\big[I_{[\lambda_1,\lambda_2]}(x)\big(\alpha_t(x)g(x) - \tfrac{1}{2}\alpha_t(x)^2\big)\big]$$

$$= \frac{1}{2}\widetilde{\mathrm{E}}[I_{[\lambda_1,\lambda_2]}(x)g(x)^2]$$

$$\text{(since } \widetilde{\mathrm{E}}|I_{[\lambda_1,\lambda_2]}(x)(\alpha_t(x) - g(x))|^2 \to 0).$$

Hence

$$\varliminf_{t\to\infty} \inf_{\alpha\in M_t} \widetilde{\mathsf{E}}\Big[I_{[\lambda_1,\lambda_2]}(x) \times \Big(\alpha(x)g(x) - \frac{1}{2}\,\alpha(x)^2 - \frac{1}{2}\,g(x)^2\Big)\Big] = 0 \quad \text{almost surely,}$$

and then, finally,

$$\overline{\lim_{t\to\infty}} \sup_{\alpha\in M_t} \widetilde{\mathsf{E}}\,|I_{[\lambda_1,\lambda_2]}(x)(\alpha(x) - g(x))|^2 = 0 \quad \text{almost surely.}$$

It remains to prove Lemmas 1 and 2.

Proof of Lemma 1. $c_1, c_2,$ etc. will refer to constants which may depend on ϵ but are independent of α, β, c or n.

$$\mathsf{P}\Big(\Big|\frac{1}{n}\sum_{k=1}^{n}\Big\{\int_{e_k}^{l_k}\alpha(x_s)\,ds + \int_{e_k}^{l_k}\beta(x_s)\,dw_s\Big\} - \widetilde{\mathsf{E}}[\alpha(x)]\Big| > \epsilon\Big)$$

$$\leqslant \mathsf{P}\Big(\Big|\frac{1}{n}\sum_{k=1}^{n}\int_{e_k}^{l_k}\alpha(x_s)\,ds - \widetilde{\mathsf{E}}[\alpha(x)]\Big| > \frac{\epsilon}{2}\Big)$$

$$+ \mathsf{P}\Big(\Big|\frac{1}{n}\sum_{k=1}^{n}\int_{e_k}^{l_k}\beta(x_s)\,dw_s\Big| > \frac{\epsilon}{2}\Big)$$

$$\leqslant \mathsf{P}\Big(\Big|\frac{1}{n}\sum_{k=1}^{n}\Big(\int_{e_k}^{l_k}\alpha(x_s)\,ds - \mathsf{E}\Big[\int_{e_k}^{l_k}\alpha(x_s)\,ds \,|\, x_{e_k}, x_{e_{k+1}}\Big]\Big)\Big| > \frac{\epsilon}{4}\Big) \tag{3.5}$$

$$+ \mathsf{P}\Big(\Big|\frac{1}{n}\sum_{k=1}^{n}\mathsf{E}\Big[\int_{e_k}^{l_k}\alpha(x_s)\,ds \,|\, x_{e_k}, x_{e_{k+1}}\Big] - \widetilde{\mathsf{E}}[\alpha(x)]\Big| > \frac{\epsilon}{4}\Big) \tag{3.6}$$

$$+ \mathsf{P}\Big(\Big|\frac{1}{n}\sum_{k=1}^{n}\Big(\int_{e_k}^{l_k}\beta(x_s)\,dw_s - \mathsf{E}\Big[\int_{e_k}^{l_k}\beta(x_s)\,dw_s \,|\, x_{e_k}, x_{e_{k+1}}\Big]\Big)\Big| > \frac{\epsilon}{4}\Big) \tag{3.7}$$

$$+ \mathsf{P}\Big(\Big|\frac{1}{n}\sum_{k=1}^{n}\mathsf{E}\Big[\int_{e_k}^{l_k}\beta(x_s)\,dw_s \,|\, x_{e_k}, x_{e_{k+1}}\Big]\Big| > \frac{\epsilon}{4}\Big). \tag{3.8}$$

I will develop bounds for (3.7) and (3.8). The treatment of (3.5) and (3.6), being analogous and somewhat simpler, is omitted.

Begin with two preliminary bounds:

1. There exists a constant c_1 such that

$$\text{(3.9)} \qquad \mathsf{E}\Big[\Big|\int_{e_1}^{l_1} \beta(x_s)\, dw_s\Big|^l \mid x_{e_1} = \lambda_i,\ x_{e_2} = \lambda_j\Big] \leqslant c_1 c^l$$

for all $i = 1, 2$, $j = 1, 2$ and $l = 1, 2, 3, 4$.

