(2)

CENTER FOR STOCHASTIC PROCESSES

Department of Statistics
University of North Carolina
Chapel Hill, North Carolina

ESTIMATION OF HILBERT SPACE

VALUED PARAMETERS BY THE METHOD OF SIEVES

by

G. Kallianpur and R.S. Selukar

Technical Report No. 278

October 1989

198. T. Hsing. On the extreme order statistics for a stationary sequence. July 87. *Stochastic Proc. Appl.* 29, 1988, 155-169.

199. T. Hsing. Characterization of certain point processes. Aug. 87. *Stochastic Proc. Appl.* 26, 1987, 297-316.

200. J.P. Nolan. Continuity of symmetric stable processes. Aug. 87. *J. Multivariate Anal.*, 29, 1989, 84-93.

201. M. Marques and S. Cambanis. Admissible and singular translates of stable processes. Aug. 88. *Probability Theory on Vector Spaces IV*, S. Cambanis and A. Weron, eds., *Lecture Notes in Mathematics* No. 1391, Springer, 1989, 239-257.

202. O. Kallenberg. One-dimensional uniqueness and convergence criteria for exchangeable processes. Aug. 87. *Stochastic Proc. Appl.* 28, 1988, 159-183.

203. R.J. Adler, S. Cambanis and G. Samorodnitsky. On stable Markov processes. Sept. 87. *Stochastic Proc. Appl.*, 1989, to appear.

204. G. Kallianpur and V. Perez-Abreu. Stochastic evolution equations driven by nuclear space valued martingales. Sept. 87. *Appl. Math. Optimization*, 17, 1988, 237-272.

205. R.L. Smith. Approximations in extreme value theory. Sept. 87.

206. E. Willekens. Estimation of convolution tails. Sept. 87.

207. J. Rosinski. On path properties of certain infinitely divisible processes. Sept. 87. *Stochastic Proc. Appl.*, 33, 1989, to appear.

208. A.H. Korezlioglu. Computation of filters by sampling and quantization. Sept. 87.

209. J. Bather. Stopping rules and observed significance levels. Sept. 87. *Sequential Anal.*, 8, 1989, 101-114.

210. S.T. Rachev and J.E. Yukich. Convolution metrics and rates of convergence in the central limit theorem. Sept. 87. *Ann. Probability*, 17, 1989, 775-788.

211. M. Fujisaki. Normed Bellman equation with degenerate diffusion coefficients and its applications to differential equations. Oct. 87.

212. G. Simons, Y.C. Yao and X. Wu. Sequential tests for the drift of a Wiener process with a smooth prior, and the heat equation. Oct. 87. *Ann. Statist.*, 17, 1989, 783-792.

213. R.L. Smith. Extreme value theory for dependent sequences via the Stein-Chen method of Poisson approximation. Oct. 87. *Stochastic Proc. Appl.*, 30, 1988, 317-327.

214. C. Houdré. A vector bimeasure integral with some applications. June 88 (Revised).

215. M.R. Leadbetter. On the exceedance random measures for stationary processes. Nov. 87 (Revised as No. 252).

216. M. Marques. A study on Lebesgue decomposition of measures induced by stable processes. Nov. 87 (*Dissertation*).

217. M.T. Alpuim. High level exceedances in stationary sequences with extremal index. Dec. 87. *Stochastic Proc. Appl.*, 30, 1988, 1-16.

218. R.F. Serfozo. Poisson functionals of Markov processes and queueing networks. Dec. 87. *Adv. Appl. Prob.*, 21, 1989, 595-611.

219. J. Bather. Stopping rules and ordered families of distributions. Dec. 87. *Sequential Anal.*, 7, 1988, 111-126.

220. S. Cambanis and M. Maejima. Two classes of self-similar stable processes with stationary increments. Jan. 88. *Stochastic Proc. Appl.*, 32, 1989, 305-329.

221. H.P. Hucke, G. Kallianpur and R.L. Karandikar. Smoothness properties of the conditional expectation in finitely additive white noise filtering. Jan. 88. *J. Multivariate Anal.*, 27, 1988, 261-269.

222. I. Mitoma. Weak solution of the Langevin equation on a generalized functional space. Feb. 88. (Revised as No. 238).

223. L. de Haan, S.I. Resnick, H. Rootzén and C. de Vries. Extremal behaviour of solutions to a stochastic difference equation with applications to arch-processes. Feb. 88. *Stochastic Proc. Appl.*, 32, 1989, 213-224.

224. O. Kallenberg and J. Szulga. Multiple integration with respect to Poisson and Lévy processes. Feb. 88. *Prob. Theor. Rel. Fields*, 1989, to appear.

225. D.A. Dawson and L.G. Gorostiza. Generalized solutions of a class of nuclear space valued stochastic evolution equations. Feb. 88. *Appl. Math. Optimization*, to appear.

226. G. Samorodnitsky and J. Szulga. An asymptotic evaluation of the tail of a multiple symmetric α-stable integral. Feb. 88. *Ann. Probability*, to appear.

227. J.J. Hunter. The computation of stationary distributions of Markov chains through perturbations. Mar. 88.

228. H.C. Ho and T.C. Sun. Limiting distribution of nonlinear vector functions of stationary Gaussian processes. Mar. 88. *Ann. Probability*, to appear.

229. R. Brigola. *On functional estimates for ill-posed linear problems*. Apr. 88.

230. M.R. Leadbetter and S. Nandagopalan. On exceedance point processes for stationary sequences under mild oscillation restrictions. Apr. 88. *Proc. Oberwolfach Conf. on Extremal Value Theory*. J. Husler and R. Reiss, eds., Springer, to appear.

231. S. Cambanis, J. P. Nolan and J. Rosinski. On the oscillation of infinitely divisible processes. Apr. 88. *Stochastic Proc. Appl.*, to appear.

232. G. Hardy, G. Kallianpur and S. Ramasubramanian. A nuclear space-valued stochastic differential equation driven by Poisson random measures. Apr. 88.

233. D.J. Daley, T. Rolski. Light traffic approximations in queues (II). May 88. *Math. Operat. Res.*, to appear.

234. G. Kallianpur, I. Mitoma, R.L. Wolpert. Nuclear space-valued diffusion equations. July 88. *Stochastics*, 1989, to appear.

235. S. Cambanis. Admissible translates of stable processes: A survey and some new models. July 88.

236. E. Platen. On a wide range exclusion process in random medium with local jump intensity. Aug. 88.

237. R.L. Smith. A counterexample concerning the extremal index. Aug. 88. *Adv. Appl. Probab.*, 20, 1988, 681-683.

UNCLASSIFIED
SECURITY CLASSIFICATION OF THIS PAGE

REPORT DOCUMENTATION PAGE

Form Approved OMB No. 0704-0188

1a. REPORT SECURITY CLASSIFICATION Unclassified	1b. RESTRICTIVE MARKINGS
2a. SECURITY CLASSIFICATION AUTHORITY N/A 2b. DECLASSIFICATION / DOWNGRADING SCHEDULE N/A	3. DISTRIBUTION / AVAILABILITY OF REPORT Approved for Public Release; Distribution Unlimited
4. PERFORMING ORGANIZATION REPORT NUMBER(S) Technical Report No. 278	5. MONITORING ORGANIZATION REPORT NUMBER(S) AFOSR-TR- 90-0273
6a. NAME OF PERFORMING ORGANIZATION University of North Carolina Center for Stochastic Processes 6b. OFFICE SYMBOL (If applicable)	7a. NAME OF MONITORING ORGANIZATION AFOSR/NM
6c. ADDRESS (City, State, and ZIP Code) Statistics Department CB #3260, Phillips Hall Chapel Hill, NC 27599-3260	7b. ADDRESS (City, State, and ZIP Code) Bldg. 410 Bolling Air Force Base, DC 20332-6448
8a. NAME OF FUNDING / SPONSORING ORGANIZATION AFOSR 8b. OFFICE SYMBOL (If applicable) NM	9. PROCUREMENT INSTRUMENT IDENTIFICATION NUMBER F49620 85C 0144
8c. ADDRESS (City, State, and ZIP Code) Bldg. 410 Bolling AFB, DC 20332-6448	10. SOURCE OF FUNDING NUMBERS PROGRAM ELEMENT NO. 6.1102F PROJECT NO. 2304 TASK NO A5 WORK UNIT ACCESSION NO.

11. TITLE (Include Security Classification)
Estimation of Hilbert space valued parameters by the method of sieves.

12. PERSONAL AUTHOR(S)
Kallianpur, G. and Selukar, R.S.

13a. TYPE OF REPORT	13b. TIME COVERED	14. DATE OF REPORT (Year, Month, Day)	15. PAGE COUNT
preprint	FROM ______ TO ______	October 1989	46

16. SUPPLEMENTARY NOTATION
None

17. COSATI CODES			18. SUBJECT TERMS (Continue on reverse if necessary and identify by block number)
FIELD	GROUP	SUB-GROUP	Hilbert space, sieve estimation.
XXXXXX	XXXXXXXXXX	XXX	

19. ABSTRACT (Continue on reverse if necessary and identify by block number)

See over.

20. DISTRIBUTION / AVAILABILITY OF ABSTRACT ☒ UNCLASSIFIED/UNLIMITED ☐ SAME AS RPT. ☐ DTIC USERS	21. ABSTRACT SECURITY CLASSIFICATION UNCLASSIFIED
22a. NAME OF RESPONSIBLE INDIVIDUAL Professor Eytan Barouch	22b. TELEPHONE (Include Area Code) (202)767-~~5026~~ 4940 22c. OFFICE SYMBOL AFOSR/NM

DD Form 1473, JUN 86 *Previous editions are obsolete.*

SECURITY CLASSIFICATION OF THIS PAGE
UNCLASSIFED

UNCLASSIFIED

19. Abstract.

By extending the ideas of Ibragimov & Hasminski in the finite dimensional parameter estimation a large deviation inequality for a sieve estimator estimating a Hilbert space valued parameter is obtained. This sieve estimator corresponds to a sieve which consists of finite dimensional, compact, convex sets. The inequality suggests a procedure of consistent estimation of Hilbert space valued parameters and naturally provides the convergence rates of the resultant estimators.

The usefulness of this approach is demonstrated by applying it to two examples; the first one deals with the estimation of the drift function in a linear stochastic differential equation and the second problem is of the intensity estimation of a nonstationary Poisson process. A detailed discussion of the convergence rates of our estimators and how they compare with the other estimators proposed in the literature is given in both cases.

Estimation of Hilbert space

valued parameters by the method of sieves.

by

G. Kallianpur and R. S. Selukar *

University of North Carolina at Chapel Hill

By extending the ideas of Ibragimov & Hasminski in the finite dimensional parameter estimation a large deviation inequality for a sieve estimator estimating a Hilbert space valued parameter is obtained. This sieve estimator corresponds to a sieve which consists of finite dimensional, compact, convex sets. The inequality suggests a procedure of consistent estimation of Hilbert space valued parameters and naturally provides the convergence rates of the resultant estimators.

The usefulness of this approach is demonstrated by applying it to two examples; the first one deals with the estimation of the drift function in a linear stochastic differential equation and the second problem is of the intensity estimation of a nonstationary Poisson process. A detailed discussion of the convergence rates of our estimators and how they compare with the other estimators proposed in the literature is given in both cases.

Key words: Hilbert space, sieve estimation.

AMS 1980 subject classification: Primary 62G05; Secondary 62M99

* Now at the University of Illinois at Urbana-Champaign

Research partially supported by the Air Force Office of Scientific Research Contract No. F49620 85C 0144.

Estimation of Hilbert space valued parameters by the method of sieves.

I. Introduction By now the method of sieves, first suggested by Grenander, has become a fairly common technique in the infinite dimensional parameter estimation. There are several specific examples (see Grenander (1981), Geman (1981), Karr (1987)) where this method has yielded consistent estimators with good asymptotic properties. However, the situation is quite different when it comes to finding general conditions which imply the consistency of a sieve estimator and give an idea about its rate of convergence.

The first step in this direction was taken by Geman & Hwang (1982) who generalized the Wald's (1949) proof of the consistency of the maximum likelihood estimator of a finite dimensional parameter to the case of infinite dimensional parameter via the method of sieves. They give conditions on the likelihood function and the parameter space (which is assumed to be a Polish space) so that there exists a sieve such that the corresponding sieve estimator is consistent. This result is quite general but it gives no information about the rate of convergence of the resultant sieve estimator and moreover, the observations are assumed to be independent and identically distributed (i.i.d.). The other results of this type (see, e.g., Karr (1987)) assume some additional structure on the likelihood function as well as on the parameter space.

Here we generalize the ideas of Ibragimov & Hasminski (in finite dimensional parameter estimation) and obtain a large deviation inequality for a sieve estimator estimating a Hilbert space valued parameter. This inequality suggests a method of consistent estimation of Hilbert space valued parameters using the sieve estimators which correspond to the sieves consisting of finite dimensional, compact, convex sets. This method naturally provides the convergence rates of the resultant estimators and the observations do not have to be i.i.d.. The usefulness of this approach is demonstrated by applying it to two examples; in the first one, the drift function in a linear stochastic differential equation is estimated and in the next, the intensity function of a nonstationary Poisson process is estimated. In both cases, a detailed discussion of the convergence rates of our estimators and how they compare with the other estimators proposed in the literature is given.

The article is arranged in the following manner: the rest of this section is devoted to the brief explanation of our approach, Section II consists of the basic large deviation inequality, its derivation and the consequences, Section III contains the first example of the drift function estimation and Section IV contains the Poisson intensity estimation.

Let $(X_1, X_2, \ldots, X_n)$ be an n-dimensional random vector (X_is can be general random objects) with joint distribution $P^n{}_\theta$. Assume that, for each n, $P^n{}_\theta$ posseses a density, $f^n(\theta)$, w.r.t. a σ-finite measure ν^n and the unknown parameter θ is an element of Θ, a subset of a real, separable, infinite dimensional Hilbert space H. The goal is to estimate θ consistently using the observations $(X_1, X_2, \ldots, X_n)$.

Let $(g_j), j \geq 1$; be an arbitrary but fixed set of independent vectors from H and $S_k, k \geq 1$, be k-dimensional, compact, convex subsets in the subspaces of Θ spanned by $(g_j), j \leq k$ such that:

i) $S_k \subset S_{k+1}$,

ii) $\cup S_k$ is dense in Θ and

iii) For every n and k, there exists a (random) point $\hat{\theta}_k{}^n$ such that

$$f^n(\hat{\theta}_k{}^n) = \sup_{\theta' \in S_k} f^n(\theta').$$

That is, (S_k) is a sieve consisting of *finite* dimensional, compact, convex sets; $\hat{\theta}_k{}^n$ is called a sieve estimator (for each n, it is merely a restricted maximum likelihood estimator).

For $\theta \in \Theta$, let θ_k denote its projection on S_k. Then, since both θ_k and $\hat{\theta}_k{}^n$ are finite dimensional, using the techniques of Ibragimov & Hasminski (1981), one can obtain an upper bound on the probability:

$$P^n{}_\theta \{ n^\beta \| \hat{\theta}_k{}^n - \theta_k \| > h \} \tag{1.1}$$

where $\|.\|$ denotes the norm in H, and $\beta \geq 0$ and $h > 0$ are constants.

Now (ii)implies that, for every $k_n \to \infty$, the deterministic difference (bias) $\|\theta_{k_n} - \theta\| \to 0$. Therefore, to estimate θ consistently, it is enough to estimate θ_{k_n} consistently for some $k_n \to \infty$ (i.e. the error in estimating θ_{k_n} should tend to zero as $n \to \infty$). Thus it is sufficient to show the existence of a subsequence k_n such that, for every $h > 0$ and some $\beta \geq 0$, the bound for the probability in (1.1) tends to zero. In fact, this also means that the rate of convergence of the sieve estimator, $\hat{\theta}_{k_n}$, is atleast n^β if, in addition, we have $n^\beta \|\theta_{k_n} - \theta\| \to 0$.

In Theorem 2.1 an exponential bound is obtained for the probability in (1.1) under conditions similar to those of Ibragimov & Hasminski (1981, Theorem 5.1, Ch.1). In Corollary 2.1 sufficient conditions are stated for the existence of a subsequence k_n such that the sieve estimator $\hat{\theta}_{k_n}$ is consistent. In the examples considered in Sections III & IV the conditions of this corollary are verified.

II. Let $\{ \chi^n, \mathbf{u}^n, \mathbf{P}_\theta{}^n \}$ be a family probability spaces indexed by $n \geq 1$ and $\theta \in \Theta$. Points (observations) from χ^n will be denoted by X^n. Assume that the parameter space Θ is a subset of some fixed infinite dimensional separable Hilbert space $\mathbf{H}$. Also assume that, for every n there exists a σ-finite measure ν^n such that $P_\theta{}^n$ is absolutely continuous w.r.t. ν^n for all $\theta \in \Theta$. Let $f^n(X^n, \theta)$ denote the corresponding density.

Definition 1 (sieve) A sequence $\{S_k\}$ of subsets of Θ is called a sieve if $\cup S_k$ is a dense set in Θ and for each n and k the maximum of $f^n(X^n, \theta)$ over S_k is attained in S_k.

Definition 2 (sieve estimator) An estimator $\hat{\theta}_k{}^n(X^n)$ is called a sieve estimator *corresponding* to the sieve $\{S_k\}$ if it is a point in S_k such that,

$$f^n(X^n, \hat{\theta}_k^n) = \sup_{t \in S_k} f^n(X^n, t).$$

Our aim is to study the asymptotic behavior of a sieve estimator $\hat{\theta}_k^n$ corresponding to a given sieve (S_k), $k \geq 1$, of the following type :

i) for each $k \geq 1$, S_k is a compact, convex set in the span of $\{g_j\}$ $1 \leq j \leq k$; where $\{g_j\}$ is some fixed set of independent vectors from Θ.

ii) $S_k \subset S_{k+1}$ and $\cup S_k$ is dense in Θ.