2. There exists a constant c_2 such that

$$\text{(3.10)} \qquad \mathsf{E}\Big[\exp\Big\{t\Big|\int_{e_1}^{l_1} \beta(x_s)\, dw_s\Big|\Big\} \mid x_{e_1} = \lambda_i,\ x_{e_2} = \lambda_j\Big] \leqslant c_2$$

for all $t \leqslant \dfrac{1}{c_2 c}$, $i = 1, 2$ and $j = 1, 2$.

Observe that for any random variable z such that $\mathsf{E}|z| < \infty$, and any $i = 1, 2$,

$$\mathsf{E}[|z| \mid x_{e_1} = \lambda_i] = \mathsf{E}[|z| \mid x_{e_1} = \lambda_i,\ x_{e_2} = \lambda_1]p_{i1}$$
$$+\ \mathsf{E}[|z| \mid x_{e_1} = \lambda_i,\ x_{e_2} = \lambda_2]p_{i2}.$$

Recall that $p_{ij} > 0$ for all i and j. Hence with $c_3 = \max\limits_{i,j} \dfrac{1}{p_{ij}}$,

$$\mathsf{E}[|z| \mid x_{e_1} = \lambda_i,\ x_{e_2} = \lambda_j] = c_3 \mathsf{E}[|z| \mid x_{e_1} = \lambda_i]$$

for all i and j. Use this to establish (3.10):

$$\text{(3.11)} \qquad \mathsf{E}\Big[\exp\Big\{t\Big|\int_{e_1}^{l_1} \beta(x_s)\, dw_s\Big|\Big\} \mid x_{e_1} = \lambda_i,\ x_{e_2} = \lambda_j\Big]$$
$$\leqslant c_3 \mathsf{E}\Big[\exp\Big\{t\Big|\int_{e_1}^{l_1} \beta(x_s)\, dw_s\Big|\Big\} \mid x_{e_1} = \lambda_i\Big]$$
$$= c_3 \sum_{k=1}^{\infty} \mathsf{E}\big[I_{[p-1,p)}(l_1 - e_1)$$

$$\times \exp\left\{t\left|\int_{e_1}^{p} I_{[e_1,l_1]}(s)\beta(x_s)\,dw_s\right|\right\} \mid x_{e_1} = \lambda_i\Big]$$

$$\leqslant c_3 \sum_{p=1}^{\infty} \Big\{\mathsf{P}((l_1 - e_1) \in [p-1, p) \mid x_{e_1} = \lambda_i)^{\frac{1}{2}}$$

$$\times\ \mathsf{E}\left[\exp\left\{2t\left|\int_{e_1}^{p} I_{[e_1,l_1]}(s)\beta(x_s)\,dw_s\right|\right\} \mid x_{e_1} = \lambda_i\right]^{\frac{1}{2}}\Big\}.$$

From (3.1)

$$\text{(3.12)} \qquad \mathsf{P}((l_1 - e_1) \in [p-1, p) \mid x_{e_1} = \lambda_i) \leqslant \rho^{p-1}$$

where $\rho < 1$, and, from Lemma B, Section 5 of Geman and Hwang [6]

$$\text{(3.13)} \qquad \mathsf{E}\left[\exp\left\{2t\left|\int_{e_1}^{p} I_{[e_1,l_1]}(s)\beta(x_s)\,dw_s\right|\right\} \mid x_{e_1} = \lambda_i\right] \leqslant \sqrt{2} + e^{(4tc)^2 p}.$$

Use (3.12) and (3.13) in (3.11):

$$\text{(3.14)} \qquad \begin{aligned} &\mathsf{E}\left[\exp\left\{t\left|\int_{e_1}^{l_1} \beta(x_s)\,dw_s\right|\right\} \mid x_{e_1} = \lambda_i,\ x_{e_2} = \lambda_j\right] \\ &\leqslant c_3 \sum_{p=1}^{\infty} \rho^{\frac{p-1}{2}} (\sqrt{2} + e^{(4tc)^2 p})^{\frac{1}{2}}. \end{aligned}$$

Provided that c_2 is sufficiently large, the expression in (3.14) is bounded by c_2 whenever $tc \leqslant \frac{1}{c_2}$. This establishes (3.10). A similar argument can be used for (3.9).

Return now to (3.7). With

$$z_k = \int_{e_k}^{l_k} \beta(x_s)\,dw_s - \mathsf{E}\left[\int_{e_k}^{l_k} \beta(x_s)\,dw_s \mid x_{e_k}, x_{e_{k+1}}\right]$$

(3.7) becomes

$$\mathsf{P}\left(\left|\frac{1}{n}\sum_{k=1}^{n} z_k\right| > \frac{\epsilon}{4}\right),$$

and this is no bigger than

$$(3.15)\qquad \mathsf{P}\Big(\frac{1}{n}\sum_{k=1}^{n} z_k > \frac{\epsilon}{4}\Big) + \mathsf{P}\Big(\frac{1}{n}\sum_{k=1}^{n} z_k < -\frac{\epsilon}{4}\Big).$$

$$\mathsf{P}\Big(\frac{1}{n}\sum_{k=1}^{n} z_k > \frac{\epsilon}{4}\Big) = (\text{for any } t > 0)\ \mathsf{P}\Big(t\sum_{k=1}^{n}\Big(z_k - \frac{\epsilon}{4}\Big) > 0\Big)$$

$$\leqslant \mathsf{E}\prod_{k=1}^{n} e^{t(z_k - \frac{\epsilon}{4})}$$

$$= \mathsf{E}\Big\{\mathsf{E}\Big[\prod_{k=1}^{n} e^{t(z_k - \frac{\epsilon}{4})} \,|\, x_{e_1}, \ldots, x_{e_{n+1}}\Big]\Big\}$$

$= $ (strong Markov property)

$$(3.16)\qquad \mathsf{E}\Big\{\prod_{k=1}^{n} \mathsf{E}[e^{t(z_k - \frac{\epsilon}{4})} \,|\, x_{e_k}, x_{e_{k+1}}]\Big\}.$$

Fix k. For any $i = 1, 2$ and $j = 1, 2$ define

$$\varphi_k(t) = \mathsf{E}[e^{t(z_k - \frac{\epsilon}{4})} \,|\, x_{e_k} = \lambda_i,\ x_{e_{k+1}} = \lambda_j].$$

Then $\varphi_k(0) = 1$, $\frac{d}{dt}\varphi_k(0) = -\frac{\epsilon}{4}$, and

$$\frac{d^2}{dt^2}\varphi_k(t) = \mathsf{E}\Big[\Big(z_k - \frac{\epsilon}{4}\Big)^2 e^{t(z_k - \frac{\epsilon}{4})} \,|\, x_{e_k} = \lambda_i,\ x_{e_{k+1}} = \lambda_j\Big]$$

$$\leqslant \mathsf{E}\Big[\Big(z_k - \frac{\epsilon}{4}\Big)^4 \,|\, x_{e_k} = \lambda_i,\ x_{e_{k+1}} = \lambda_j\Big]^{\frac{1}{2}}$$

$$\times\ \mathsf{E}[e^{2t(z_k - \frac{\epsilon}{4})} \,|\, x_{e_k} = \lambda_i,\ x_{e_{k+1}} = \lambda_j]^{\frac{1}{2}}$$

$\leqslant$ (use 3.9 and 3.10) $c_4 c^2$

provided $t \leqslant \frac{1}{c_4 c}$, where c_4 is some sufficiently large constant. Integrate $\frac{d^2}{dt^2}\varphi_k(t)$ and then $\frac{d}{dt}\varphi_k(t)$:

$$\frac{d}{dt}\varphi_k(t) \leqslant -\frac{\epsilon}{4} + c_4 c^2 t \Rightarrow \varphi_k(t) \leqslant 1 - \frac{\epsilon}{4} t + \frac{1}{2} c_4 c^2 t^2$$

provided $t \leqslant \frac{1}{c_4 c}$. Set $t = \frac{1}{c_5 c^2}$ where $c_5 \geqslant c_4$ (recall that $c \geqslant 1$):

$$\mathsf{E}[e^{t(z_k - \frac{\epsilon}{4})} \mid x_{e_k} = \lambda_i,\ x_{e_{k+1}} = \lambda_j]$$

$$= \varphi_k(t) \leqslant 1 - \frac{1}{c_5 c^2}\left[\frac{\epsilon}{4} - \frac{c_4}{2c_5}\right] \leqslant 1 - \frac{\epsilon}{8 c_5 c^2}$$

for c_5 sufficiently large. Finally, put this back into (3.16):

$$\mathsf{P}\left(\frac{1}{n}\sum_{k=1}^{n} z_k > \frac{\epsilon}{4}\right) \leqslant \left(1 - \frac{\epsilon}{8 c_5 c^2}\right)^n.$$

Since the same argument applies to the second term in (3.15), this establishes the required bound on (3.7).