For any $\theta \in \Theta$, let θ_k denote its projection in S_k i. e.

$$\| \theta - \theta_k \| = \min_{t \in S_k} \| \theta - t \|.$$

Remark 2.1 : Of course, our setup does not guarantee the existence of such a sieve, we are *assuming* that we already have such a sieve.

Since S_k is compact and convex, θ_k always exists and is unique. Furthermore, $\|\theta - \theta_k\| \to 0$ as $k \to \infty$.

Let $\theta \in \Theta$ be the true parameter.

Let $\Phi^n{}_k$ be $k \times k$ nondegenerate normalizing matrices with real elements such that, $|\Phi^n{}_k| \to 0$ as $n \to \infty$ where $|\Phi^n{}_k|$ denotes the operator norm of $\Phi^n{}_k$. Note that $\Phi^n{}_k$ can be considered as a linear operator on the span of $\{g_j\}$, $1 \leq j \leq k$ in an obvious way. With this understanding let us define real valued random functions $Z^n{}_k(u)$ with domains $U^n{}_k$ as follows :

$$U^n{}_k = (\Phi^n{}_k)^{-1}(S_k - \theta_k); \tag{2.1}$$

and for $u \in U^n{}_k$,

$$Z^n{}_k(u) = l_n(\theta_k + \Phi^n{}_k u) - l_n(\theta_k) \tag{2.2}$$

where, for $\theta \in \Theta$, $l(X^n, \theta) = l_n(\theta)$ is the loglikelihood function of the data at θ. Then, since $\hat{\theta}_k^n$ maximizes $l_n(\theta)$ in S_k and $Z^n{}_k(0) = 0$ it is easy to verify that $\{ \|(\Phi^n{}_k)^{-1}(\hat{\theta}_k^n - \theta_k)\| \geq h \} \subset \{ \sup_{\|u\| > h} Z^n{}_k(u) \geq 0 \}$. Therefore the following inequality is true:

$$P_\theta^n\{\,||(\Phi^n{}_k)^{-1}(\hat{\theta}_k{}^n-\theta_k)\,||\geq h\,\}\leq P_\theta^n\{\sup_{||u||>h} Z^n{}_k(u)\geq 0\}. \quad (2.3)$$

Under suitable assumptions on the functions $Z^n{}_k(u)$ the probability on the RHS above can be bounded.

Let $\mathbf{G}$ denote the following class of functions :

$g\in\mathbf{G}$ if and only if,

i) $g:[0,\infty)\to R^1$ and there exists x_0 such that g is positive and strictly increasing on $[x_0,\infty)$.

ii) For all $N>0$,

$$\int_0^\infty y^N\exp(-g(y))\,dy<\infty.$$

Throughout this article C, with or without a subscript, will denote a positive constant <u>independent</u> of n and k; it need not always be the same. Also, $[x]$ will denote the largest integer smaller than x.

Theorem 2.1 *Suppose that for each $k\in N$ the following conditions hold :*

C1) There exist numbers $\alpha=\alpha(k)>k$, $m=m(k)\geq\alpha$ and positive constants $B^n{}_k$ and $p=p(k)$ such that for all $R>0$,

$$\sup_{\substack{||u||,||v||\leq R,\\ u,v\in U^n{}_k}} E^n{}_\theta\,|Z^n{}_k(u)-Z^n{}_k(v)|^m\leq B^n{}_k\,(1+R^p)\,||u-v||^\alpha.$$

C2) There exists $g^n{}_k\in\mathbf{G}$ such that for some $\eta>0$,

$$E^n{}_\theta\exp(\eta Z^n{}_k(u))\leq\exp(-g^n{}_k(||u||))$$

for all $u\in U^n{}_k$.

Then, for $h\in[x_0(n,k),\infty)$,

$$P_\theta^n\{\,||(\Phi^n{}_k)^{-1}(\hat{\theta}_k{}^n-\theta_k)\,||\geq h\}\leq B_1\sum_{r=0}^\infty(1+h+r)^{B_2}g^n{}_{kr}\exp(-b_1 g^n{}_{kr})$$

where

i) $x_0(n,k)$ *is a point such that* $g^n{}_k$ *is positive and increasing on* $[x_0(n,k), \infty)$;

ii) $g^n{}_{kr} = g^n{}_k(h+r)$,

iii) $b_1 = b_1(k) = \dfrac{\alpha-k}{\alpha-k+mk} > 0$,

iv) $B_1 = B_1(n,k) = C\,(B^n{}_k)^{\frac{1}{m}} D$,

$$D = (24)^{2k+4}.\left(\frac{1}{1-2^{\frac{-(\alpha-k)}{m}}}+\frac{1}{2^{\frac{m-\alpha}{m}}-1}\right)$$

if $m > \alpha$ *and*

$$D = (24)^{2k+4}.\left(\frac{1}{1-2^{\frac{-(\alpha-k)}{m}}}\right)$$

if $m = \alpha$.

v) $B_2 = B_2(k) = \dfrac{2\alpha+p}{m}$.

From this result a useful corollary can be deduced which gives sufficient conditions for the existence of a consistent sieve estimator.

Corollary 2.1 *Assume the conditions of Theorem 2.1. Furthermore assume that:*

i) The constants $m=m(k)$, $p=p(k)$ *and* $\alpha=\alpha(k)$ *in condition C1 are constant multiples of* k, *the multiplying constant being independent of* k.

ii) $B^n{}_k \le C^k\, n^{Ck}\, k^{Ck}$ *for some* C

iii) $g^n{}_k(x) = n^{\delta} k^{\mu} |x - a(n,k)|^{\nu} - t(n,k)$

where δ, μ *and* ν *are fixed positive numbers and* $t(n,k)$ *and* $a(n,k)$ *are non-negative constants which may depend on* n *and* k.

iv) For some number $s > 0$ *the following holds:*

a) $\delta - \mu s - 2s > 0$

b) $a(n, n^s)$ *and* $\frac{t(n, n^s)}{n^s}$ *tend to zero as* $n \rightarrow \infty$.

Then $\hat{\theta}_{n'}{}^{n}$ *is a consistent estimator of* θ *and moreover, for* n *sufficiently large,* $n > n_0(h)$,

$$P^n{}_\theta\{\ |(\Phi^n{}_{n'})^{-1}(\hat{\theta}_{n'}{}^{n} - \theta_{n'})| > h\ \} \leq C \exp(-C_1 n^{\delta - s\mu - s} h^{\nu})$$

where the constants C *and* $C_1 > 0$ *depend on* θ *alone.*

Proofs : First a result from Ibragimov & Hasminski (1981) (Theorem 19, Appendix) is stated. This result gives conditions on a random function ξ, defined on R^k, so that it will have continuous paths. Furthermore, the conditions give a very useful bound on the expectation of the modulus of continuity of ξ.

Theorem A.1 *Let* $\xi(t)$ *be a real valued random function defined on* F, *a closed set in* R^k. *We shall assume that the process* $\xi(t)$ *is measurable and separable. Also assume that the following conditions are fulfilled :*

I) There exist $m \geq \alpha > k$ *such that for all* $x \in F$,

$$E\,|\xi(x)|^m \leq H(x)$$

II) For $h \in R^k$ *such that* $(x+h) \in F$,

$$E\,|\,\xi(x+h) - \xi(x)\,|^m \leq H(x)\ |h|^\alpha$$

where $H(x)$ *is some positive continuous function defined on* F .

Then with probability one the realizations of $\xi(t)$ *are continuous functions.*

Moreover, if $\omega(h, \xi, L) = \sup\{\,|\,\xi(x) - \xi(y)\,|\,\}$ *where the sup is taken over* $x, y \in F$ *with* $|x-y| < h$, $|x|, |y| < L$;

then

$$E\,\omega(h, \xi, L) \leq D(k, \alpha, m)\,(L^k \sup_{|x| \leq L} H(x))^{\frac{1}{m}}\ h^{\frac{\alpha - k}{m}}\ \ln h^{-1}$$

where the constant $D(k, \alpha, m)$ *can be bounded by*

$$24^{2k+4} \cdot \left(\frac{1}{1-2^{\frac{-(\alpha-k)}{m}}} + \frac{1}{2^{(\frac{m-\alpha}{m})}-1} \right)$$

if $m > \alpha$ *and if* $m = \alpha$ *then the second term in the above product is* $(1-2^{\frac{-(\alpha-k)}{m}})^{-1}$.

This bound on D is not explicitly given in Ibragimov & Hasminski (1981) but can be obtained by carefully going over their proof; the additional factor of $\ln h^{-1}$ in the bound on the modulus of continuity also seems to be necessary when $m = \alpha$. The numerical value of this bound in terms of k, m and α is very important since this bound will have to be taken into account while determining the subsequence k_n which controls the growth of the sets in the sieve. The details of the derivation of this bound are quite lengthy and so they are omitted, interested reader can find them in Selukar (1989, Appendix).

Proof of Theorem 2.1 : For $r=0,1,2,\ldots.$, let

$$\gamma_r^n = \{ u \in U^n{}_k : h+r < \|u\| \le h+r+1 \}. \qquad (2.4)$$

We will show that,

$$P_\theta^n [\sup_{u \in \gamma_r^n} Z^n{}_k(u) \ge 0] \le B_1 (1+h+r)^{B_2} g^n{}_{kr} \exp(-b_1 g^n{}_{kr}) \qquad (2.5)$$

where b_1, B_1 & B_2 are as in Theorem A.1. (2.5) will prove the assertion of the theorem since, from (2.3),

$$P_\theta^n \{ \| (\Phi^n{}_k)^{-1} (\hat\theta_k{}^n - \theta_k) \| \ge h \} \le P_\theta^n [\sup_{\|u\| > h} Z^n{}_k(u) \ge 0]$$

$$\le \sum_{r=0}^{\infty} P_\theta^n [\sup_{u \in \gamma_r^n} Z^n{}_k(u) \ge 0]$$

$$\le B_1 \sum_{0}^{\infty} B_1 (1+h+r)^{B_2} g^n{}_{kr} \exp(-b_1 g^n{}_{kr});$$

which is the assertion of the theorem. Now we proceed to prove (2.4) :

We subdivide the set

$\gamma_r{}^n = \{ u \in U^n{}_k : h+r < \|u\| \leq h+r+1 \}$ into N regions, each with diameter at most $2(1+h+r)\delta$, δ to be specified later. This subdivision can be done such that, $N \leq (\delta)^{-k}$. Denote the above subdivision of $\gamma_r{}^n$ by, $\gamma_{r1}{}^n, \gamma_{r2}{}^n, \ldots \gamma_{rN}{}^n$ and let u_j be any fixed point in $\gamma_{rj}{}^n$. Then,

$$P_\theta{}^n [\sup_{u \in \gamma_r{}^n} Z^n{}_k(u) \geq 0] \leq \sum_{j=1}^{N} P_\theta{}^n \{ Z^n{}_k(u_j) \geq -\frac{1}{2} \}$$

$$+ P_\theta{}^n \{ \sup_{\substack{\|u-v\| < 2\delta^* \\ \|u\|, \|v\| < 1+h+r}} (| Z^n{}_k(u) - Z^n{}_k(v) |) \geq \frac{1}{2} \}. \tag{2.6}$$

(here, $\delta^* = (1+h+r)\,\delta$)

The first term on the RHS above can be bounded using condition C2 and the Chebyshev's inequality as follows :

$$\sum_{j=1}^{N} P_\theta{}^n \{ Z^n{}_k(u_j) \geq -\frac{1}{2} \} \leq N\, e^{\frac{n}{2}} \exp(-g^n{}_{kr}) \tag{2.7}$$

($\|u_j\| \in (h+r, 1+h+r]$ implies that $g^n{}_k(\|u_j\|) \geq g^n{}_k(h+r) = g^n{}_{kr}$)

Also since, $Z^n{}_k(0) = 0$ condition C1 implies that

$$\sup_{\|u\| \leq R} E^n{}_\theta | Z^n{}_k(u) |^m \leq B^n{}_k (1+R^p) R^\alpha. \tag{2.8}$$

Hence, in view of condition C1, (2.8) and Theorem A.1 we get that,

$$E^n{}_\theta\, \omega(2\delta(1+h+r); Z^n{}_k, (1+h+r)) \leq D\, [B^n{}_k (1+(1+h+r))^{p+\alpha}]^{\frac{1}{m}}$$

$$(1+h+r)^{\frac{k}{m}} [2\delta(1+h+r)]^{\frac{\alpha-k}{m}} \ln[2\delta(1+h+r)]^{-1}$$

$$\leq D\, [B^n{}_k (2+h+r)^{p+2\alpha}]^{\frac{1}{m}} [2\delta]^{\frac{\alpha-k}{m}} \ln\delta^{-1}. \tag{2.9}$$

Using (2.9) and the Chebyshev's inequality the second term in the RHS of (2.6) can be bounded as follows

$$P_\theta^n \{ \sup_{\substack{|u-v|< 2\delta^* \\ |u|,|v|< 1+h+r}} |Z^n{}_k(u) - Z^n{}_k(v)| \geq \frac{1}{2} \}$$

$$\leq 4D\, [B^n{}_k\, (2+h+r)^{p+2\alpha}]^{\frac{1}{m}}\, \delta^{\frac{\alpha-k}{m}}\, \ln\delta^{-1}. \tag{2.10}$$

Thus from (2.6), (2.7) and (2.10) we get,

$$P_\theta^n [\sup_{u\in\gamma_r^n} Z^n{}_k(u) \geq 0]$$

$$\leq N\, e^{\frac{\eta}{2}} \exp(-g^n{}_{kr}) + 4D\,[B^n{}_k(2+h+r)^{p+2\alpha}]^{\frac{1}{m}}\,\delta^{\frac{\alpha-k}{m}}\,\ln\delta^{-1}. \tag{2.11}$$

So far we have not chosen δ. We will choose δ such that $N\ \exp(-g^n{}_{kr})$ and $\delta^{\frac{\alpha-k}{m}}$ are of the same order of magnitude. Let

$\delta = \exp(-\frac{m}{(\alpha-k)+mk} g^n{}_{kr})$. Then $\delta^{\frac{\alpha-k}{m}} = \exp(-\frac{\alpha-k}{(\alpha-k)+mk} g^n{}_{kr})$; and since, $N \leq \delta^{-k}$,

$$N\ \exp(-g^n{}_{kr}) \leq \delta^{-k}\ \exp(-g^n{}_{kr})$$

$$= \exp(-\frac{\alpha-k}{(\alpha-k)+mk} g^n{}_{kr}). \tag{2.12}$$

From (2.11) and (2.12) we get,

$$P_\theta^n [\sup_{u\in\gamma_r^n} Z^n{}_k(u) \geq 0]$$

$$\leq \exp(-\frac{\alpha-k}{(\alpha-k)+mk} g^n{}_{kr})\,[e^{\frac{\eta}{2}} + 4D\,[B^n{}_k\,(2+h+r)^{p+2\alpha}]^{\frac{1}{m}}\, g^n{}_{kr}].$$

Hence

$$P_\theta^n [\sup_{u\in\gamma_r^n} Z^n{}_k(u) \geq 0] \leq B_1(1+h+r)^{B_2}\exp(-b_1 g^n{}_{kr})\, g^n{}_{kr}$$

which completes the proof.

Proof of Corollary 2.1: Let $g_n(x) = g^n{}_{[n^s]}(x)$; then

$$g_n(x) = C\, n^{\delta-\mu s}\, |x - a(n)|^{\nu} - t(n) \tag{2.13}$$

where

$a(n) = a(n,[n^s])$ and $t(n) = t(n,[n^s])$. It is easy to check that $g_n(x)$ is *positive* and *increasing* on $[x_0(n), \infty)$ where

$$x_0(n) = a(n) + \left\{ \frac{n^{-(\delta-\mu s - s)}}{C} \cdot \frac{t(n)}{[n^s]} \right\}^{\frac{1}{\nu}}.$$

In view of assumption (iv) (a) and (b),

$\delta - \mu s - s > \delta - \mu s - 2s > 0$ and $a(n)$ and $t(n)$ are such that

$$a(n) \to 0 \quad \text{and} \quad \frac{t(n)}{[n^s]} \to 0.$$

Therefore $x_0(n) \to 0$ as $n \to \infty$. Hence for every $h > 0$, there exists $n_0(h)$ such that for all $n > n_0(h)$, $x_0(n) < h$. Then the following inequality is a consequence of Theorem 2.1.

$$P^n{}_\theta \{ \| (\Phi^n{}_{n'})^{-1} (\hat{\theta}_{n'}{}^n - \theta_{n'}) \| > h \}$$

$$\leq B_1 \sum_{r=0}^{\infty} (1+h+r)^{B_2} \exp(-b_1 g_n(h+r)) \tag{2.14}$$

where because of the assumptions (i) & (ii) of the Corollary,

$$b_1 = b_1(k) = \frac{\alpha - k}{\alpha - k + mk} - \frac{C}{k} = \frac{C}{[n^s]},$$

$$B_2 = \frac{p + 2\alpha}{m} - C,$$

$$B_1 = C\, (B^n{}_k)^{\frac{1}{m}} D(k,r,m) \leq C\, (24)^{Ck} k^C n^C$$

$$= C\, (24)^{C[n^s]} n^C$$

and $|g_n(h+r)| \le n^C k^C (h+r)^C$. (In all of the above expressions the constant C denotes a positive number independent of n but it could be different in each of them.)