For (3.8), define for each i and j a random variable n_{ij} by

$$n_{ij} = \text{number of times } (x_{e_k}, x_{e_{k+1}}) = (\lambda_i, \lambda_j)$$

$$(k = 1, 2, \ldots, n),$$

and write

$$\frac{1}{n}\sum_{k=1}^{n} \mathsf{E}\left[\int_{e_k}^{l_k} \beta(x_s)\, dw_s \mid x_{e_k}, x_{e_{k+1}}\right]$$

$$= \sum_{i=1}^{2}\sum_{j=1}^{2} \frac{n_{ij}}{n}\, \mathsf{E}\left[\int_{e_1}^{l_1} \beta(x_s)\, dw_s \mid x_{e_1} = \lambda_i,\ x_{e_2} = \lambda_j\right].$$

Observe that for $i = 1$ or 2

$$\mathsf{E}\left[\int_{e_1}^{l_1} \beta(x_s)\, dw_s \mid x_{e_1} = \lambda_i,\ x_{e_2} = \lambda_1\right] p_{i1}$$

$$+\ \mathsf{E}\left[\int_{e_1}^{l_1} \beta(x_s)\, dw_s \mid x_{e_1} = \lambda_i,\ x_{e_2} = \lambda_2\right] p_{i2}$$

$$= \mathrm{E}\Big[\int_{e_1}^{l_1} \beta(x_s)\,dw_s \mid x_{e_1} = \lambda_i\Big] = 0$$

$$\Rightarrow \frac{1}{n}\sum_{k=1}^{n} \mathrm{E}\Big[\int_{e_k}^{l_k} \beta(x_s)\,dw_s \mid x_{e_k}, x_{e_{k+1}}\Big]$$

$$= \sum_{i=1}^{2}\sum_{j=1}^{2} \Big(\frac{n_{ij}}{n} - \pi_i p_{ij}\Big)\, \mathrm{E}\Big[\int_{e_1}^{l_1} \beta(x_s)\,dw_s \mid x_{e_1} = \lambda_i,\ x_{e_2} = \lambda_j\Big].$$

Using this, and (3.9),

$$\mathrm{P}\Big(\Big|\frac{1}{n}\sum_{k=1}^{n} \mathrm{E}\Big[\int_{e_k}^{l_k} \beta(x_s)\,dw_s \mid x_{e_k}, x_{e_{k+1}}\Big]\Big| > \frac{\epsilon}{4}\Big)$$

$$(3.17) \qquad \leqslant \mathrm{P}\Big(\sum_{i=1}^{2}\sum_{j=1}^{2}\Big|\frac{n_{ij}}{n} - \pi_i p_{ij}\Big| > \frac{\epsilon}{4c_1 c}\Big)$$

$$\leqslant \sum_{i=1}^{2}\sum_{j=1}^{2} \mathrm{P}\Big(\Big|\frac{n_{ij}}{n} - \pi_i p_{ij}\Big| > \frac{\epsilon}{16 c_1 c}\Big).$$

Consider, for example,

$$\mathrm{P}\Big(\Big|\frac{n_{11}}{n} - \pi_1 p_{11}\Big| > \frac{\epsilon}{16 c_1 c}\Big).$$

Define for each i

$$n_i = \text{number of times } x_{e_k} = \lambda_i \qquad (k = 1, 2, \ldots, n).$$

Omitting details, a combinatorial-type argument demonstrates the existence of a constant c_6 sufficiently large that for all $\delta < 1$ and $n \geqslant 1$

$$(3.18) \qquad \Big\{\Big|\frac{n_{11}}{n} - \pi_1 p_{11}\Big| > \delta\Big\} \Rightarrow \bigcup_{i=1}^{2}\Big\{n_i > \frac{n}{c_6} \ \text{ and } \ \Big|\frac{n_{ii}}{n_i} - p_{ii}\Big| > \frac{\delta}{c_6} - \frac{c_6}{n}\Big\}.$$