This means that the expression in (2.14) can be written as

$$P^n{}_\theta \{ \| (\Phi^n{}_{n^s})^{-1} (\hat\theta_{n^s}{}^n - \theta_{n^s}) \| > h \}$$

$$\le C n^C (24)^{C[n^s]} \sum_{r=0}^{\infty} (1+h+r)^{B_2} \exp(-\frac{C}{[n^s]} g_n(h+r))$$

$$\le (C n^C (24)^{C[n^s]} \exp(-\frac{C}{2[n^s]} g_n(h)))$$

$$\times \sum_{r=0}^{\infty} (1+h+r)^{B_2} \exp(-\frac{C}{[n^s]} g_n(h+r)) \tag{2.15}$$

(since $g_n(h+r) \ge g_n(h)$).

For large n the first factor in (2.15) can be made arbitrarily small, in particular smaller than 1. To see this, note that for large enough n

$$a(n) < \frac{h}{2} \quad \text{and} \quad \frac{t(n)}{[n^s]} < \frac{1}{2}$$

then

$$\frac{1}{[n^s]} g_n(h) = \frac{1}{[n^s]} \{ C n^{\delta-\mu s} |h-a(n)|^{\nu} - t(n) \}$$

$$\ge C n^{(\delta-\mu s-s)} h^{\nu} - \frac{1}{2}.$$

Therefore

$$C n^C (24)^{C[n^s]} \exp(-\frac{C}{2[n^s]} g_n(h))$$

$$\le C n^C (24)^{C[n^s]} \exp(-C n^{(\delta-\mu s-s)} h^{\nu} + \frac{1}{2})$$

$$= C\, n^{C}\, (24)^{C[n']} \exp(-C\, n^{(\delta-\mu s-s)}\, h^{\nu}).$$

However since $h > 0$ and $\delta-\mu s-s > s$ the RHS above tends to zero.

Thus for large n,

$$P^{n}{}_{\theta}\{\, \| (\Phi^{n}{}_{n'})^{-1} (\hat{\theta}_{n'}{}^{n} - \theta_{n'}) \| > h \,\}$$

$$\leq \sum_{r=0}^{\infty} (1+h+r)^{B_2} \exp(-\frac{C}{[n^{s}]} g_n(h+r))$$

$$\leq \exp(-\frac{C}{2} n^{(\delta-\mu s-s)} h^{\nu}) \sum_{r=0}^{\infty} (1+h+r)^{B_2} \exp(-\frac{C}{2[n^{s}]} g_n(h+r))$$

$$\leq \exp(-\frac{C}{2} n^{(\delta-\mu s-s)} h^{\nu})$$

$$\times \sum_{r=0}^{\infty} (1+h+r)^{B_2} \exp(-\frac{C}{2[n^{s}]} n^{(\delta-\mu s-s)} (h+r)^{\nu})$$

(because for large n, $\dfrac{g_n(h+r)}{[n^{s}]} \geq C\, n^{(\delta-\mu s-s)} (h+r)^{\nu}$.)

$$\leq \exp(-\frac{C}{2} n^{(\delta-\mu s-s)} h^{\nu}). \qquad (2.16)$$

The last step follows because, using the Dominated convergence theorem it is easy to see that the second factor in the product above tends to zero as $n \to \infty$. The assertion in (2.16) essentially proves the Corollary since, $|(\Phi^{n}{}_{n'})^{-1}| \to 0$ as $n \to \infty$ implies that

$$|\hat{\theta}_{n'}{}^{n} - \theta_{n'}| \to 0 \text{ as } n \to \infty$$

and then we only need to observe that the deterministic difference $|\theta - \theta_{n'}|$ always $\to 0$ as $n \to \infty$.

III. Drift function estimation in linear SDE:

Consider the following model:

$$dx_t = \theta_t \, x_t \, dt + dW_t, \quad x_0 = X_0, \; 0 \le t \le 1; \tag{3.1}$$

where (W_t) is the standard Weiner process and the initial value X_0 is a zero mean Gaussian random variable with varience equal to σ_0^2. Furthermore, X_0 is assumed to be independent of (W_t).

The goal is to estimate (θ_t), an unknown deterministic function, using a sample of n independent trajectories of (x_t).

We assume that $\theta \in L^2([0,1],dt)$, that is, the parameter space $\Theta = L^2([0,1],dt)$. At this point we make an assumption that $\sigma_0^2 > 0$. This assumption is not very essential except that, in the presence of this assumption the consistency can be obtained in $L^2([0,1],dt)$ and the statements of the results are simpler whereas if this assumption is violated i.e. if $X_0 = 0$ *a.s.*, then the consistency is obtained in $L^2([0,1],\sigma_\theta^2(t)dt)$ where $\sigma_\theta^2(t) = E(x_t^2)$; (see Nguyen & Pham (1982) Remark pp 609).

The above problem was first discussed in a paper by Nguyen and Pham (1982). They proposed a sieve estimator and showed its *weak* mean square consistency; no regularity assumptions on θ were made. However, the rate of convergence of this estimator was not discussed. Later Kutoyants (1984) proposed a kernel estimator and proved its *weak pointwise* consistency under some mild differentiability assumptions on θ. He also discusses the rate of convergence of this estimator. Here we propose a sieve estimator which is based on the method suggested in the last section. It turns out that our estimator is a smoothed version of the estimator proposed by Nguyen & Pham. We will see that unless some kind of regularity assumption on θ is made the method in Section II does not yield the consistency of the proposed estimator, however, under mild regularity conditions on θ this estimator is shown to be mean square consistent as well as uniform norm consistent; the rate of convergence being dependent upon the amount of regularity assumed. We first define the necessary notation, state our results and then compare them with the results of Nguyen & Pham and Kutoyants. The proofs are given at the end of the section.

It is easy to verify that (3.1) has a unique strong solution (x_t) given by

$$x_t = x_0 \exp(\int_0^t \theta_u du) + \int_0^t \exp(\int_s^t \theta_u du) dW_s. \tag{3.2}$$

Note that (x_t) is a continuous Gaussian process with the mean fuction identically equal to zero, and if $\sigma_\theta^2(t) = Ex_t^2$ then it can be shown that

$$\sigma_\theta^2(t) = \sigma_0^2 \exp(2\int_0^t \theta_u du) + \int_0^t \exp(2\int_s^t \theta_u du) ds.$$

Obviously $\sigma_\theta^2(t)$ is a continuous function on $[0,1]$. Let

$$M_\theta = \sup_{0 \le t \le 1} \sigma_\theta^2(t) \quad \text{and} \quad m_\theta = \inf_{0 \le t \le 1} \sigma_\theta^2(t);$$

then, since $[0,1]$ is compact and $\sigma_\theta^2(t)$ continuous, $M_\theta < \infty$. Also, $m_\theta > 0$ since $\sigma_0^2 > 0$. This means that, for $f \in \Theta$,

$$m_\theta ||f||^2 \le ||f||^2_\theta \le M_\theta ||f||^2 \tag{3.3}$$

where $||f||^2_\theta = \int_0^1 f_t^2 \sigma_\theta^2(t) dt$ and $||f||^2 = \int_0^1 f_t^2 dt$.

For $\theta \in \Theta$, let P_θ denote the measure induced by the paths of (x_t) on $(C, \mathbf{C})$ where C is the space of continuous functions on $[0,1]$ endowed with the supremum norm topology and $\mathbf{C}$ is the Borel σ-field of subsets of C (that is, it is generated by the open sets in C). Then it is well known that (see, for example, Liptser and Shiryayev (1977), vol. I, Theorem 7.19, pp 277) P_θ and P_0 are mutually absolutely continuous (P_0 denoting the measure P_θ when $\theta \equiv 0$) and,

$$\frac{dP_\theta}{dP_0}(\omega) = \exp(\int_0^1 \theta_t \omega_t d\omega_t - \frac{1}{2}\int_0^1 \theta_t^2 \omega_t^2 dt) \tag{3.4}$$

where $\omega = (\omega_t) \in C$.

Suppose $\theta = (\theta_t)$, a point in Θ, is the true parameter.

Our aim is to estimate θ, using a sample of n *independent* trajectories of (x_t).

Let C^n be the n-fold product of C, $\mathbf{C}^n$ the n-fold product of $\mathbf{C}$ and $P_\theta{}^n$ the n-fold product of P_θ. Also, ω^n will denote a typical point in C^n.

With this notation, the family of probability spaces under consideration can be described as the triplet, $\{ C^n, \mathbf{C}^n, \mathrm{P}_\theta{}^n, \theta \in \Theta \}$.