Take c_7 so large that $\frac{\epsilon}{16 c_1} > \frac{1}{c_7}$. Then, using (3.18),

$$\mathrm{P}\left(\left|\frac{n_{11}}{n}-\pi_1 p_{11}\right|>\frac{\epsilon}{16c_1 c}\right)\leqslant \mathrm{P}\left(\left|\frac{n_{11}}{n}-\pi_1 p_{11}\right|>\frac{1}{c_7 c}\right)$$

$$\leqslant \mathrm{P}\left(n_1>\frac{n}{c_6}\ \text{ and }\ \left|\frac{n_{11}}{n_1}-p_{11}\right|>\frac{1}{c_6 c_7 c}-\frac{c_6}{n}\right)$$

$$+\ \mathrm{P}\left(n_2>\frac{n}{c_6}\ \text{ and }\ \left|\frac{n_{22}}{n_2}-p_{22}\right|>\frac{1}{c_6 c_7 c}-\frac{c_6}{n}\right)$$

$$\leqslant \mathrm{P}\left(\left|\frac{b(r,p_{11})}{r}-p_{11}\right|>\frac{1}{c_6 c_7 c}-\frac{c_6}{n}\right)$$

$$+\ \mathrm{P}\left(\left|\frac{b(r,p_{22})}{r}-p_{22}\right|>\frac{1}{c_6 c_7 c}-\frac{c_6}{n}\right)$$

where r is the smallest integer greater than or equal to $\frac{n}{c_6}$ and $b(r, p_{ii})$ is a binomial random variable with r "trials" and a p_{ii} "probability of success" in each trial. Each of the probabilities in (3.19) is easily bounded by

$$\left(1-\frac{1}{c_8 c^2}+\frac{c_8}{n}\right)^{\frac{n}{c_6}}$$

for c_8 sufficiently large.

The proof of Lemma 1 can now be completed by the analogous treatment of the remaining terms in (3.17).

Proof of Lemma 2. Because of (3.4), and the assumed relation between m_t and $\mathcal{I}(t)$, there is a constant c_9 such that

$$\overline{\lim_{t\to\infty}}\ \frac{m_t}{[c_9 n^{1-\delta}]}<1 \quad \text{almost surely,}$$

where $[x]$ denotes the greatest integer less than or equal to x. Define $m_n = [c_9 n^{1-\delta}]$ and

$$\tilde{S}_m=\left\{\sum_{j=1}^{m} a_j\psi_j(x):\ \sum_{j=1}^{m}|a_j|\leqslant k_1(\log m)^{k_2}\right\}.$$

Then

$$\overline{\lim_{t\to\infty}}\ \sup_{\alpha\in S_t} \Big| \frac{1}{n_t} \log f_t(x\,;\alpha) - \widetilde{\mathrm{E}}\big[I_{[\lambda_1,\lambda_2]}(x)\big(\alpha(x)g(x) - \tfrac{1}{2}\,\alpha(x)^2\big)\big]\Big|$$

$$= \overline{\lim_{t\to\infty}}\ \sup_{\alpha\in S_t} \Big| \frac{1}{n_t} \log f_{l_{n_t}}(x\,;\alpha) - \widetilde{\mathrm{E}}\big[I_{[\lambda_1,\lambda_2]}(x)\big(\alpha(x)g(x) - \tfrac{1}{2}\,\alpha(x)^2\big)\big]\Big| \quad \text{almost surely}$$

(3.20)

$$\leqslant \overline{\lim_{n\to\infty}}\ \sup_{\alpha\in S_{m_n}} \Big| \frac{1}{n} \log f_{l_n}(x\,;\alpha) - \widetilde{\mathrm{E}}\big[I_{[\lambda_1,\lambda_2]}(x)\big(\alpha(x)g(x) - \tfrac{1}{2}\,\alpha(x)^2\big)\big]\Big| \quad \text{almost surely.}$$

For each n, let $\beta_{n1}, \beta_{n2}, \ldots, \beta_{nd_n}$ denote the collection of functions $\beta \in \widetilde{S}_{m_n}$ of the form

$$\beta(x) = \sum_{j=1}^{m_n}{}' \, b_j\, \psi_j(x)$$

where, for each j,

$$b_j = \frac{q_j}{m_n\,(\log m_n)^{k_2+1}}$$

for some integer q_j (positive, negative, or zero). Because

$$\sum_{k=1}^{m_n}{}' \, |b_j| \leqslant k_1\,(\log m_n)^{k_2}$$

for any $\beta \in \widetilde{S}_{m_n}$, $d_n = O(m_n^{2m_n})$. Associated with each β_{ni} (say

$$\beta_{ni} = \sum_{j=1}^{m_n}{}' \, a_j\, \psi_j(x)),$$

define a set of functions B_{ni}

$$B_{ni} = \Big\{\alpha(x) = \sum_{j=1}^{m_n} a_j \psi_j(x) \colon \alpha \in \widetilde{S}_{m_n} \quad \text{and}$$

$$\sup_{1 \leq j \leq m_n} |a_j - b_j| \leq \frac{1}{m_n (\log m_n)^{k_2+1}}\Big\}.$$

Then $\bigcup_{i=1}^{d_n} B_{ni}$ includes all of $\widetilde{S}_{m_n}$, and since $|\psi_j(x)| \leq 1$ on $[\lambda_1, \lambda_2]$,

$$\sup_{x \in [\lambda_1, \lambda_2]} |\alpha(x) - \beta_{ni}(x)| \leq \frac{1}{(\log m_n)^{k_2+1}}$$

for all $\alpha \in B_{ni}$.

We can now rewrite (3.20):

$$\varlimsup_{n\to\infty} \sup_{i=1,\ldots,d_n} \sup_{\alpha \in B_{ni}} \Big| \frac{1}{n} \log f_{l_n}(x.; \alpha)$$

$$- \mathrm{E}\Big[I_{[\lambda_1,\lambda_2]}(x)\Big(\alpha(x)g(x) - \frac{1}{2}\alpha(x)^2\Big)\Big]\Big|$$

which is bounded by

$$(3.21) \qquad \varlimsup_{n\to\infty} \sup_{i=1,\ldots,d_n} \Big| \frac{1}{n} \log f_{l_n}(x.; \beta_{ni}) - \widetilde{\mathrm{E}}\Big[I_{[\lambda_1,\lambda_2]}(x)\Big(\beta_{ni}(x)g(x) - \frac{1}{2}\beta_{ni}(x)^2\Big)\Big]\Big|$$

$$(3.22) \qquad + \varlimsup_{n\to\infty} \sup_{i=1,\ldots,d_n} \sup_{\alpha \in B_{ni}} \Big| \frac{1}{n} \log f_{l_n}(x.; \alpha) - \frac{1}{n} \log f_{l_n}(x.; \beta_{ni})\Big|$$

$$(3.23) \qquad + \varlimsup_{n\to\infty} \sup_{i=1,\ldots,d_n} \sup_{\alpha \in B_{ni}} \Big| \widetilde{\mathrm{E}}\Big[I_{[\lambda_1,\lambda_2]}(x) \times \Big(\alpha(x)g(x) - \frac{1}{2}\alpha(x)^2\Big)\Big] - \widetilde{\mathrm{E}}\Big[I_{[\lambda_1,\lambda_2]}(x)\Big(\beta_{ni}(x)g(x) - \frac{1}{2}\beta_{ni}(x)^2\Big)\Big]\Big|$$

The expression in (3.23) equals zero, since

$$\sup_{\alpha \in B_{ni}} \sup_{x \in [\lambda_1, \lambda_2]} |\alpha(x) - \beta_{ni}(x)| \leqslant \frac{1}{(\log m_n)^{k_2+1}} \to 0$$

as $n \to \infty$.

Next, argue that the expression in (3.21) also equals zero (almost surely): fix an arbitrary $\epsilon > 0$. Notice that $\alpha \in S_{m_n} \Rightarrow \sup_{x \in [\lambda_1, \lambda_2]} |\alpha(x)| \leqslant k_1 (\log m_n)^{k_2}$. Apply Lemma 1 with

$$\alpha(x) = I_{[\lambda_1, \lambda_2]}(x)\big(\beta_{ni}(x) g(x) - \tfrac{1}{2}\, \beta_{ni}(x)^2\big)$$

$$\beta(x) = \sigma I_{[\lambda_1, \lambda_2]}(x) \beta_{ni}(x) \quad \text{and} \quad c = c_n = O((\log m_n)^{2k_2}):$$

$$\mathrm{P}\Big(\sup_{i=1,\dots,d_n} \Big| \frac{1}{n} \log f_{l_n}(x.; \beta_{ni}) - \tilde{\mathrm{E}}\big[I_{[\lambda_1, \lambda_2]}\big(\beta_{ni}(x) g(x) - \tfrac{1}{2}\, \beta_{ni}(x)^2\big)\big] \Big| > \epsilon \Big)$$

$$\leqslant \sum_{i=1}^{d_n} \mathrm{P}\Big(\Big| \frac{1}{n} \log f_{l_n}(x.; \beta_{ni}) - \tilde{\mathrm{E}}\big[I_{[\lambda_1, \lambda_2]}\big(\beta_{ni}(x) g(x) - \tfrac{1}{2}\, \beta_{ni}(x)^2\big)\big] \Big| > \epsilon \Big)$$

$$\leqslant d_n k \Big(1 - \frac{1}{k c_n^2} + \frac{k}{n}\Big)^{\frac{n}{k}}$$

$$= O\Big(n^{2c_9 n^{1-\delta}} \Big(1 - \frac{1}{k (\log n)^{4k_2}} + \frac{k}{n}\Big)^{\frac{n}{k}}\Big).$$

Since

$$\sum_{n=1}^{\infty} n^{2c_9 n^{1-\delta}} \Big(1 - \frac{1}{k (\log n)^{4k_2}} + \frac{k}{n}\Big)^{\frac{n}{k}} < \infty,$$

the Borel – Cantelli Lemma implies that the expression in (3.21) is equal to zero with probability one.