The sieve S_k to be used is defined as follows:

Let $\{ f_j \}$ be an arbitrary but fixed complete orthonormal system (CONS) in $\Theta = L^2([0,1],dt)$ and for $k \geq 1$,

$$S_k = \{ \theta \in \Theta : \theta = \sum_{j=1}^{k} \theta_j f_j \; ; \; \sum_{j=1}^{k} \theta_j{}^2 \leq k^{2q} \}; \tag{3.5}$$

$q > 0$ will be determined later. Let $\hat{\theta}_k{}^n = \hat{\theta}_k{}^n(\omega^n)$ be the sieve estimator corresponding to this sieve. Now we are in position to state the main result of this section, the large deviation inequality for $\hat{\theta}_k{}^n$. However, we first have to define the domain of applicability of this result.

For $\theta \in \Theta$, let $P_k\theta$ and $\underline{\theta}_k$ be the projections of θ on the linear span of $\{f_j, 1 \leq j \leq k\}$ and S_k respectively, $\{f_j\}$ being the CONS used in defining the sieve S_k. (clearly, for large k, $k > K_\theta$, $P_k\theta = \underline{\theta}_k$) For $0 < \varepsilon < \frac{1}{2}$ let

$$\Theta_\varepsilon = \{ \theta \in \Theta : n^{1-\varepsilon} \, || \theta - P_{n^\varepsilon}\theta ||^2 \to 0 \text{ as } n \to \infty \}. \tag{3.6}$$

That is, $\theta \in \Theta_\varepsilon$ if and only if

$$n^{1-\varepsilon} \sum_{j=n^\varepsilon}^{\infty} \theta_j{}^2 \to 0 \text{ as } n \to \infty \tag{3.7}$$

where θ_j's are the Fourier coefficients of θ w.r.t. $\{ f_j \}$.

From the definition it is easy to see that,

$$\varepsilon_1 < \varepsilon_2 \Rightarrow \Theta_{\varepsilon_1} \subset \Theta_{\varepsilon_2}; \tag{3.8}$$

that is, smaller ε implies higher "smoothness". It is clear that the type of functions which belong to Θ_ε depends on the CONS used in defining the sieve S_k and for an arbitrary CONS it is difficult to ascertain whether a given function is a member of Θ_ε or not. However, for some special choices of CONSs it is possible to identify large classes of functions which belong to Θ_ε. One such choice of CONS $\{ f_j \}$ is illustrated below:

Set $f_1 \equiv 1$ and for $j \geq 2$, let

$$f_j(t) = \sqrt{2} \cos(\pi jt) \quad \text{for even } j\text{'s and}$$

$$f_j(t) = \sqrt{2} \sin[\pi (j+1) t] \quad \text{for odd } j\text{'s}; \quad t \in [0,1]. \tag{3.9}$$

For $m \geq 0$ and $0 \leq \alpha \leq 1$ let $\Theta_{m,\alpha}$ be the class of m-times differentiable functions with m-th derivative Lipschitz continuous of order α. More precisely,

$$\Theta_{m,\alpha} = \{ \theta \in \Theta : \theta \text{ is m times continuously differentiable}$$

$$\text{with m-th derivative Lipschitz continuous of order } \alpha.$$

$$\text{Furthermore, for } 0 \leq j \leq m,$$

$$\theta^{(j)}(0) = \theta^{(j)}(1) \text{ where, } \theta^{(j)}(t) = \frac{d^j}{dt^j} \theta(t) \} \tag{3.10}$$

If the CONS in (3.9) is used in defining the sieve S_k then it can be shown that (see Corollary 3.5),

$$\Theta_{m,\alpha} \subset \Theta_\varepsilon \quad \text{when} \quad \varepsilon > \frac{1}{1+2(m+\alpha)}. \tag{3.11}$$

Theorem 3.1 *Assume that the true parameter θ belongs to Θ_ε for some ε, $0 < \varepsilon < \frac{1}{2}$. Choose $\beta \geq 0$ and $q > 0$ such that*

$$b = 1 - 2\beta - \varepsilon - 2\varepsilon q > \varepsilon > 0.$$

Then, for all n *sufficiently large and* $h > 0$,

$$P^n{}_\theta \{ n^\beta \, || \, \hat{\theta}_{n^\varepsilon}{}^n - \underline{\theta}_{n^\varepsilon} \, || > h \} \leq \exp(-C\, n^b\, h^2)$$

where $C > 0$, *depends on* θ *alone.*

Now the Borel-Cantelli lemma and the above result give us the next theorem.

Theorem 3.2 *Assume that* $\theta \in \Theta_\varepsilon$ *for some* ε, $0 < \varepsilon < \frac{1}{2}$. *Choose* $\beta \geq 0$ *and* $q > 0$ *such that*

$1 - 2\beta - 2\varepsilon - 2\varepsilon q > 0$. *Then,*

$$n^\beta \, || \, \hat{\theta}_{n^\varepsilon}{}^n - \theta \, || \to 0 \text{ a.s. as } n \to \infty .$$

For the following results it is assumed that the CONS $\{ f_j \}$ used in defining the sieve S_k is the one given in (3.9). Then the next result is just a restatement of the last theorem using the fact given in (3.11).

Theorem 3.3 *Let* $\theta \in \Theta_{m,\alpha}$, $m + \alpha > \frac{1}{2}$. *Then for every* β, $\beta < \dfrac{(m+\alpha) - \frac{1}{2}}{2(m+\alpha)+1}$, *there exist* $\varepsilon > 0$ *and* $q > 0$ *(which depend on* β*) such that,*

$$n^\beta \, || \, \hat{\theta}_{n^\varepsilon}{}^n - \theta \, ||_\theta \to 0 \text{ a.s. as } n \to \infty .$$

It is well known that if $\theta \in \Theta_{m,\alpha}$ and $m + \alpha > \frac{1}{2}$, the Fourier series of θ converges absolutely and uniformly. Then a few additional computations give us the following relation :

$$\sup_{0 \le t \le 1} | \hat{\theta}_{n^\varepsilon}{}^{n}(t) - \underline{\theta}_{n^\varepsilon}(t) | \le n^{\frac{\varepsilon}{2}} \| \hat{\theta}_{n^\varepsilon}{}^{n} - \underline{\theta}_{n^\varepsilon} \| \tag{3.12}$$

(for $n > n_0(\theta)$).

Therefore, in order to show the uniform norm consistency of $\hat{\theta}_{n^\varepsilon}{}^{n}$ at the rate n^β, it suffices to obtain L^2-consistency at the rate $n^{\beta + \frac{\varepsilon}{2}}$. The next result gives sufficient conditions for this to happen (Roughly speaking, we put $\beta + \frac{\varepsilon}{2}$ in place of β in Theorem 3.3)

Theorem 3.4 *Let* $\theta \in \Theta_{m,\alpha}$, $m + \alpha > 1$. *Then for every* β, $\beta < \dfrac{(m+\alpha)-1}{2(m+\alpha)+1}$, *there exist* $\varepsilon > 0$ *and* $q > 0$ *(which depend on* β*) such that,*

$$n^\beta \sup_{0 \le t \le 1} | \hat{\theta}_{n^\varepsilon}{}^{n}(t) - \theta(t) | \to 0 \ \ a.s., \ \ as \ \ n \to \infty .$$

The next result is an obvious consequence of the above when θ is assumed to be infinitely differentiable, that is if, $\theta \in \Theta_{m,\alpha}$ for *every* m.

Corollary 3.4 *Let* $\theta \in \Theta_{m,\alpha}$, *for every* m; *then,*
for every $\beta < \dfrac{1}{2}$ *there exist* $\varepsilon > 0$ *and* $q > 0$ *such that*

$$n^\beta \sup_{0 \le t \le 1} | \hat{\theta}_{n^\varepsilon}{}^{n}(t) - \theta(t) | \to 0 \ \ a.s., \ \ as \ \ n \to \infty .$$

Comparison of results: Let $\{f_j\}, j \ge 1$ be an arbitrary but fixed CONS and $\ddot{\theta}_k{}^n$ be the sieve estimator corresponding to the sieve $\{\overline{S}_k, k \ge 1\}$; $\overline{S}_k$ being the k-dimensional subspaces spanned by f_j, $1 \le j \le k$. Without assuming any regularity on θ, Nguyen & Pham (1982) proved that $\ddot{\theta}_{k_n}{}^n$ converges to θ in *probability* in L^2

norm, when $k_n \to \infty$ slower than $\sqrt{n}$. The rate of covergence of $\ddot{\theta}_{k_n}{}^n$, however, was not discussed. We were not able to obtain a result of this type using the method in Section II but, in view of Theorems 3.2 - 3.3 it is clear that in the presence of mild regularity on θ this method works quite well. Next let us consider the consistency result of Kutoyants.