All that remains is the expression in (3.22).

$$\sup_{i=1,\ldots,d_n} \sup_{\alpha \in B_{ni}} \left| \frac{1}{n} \log f_{l_n}(x_.;\alpha) - \frac{1}{n} \log f_{l_n}(x_.;\beta_{ni}) \right|$$

$$\leqslant \sup_{i=1,\ldots,d_n} \sup_{\alpha \in B_{ni}} \left| \frac{1}{n} \sum_{k=1}^{n} \int_{e_k}^{l_k} I_{[\lambda_1,\lambda_2]}(x_s) \right.$$

(3.24) $$\times \left[g(x_s)(\alpha(x_s) - \beta_{ni}(x_s)) \right.$$

$$\left. \left. - \frac{1}{2}(\alpha(x_s) - \beta_{ni}(x_s))(\alpha(x_s) + \beta_{ni}(x_s)) \right] ds \right|$$

(3.25) $$+ \sup_{i=1,\ldots,d_n} \sup_{\alpha \in B_{ni}} \left| \frac{\sigma}{n} \sum_{k=1}^{n} \int_{e_k}^{l_k} I_{[\lambda_1,\lambda_2]}(x_s) \right.$$

$$\left. \times (\alpha(x_s) - \beta_{ni}(x_s))\, dw_s \right|.$$

If c_{10} is chosen large enough, then the expression in (3.24) is bounded by

(3.26) $$\frac{c_{10}}{\log m_n} \frac{1}{n} \sum_{k=1}^{n} (l_k - e_k).$$

Lemma 1 (with $\alpha(x) = 1$ and $\beta(x) = 0$) implies that

$$\frac{1}{n} \sum_{k=1}^{n} (l_k - e_k) \to \tilde{\mathrm{E}}[l] \quad \text{almost surely}$$

so ((3.26) (and therefore (3.24)) converges to zero almost surely as $n \to \infty$. In (3.25), write

$$\alpha(x) = \sum_{j=1}^{m_n} a_j \psi_j(x) \quad \text{and} \quad \beta_{ni}(x) = \sum_{j=1}^{m_n} b_j \psi_j(x):$$

$$\left| \frac{\sigma}{n} \sum_{k=1}^{n} \int_{e_k}^{l_k} I_{[\lambda_1,\lambda_2]}(x_s)(\alpha(x_s) - \beta_{ni}(x_s))\, dw_s \right|$$

$$= \left| \frac{\sigma}{n} \sum_{k=1}^{n} \sum_{j=1}^{m_n} (a_j - b_j) \int_{e_k}^{l_k} I_{[\lambda_1,\lambda_2]}(x_s) \psi_j(x_s)\, dw_s \right|$$

$\leqslant$ (since $\alpha \in B_{ni}$)

$$\frac{\sigma}{(\log m_n)^{k_2+1}} \frac{1}{n} \sum_{k=1}^{n} \frac{1}{m_n} \sum_{j=1}^{m_n} \left| \int_{e_k}^{l_k} I_{[\lambda_1,\lambda_2]}(x_s) \psi_j(x_s)\, dw_s \right|.$$

Since this last expression does not involve i or α, it is a bound on (3.25). It will therefore be sufficient to show that

$$\frac{1}{n}\sum_{k=1}^{n}\frac{1}{m_n}\sum_{j=1}^{m_n}\left|\int_{e_k}^{l_k} I_{[\lambda_1,\lambda_2]}(x_s)\,\psi_j(x_s)\,dw_s\right|$$

is with probability one bounded for all n.