Assume that $\theta \in \Theta_{m,\alpha}$, for some m and α such that $m+\alpha > 0$. Suppose $(x^1{}_t, \ldots x^n{}_t)$ are n i.i.d. copies of $(x_t), 0 \le t \le 1$. For $j = 1, \ldots n$ and $z_n = n^{-\frac{m+\alpha}{2(m+\alpha)+1}}$ define

$$Y_{n,j}(t) = (x_t{}^j)^{-1} \quad \text{if} \quad |x_t{}^j| \ge z_n$$

$$= z_n{}^{-1} \quad \text{otherwise.}$$

Then the kernel estimator of Kutoyants, $\overline{\theta}_n(t)$ say, is defined by

$$\overline{\theta}_n(t) = \frac{1}{n\, a_n} \sum_{j=1}^{n} \int_0^1 K\left(\frac{\tau - t}{a_n}\right) Y_{n,j}(\tau)\, dx^j{}_\tau$$

where $a_n = n^{-\frac{1}{2(m+\alpha)+1}}$ and $K(.)$ is some bounded kernel. The *weak pointwise* consistency of $\overline{\theta}_n(t)$ is proved by showing that, for every $c > 0$ and $d \le 1$,

$$\lim_{n \to \infty} \sup_{c \le t \le d} E\{n^{\frac{(m+\alpha)}{4(m+\alpha)+2}} |\overline{\theta}_n(t) - \theta_t|\}^2 < \infty.$$

From the above it is clear that for every β, $\beta < \frac{(m+\alpha)}{4(m+\alpha)+2}$, there exists a choice of z_n and a_n such that,

$$n^{\beta} |\overline{\theta}_n(t) - \theta_t| \to 0 \text{ as } n \to \infty \text{ in probability.}$$

Let us compare this result with Theorem 3.4. Clearly, Kutoyants's result is more widely applicable than Theorem 3.4; it is applicable for $\theta \in \Theta_{m,\alpha}, m+\alpha > 0$ whereas for Theorem 3.4 we require that $m+\alpha > 1$. However, the possible rate of convergence of $\hat{\theta}_k{}^n$ exceeds the possible rate of convergence of the kernel estimator $\overline{\theta}_n(t)$ when it is assumed that $\theta \in \Theta_{m,\alpha}, m+\alpha > 2$. It is interesting to note that even if θ is assumed to be very smooth i.e. if $\theta \in \Theta_{m,\alpha}$ for every m, the largest possible (limiting) value of β for $\overline{\theta}_n(t)$ is $\frac{1}{4}$ where as for $\hat{\theta}_k{}^n$ it is $\frac{1}{2}$ (Corollary 3.4). It should also be emphasized that the convergence in Theorem 3.4 is in the uniform norm.

Proofs: First we prove Theorem 3.1. In view of (3.3) the statement of the theorem is equivalent to the statement with $\|.\|$ norm replaced by $\|.\|_\theta$ norm. This equivalent statement is proved by verifying the conditions of Corollary 2.1. The verification is done with the help of several lemmas The first two lemmas are technical. Lemma 3.1 bounds the higher moments of the L^2 norm of a square integrable Gaussian process in terms of its second moment and Lemma 3.2 gives bounds on its moment generating function. Both these results are important in their own right and their proofs must be available in the literature but since the auther was unable to find the original sources brief proofs are sketched.

Lemma 3.1 *Let $(Y_t), 0 \le t \le T$ be a zero mean Gaussian process such that*

$$E \int_0^T Y_t^2 \, dt < \infty . \tag{3.13}$$

Then, for all $k \ge 1$,

$$E \left(\int_0^T Y_t^2 \, dt \right)^k < k^k \left[E \left(\int_0^T Y_t^2 \, dt \right) \right]^k .$$

Proof Let $R(t,s) = E\, Y_t Y_s$ and let R be the corresponding covariance operator defined on $L^2[0,T]$. The condition (3.13) implies that $(Y_t) \in L^2[0,T]$ a.s. and that R is a self adjoint, non-negative definite, trace class operator. In fact,

$$tr(R) := Trace\,(R) = \int_0^T R(t,t)\, dt = E \int_0^T Y_t^2 \, dt. \tag{3.14}$$

Since $(Y_t) \in L^2[0,T]$ a.s., we can expand it in terms of any fixed complete orthonormal system (CONS) in $L^2[0,T]$. If the CONS chosen is the one consisting of the eigenfunctions of the covariance operator R, the resulting expansion is known as the Karhunen-Loeve expansion. Let $\{f_j\}$ be the system of eigenfunctions of R and λ_j's be the corresponding eigenvalues then the Karhunen-Loeve expansion of (Y_t) can be written as

$$Y_t = \sum_{j=1}^{\infty} \sqrt{\lambda_j}\, X_j\, f_j(t) \quad a.s. \tag{3.15}$$

where the coefficients (X_j) are i.i.d. $N(0,1)$ random variables.

From the above the following identities are immediate :

$$\int_0^T Y_t^2\, dt = \sum_{j=1}^{\infty} \lambda_j X_j^2 \text{ and} \tag{3.16}$$

$$E \int_0^T Y_t^2\, dt = \sum_{j=1}^{\infty} \lambda_j = tr(R). \tag{3.17}$$

Therefore,

$$E\left(\int_0^T Y_t^2\, dt\right)^k = E\left(\sum_{j=1}^{\infty} \lambda_j X_j^2\right)^k$$

$$= E\left\{\left(\sum_1^{\infty} \lambda_j\right)^k \left[\left(\sum_1^{\infty} \lambda_j\right)^{-1} \sum_{j=1}^{\infty} \lambda_j X_j^2\right]^k\right\}.$$

Applying Jensen's inequality for the quantity in the square bracket we get

$$E\left(\int_0^T Y_t^2\, dt\right)^k \le \left(\sum_1^{\infty} \lambda_j\right)^{k-1} E\left(\sum_{j=1}^{\infty} \lambda_j X_j^{2k}\right)$$

$$= \left(\sum_1^{\infty} \lambda_j\right)^k E X_1^{2k}$$

since X_j's are i.i.d.. Moreover because X_1 is $N(0,1)$, using Stirling's approximation (see Rao (1973) pp 59),

$$E\,X_1^{2k} = \frac{(2k)!}{2^k\,k!} \le k^k .$$

So finally we have

$$E\left(\int_0^T Y_t^2\,dt\right)^k \le k^k \left(\sum_1^\infty \lambda_j\right)^k ;$$

which completes the proof (see (3.17)).

Lemma 3.2 *Assume that the Gaussian process (Y_t) is as in Lemma 3.1 and R is the corresponding covariance operator, also, let $\|R\|$ denote the operator norm of R then,*

i) For $0 \le \alpha < (2\|R\|)^{-1}$,

$$E \exp\left(\alpha \int_0^T Y_t^2\,dt\right) \le \exp\left(\frac{\alpha\, tr(R)}{1-2\alpha \|R\|}\right). \tag{3.18}$$

If $0 \le \alpha < (2\, tr(R))^{-1}$ the following weaker version of the above inequality can be useful.

$$E \exp\left(\alpha \int_0^T Y_t^2\,dt\right) \le \exp\left(\frac{\alpha\, tr(R)}{1-2\alpha\, tr(R)}\right). \tag{3.19}$$

ii)

$$E \exp\left(-\int_0^T Y_t^2\,dt\right) \le \exp\left(\frac{-\,tr(R)}{1+2\|R\|}\right)$$

$$\le \exp\left(\frac{-\,tr(R)}{1+2tr(R)}\right). \tag{3.20}$$

Note that $tr(R) = \int_0^T EY_t^2\,dt$.

Proof From (3.15)

$$\int_0^T Y_t^2\,dt = \sum_{j=1}^\infty \lambda_j X_j^2$$

where λ_j's are eigenvalues of R and X_j's are i.i.d. $N(0,1)$. Note that if X is a $N(0,1)$ variable then, for $\beta < \frac{1}{2}$, $E \exp(\beta X^2) = (1-2\beta)^{-\frac{1}{2}}$. Therefore, using the fact that $\|R\| \geq \lambda_j$ for all j and $\alpha < (2\|R\|)^{-1}$ we get

$$E \exp(\alpha \lambda_j X_j^2) = (1-2\alpha \lambda_j)^{-\frac{1}{2}}.$$

Hence

$$E \exp\left(\alpha \int_0^T Y_t^2 \, dt\right) = \prod_{j=1}^{\infty} (1-2\alpha \lambda_j)^{-\frac{1}{2}}$$

$$= \exp\left(-\frac{1}{2} \sum_1^{\infty} \ln(1-2\alpha \lambda_j)\right)$$

$$= \exp\left(\frac{1}{2} \sum_1^{\infty} \ln\left(1+\frac{2\alpha \lambda_j}{1-2\alpha \lambda_j}\right)\right)$$

$$\leq \exp\left(\sum_1^{\infty} \frac{\alpha \lambda_j}{1-2\alpha \lambda_j}\right)$$

since $\ln(1+x) < x$ for $x > 0$.

So

$$E \exp\left(\alpha \int_0^T Y_t^2 \, dt\right) \leq \exp\left(\alpha \sum_1^{\infty} \frac{\lambda_j}{1-2\alpha \lambda_j}\right).$$

The conclusions in (i) follow because of the following observation

$$\frac{1}{1-2\alpha \lambda_j} < \frac{1}{1-2\alpha \|R\|} < \frac{1}{1-2\alpha \, tr(R)}.$$

Proof of (ii) is also almost identical.