By using the fact that $|\psi_j(x)| \leqslant 1$ on $[\lambda_1, \lambda_2]$, and imitating the argument used in the proof of Lemma 1 for bounding (3.7), we can show that for fixed $\epsilon > 0$ there exists a constant $c_{11} > 0$ such that

$$\begin{aligned}
&\mathsf{P}\Big(\Big|\frac{1}{n}\sum_{k=1}^{n}\Big\{\frac{1}{m_n}\sum_{j=1}^{m_n}\Big|\int_{e_k}^{l_k} I_{[\lambda_1,\lambda_2]}(x_s)\,\psi_j(x_s)\,dw_s\Big| \\
(3.27)\qquad &-\frac{1}{m_n}\sum_{j=1}^{m_n}\mathsf{E}\Big[\Big|\int_{e_k}^{l_k} I_{[\lambda_1,\lambda_2]}(x_s)\,\psi_j(x_s)\,dw_s\Big|\,\Big|\,x_{e_k}, x_{e_{k+1}}\Big]\Big\}\Big| > \epsilon\Big) \\
&\leqslant 2\Big(1-\frac{1}{c_{11}}\Big)^n
\end{aligned}$$

for all $n \geqslant 1$. (The only modification to the previous argument is in developing an analogue to (3.10). For this purpose, first observe that

$$\begin{aligned}
&\mathsf{E}\Big[\exp\Big\{t\,\frac{1}{m_n}\sum_{j=1}^{m_n}\Big|\int_{e_k}^{l_k} I_{[\lambda_1,\lambda_2]}(x_s)\,\psi_j(x_s)\,dw_s\Big|\Big\}\,\Big|\,x_{e_k}, x_{e_{k+1}}\Big] \\
(3.28)\qquad &\leqslant \prod_{j=1}^{m_n}\mathsf{E}\Big[\exp\Big\{t\int_{e_k}^{l_k} I_{[\lambda_1,\lambda_2]}(x_s)\,\psi_j(x_s)\,dw_s\Big|\Big\}\,\Big|\,x_{e_k}, x_{e_{k+1}}\Big]^{\frac{1}{m_n}},
\end{aligned}$$

and the apply (3.10) to conclude that the expression in (3.28) is no bigger than c_2 provided $t \leqslant \frac{1}{c_2}$.)

One consequence of (3.27) is that

$$\frac{1}{n}\sum_{k=1}^{n}\frac{1}{m_n}\sum_{j=1}^{m_n}\left|\int_{e_k}^{l_k} I_{[\lambda_1,\lambda_2]}(x_s)\,\psi_j(x_s)\,dw_s\right|$$

$$-\frac{1}{n}\sum_{k=1}^{n}\frac{1}{m_n}\sum_{j=1}^{m_n} \mathrm{E}\Big[\Big|\int_{e_k}^{l_k} I_{[\lambda_1,\lambda_2]}(x_s) \times \psi_j(x_s)\,dw_s\Big|\,\Big|\, x_{e_k}, x_{e_{k+1}}\Big] \to 0 \quad \text{almost surely.}$$

Since, by (3.9)

$$\frac{1}{n}\sum_{k=1}^{n}\frac{1}{m_n}\sum_{j=1}^{m_n} \mathrm{E}\Big[\Big|\int_{c_k}^{l_k} I_{[\lambda_1,\lambda_2]}(x_s) \times \psi_j(x_s)\,dw_s\Big|\,\Big|\, x_{e_k}, x_{e_{k+1}}\Big] \leqslant c_1;$$

this completes the proof of Lemma 2.3

REFERENCES

[1] G. Banon, Nonparametric identification for diffusion processes, *SIAM J. on Control and Optimization,* 16 (1978), 380-395.

[2] G. Banon – H.T. Nguyen, Recursive estimation in diffusion model, Preliminary report.

[3] B.M. Brown – J.I. Hewitt, Asymptotic likelihood theory for diffusion processes, *J. Appl. Prob.*, 12 (1975), 228-238.

[4] P.D. Feigin, Maximum likelihood estimation for continuous time stochastic processes, *Adv. Appl. Prob.*, 8 (1976), 712-736.

[5] A. Friedman, *Stochastic differential equations and applications,* Vol. 1, Academic Press, New York, 1975.

[6] S. Geman – C.-R. Hwang, Nonparametric maximum likelihood estimation by the method of sieves, ~~Reports in Pattern Analysis, No. 80.~~ Annals of Statistics (to appear)

[7] U. Grenander, Abstract inference, in preparation.

[8] T.S. Lee – F. Kozin, Almost sure asymptotic likelihood theory for diffusion processes, *J. Appl. Prob.*, 14 (1977), 527-537.

[9] R.S. Lipster – A.N. Shiryayev, *Statistics of random processes* II, (Chapter 17), Springer-Verlag, New York, 1978.

[10] H.T. Nguyen – T.D. Pham, On the law of large numbers for continuous time martingales and applications to statistics, Preliminary report.

S. Geman

Division of Applied Mathematics, Brown University, Providence, Rhode Island 02912, USA.