Now, following the notation in Section II, let us define random functions $Z_k^n(u)$ with domains U_k^n as follows :

For $n \geq 1$, $k \geq 1$ and $\beta \geq 0$ let

$$U_k^n = n^{\beta} (S_k - \underline{\theta}_k) \tag{3.21}$$

and

$$Z_k^n(u) = l_n(\underline{\theta}_k + n^{-\beta} u) - l_n(\underline{\theta}_k) \tag{3.22}$$

where $\underline{\theta}_k$ is the projection of θ in S_k and $l_n(\theta)$ is the *loglikelihood function* at θ when the sample size is n.

From (3.4) and the fact that $P^n{}_\theta$ is a product measure the following expression for the loglikelihood function is obtained :

$$l_n(\theta, \omega^n) = \ln \frac{dP^n{}_\theta}{dP^n{}_0} (\omega^n)$$

$$= \sum_{i=1}^{n} \int_0^1 \theta_t \, \omega^i{}_t \, d\omega^i{}_t - \frac{1}{2} \int_0^1 \theta_t{}^2 \, (\omega^i{}_t)^2 \, dt. \tag{3.23}$$

The next lemma verifies the first condition of Corollary 2.1.

Lemma 3.3 *For $u \in U_k^n$,*

$$E \, | Z_k^n(u) - Z_k^n(v) |^{2k} \leq B_k^n \, || u - v ||_\theta^{2k}$$

where

$$B_k^n = n^{2k(1-\beta)+1} \, k^{3k} \, ||\theta||_\theta^{2k} \, (16 M_\theta)^{2k} .$$

Proof From (3.23),

$$Z_k^n(u) - Z_k^n(v) = l_n(\underline{\theta}_k + n^{-\beta} u) - l_n(\underline{\theta}_k + n^{-\beta} v)$$

$$= n^{-\beta} \sum_{i=1}^{n} \{ \int_0^1 (u_t - v_t) \, \omega^i{}_t \, d\omega^i{}_t$$

$$- \frac{1}{2} \int_0^1 [(\underline{\theta}_k + n^{-\beta} u)_t{}^2 - (\underline{\theta}_k + n^{-\beta} v)_t{}^2] \, (\omega^i{}_t)^2 \, dt \, \}.$$

Hence, using the elementary inequality, $(\sum_1^m a_i)^k \leq m^k (\sum_1^m a_i{}^k)$, for positive numbers a_i repeatedly, and the fact that $(\omega^i{}_t)$ are i.i.d., we can bound $E \mid Z_k{}^n (u) - Z_k{}^n (v) \mid^{2k}$ by the following expression:

$$n\, 2^{2k}\, n^{2k(1-\beta)}\, E\Big(\int_0^1 (u_t - v_t)\, \omega_t\, d\omega_t \Big)^{2k}$$

$$+ n\, n^{2k}\, E\Big(\int_0^1 [\, (\underline{\theta}_k + n^{-\beta} u)_t{}^2 - (\underline{\theta}_k + n^{-\beta} v)_t{}^2 \,]\, (\omega_t)^2\, dt \Big)^{2k} . \qquad (3.24)$$

We will bound each term in (3.24) separately. First consider $E\Big(\int_0^1 (u_t - v_t)\, \omega_t\, d\omega_t \Big)^{2k}$:

Note that, under P_θ, (ω_t) satisfies the following SDE :

$$d\omega_t = \theta_t\, \omega_t\, dt + dW_t\, (\omega) \quad , \omega_0 = X_0$$

where $W_t\, (\omega)$ is the standard Wiener process under P_θ, and X_0 is a $N(0, \sigma_0{}^2)$ r.v. independent of $W_t\, (\omega)$. Therefore

$$\int_0^1 (u_t - v_t)\, \omega_t\, d\omega_t = \int_0^1 (u_t - v_t)\, \theta_t\, \omega^2{}_t\, dt$$

$$+ \int_0^1 (u_t - v_t)\, \omega_t\, dW_t(\omega).$$

Hence $E\Big(\int_0^1 (u_t - v_t)\, \omega_t\, d\omega_t \Big)^{2k}$ is bounded by

$$2^{2k}\, \Big\{ E\Big(\int_0^1 (u_t - v_t)\, \theta_t\, \omega^2{}_t\, dt \Big)^{2k}$$

$$+ E\Big(\int_0^1 (u_t - v_t)\, \omega_t\, dW_t\, (\omega) \Big)^{2k} \Big\}. \qquad (3.25)$$

The first term in the bracket above can be bounded by the repeated application of

Cauchy-Schwartz inequality and finally, of Lemma 3.1 by

$$\{ k^{2k} \; ||u-v||_\theta^{4k} \; ||\theta||_\theta^{4k} \}^{\frac{1}{2}}.$$

Thus,

$$E\left(\int_0^1 (u_t - v_t)\, \theta_t\, \omega^2_t\, dt \right)^{2k} \leq k^k \; ||u-v||_\theta^{2k} \; ||\theta||_\theta^{2k}. \tag{3.26}$$

Next consider $E\left(\int_0^1 (u_t - v_t)\, \omega_t\, dW_t\, (\omega) \right)^{2k}$ (the second term in (3.25)); it can be bounded by

$$(8k)^{2k}\, E\left(\int_0^1 (u_t - v_t)^2\; \omega^2_t\, dt \right)^k .$$

The above bound follows from the Burkholder - Davis - Gundy (BDG) inequality stated here for the reader's convenience (see Dellacherie and Meyer (1982), pp 287 where a general inequality is given).

BDG inequality *Let* (M_t), $0 \leq t \leq T$ *be a square integrable martingale with squared variation process* $< M >_t$ *then,*

$$E \sup_{0 \leq t \leq T} |M_t|^{2p} \leq (8p)^{2p}\, E < M >_T^{\,p}$$

for all $p \geq 1$. Therefore, using Lemma 3.1, we get

$$E\left(\int_0^1 (u_t - v_t)\, \omega_t\, dW_t\, (\omega) \right)^{2k} \leq 8^{2k}\, k^{3k} \; ||u-v||_\theta^{2k}. \tag{3.27}$$

Hence, in view of (3.25), (3.26) and (3.27), the first term in (3.24) can be bounded by

$$||u-v||_\theta^{2k} \; n^{2k(1-\beta)+1} \; k^{3k} \; 8^{4k} \; ||\theta||_\theta^{2k}.$$

The second term in (3.24) can be bounded in the same fashion and then combining these two bounds the lemma follows.

The second condition of Corollary 2.1 is verified in Lemma 3.6, the following two lemmas are preparatory.

Lemma 3.4

$$E\exp\left(\frac{n^{-\beta}}{4}\int_0^1 u_t\omega_t\,d\omega_t-\frac{1}{8}\int_0^1 [(\underline{\theta}_k+n^{-\beta}u)_t{}^2-(\underline{\theta}_k)_t{}^2]\omega_t{}^2dt\right)$$

$$\leq \left(E\exp\left(-\frac{1}{4}\int_0^1 [n^{-\beta}u-2(\theta-\underline{\theta}_k)]_t{}^2\omega_t{}^2dt\right)\right)^{\frac{1}{4}}.\left(E\exp\left(\int_0^1 (\theta-\underline{\theta}_k)_t{}^2\omega_t{}^2dt\right)\right)^{\frac{1}{4}}.$$

Proof Recall that, under P_θ the paths (ω_t) satisfy the following SDE:

$$d\omega_t = \theta_t\ \omega_t\ dt + dW_t\ (\omega)\quad ,\ \omega_0 = X_0.$$

Hence, writing $\theta_t = (\theta-\underline{\theta}_k)_t+(\underline{\theta}_k)_t$ we get

$$\frac{n^{-\beta}}{4}\int_0^1 u_t\omega_t\,d\omega_t-\frac{1}{8}\int_0^1 [(\underline{\theta}_k+n^{-\beta}u)_t{}^2-(\underline{\theta}_k)_t{}^2]\omega_t{}^2dt$$

$$= \frac{n^{-\beta}}{4}\int_0^1 u_t\omega_t\,dW_t(\omega)+\frac{n^{-\beta}}{4}\int_0^1 u_t(\theta-\underline{\theta}_k)_t\omega_t{}^2dt-\frac{n^{-2\beta}}{8}\int_0^1 u_t{}^2\omega_t{}^2dt.$$

Moreover,

$$E_\theta\ \{\exp\left(\frac{n^{-\beta}}{2}\int_0^1 u_t\omega_t\,dW_t(\omega)-\frac{n^{-2\beta}}{8}\int_0^1 u_t{}^2\omega_t{}^2dt\right)\} = 1$$

since the expression inside the bracket is in fact $\dfrac{dP_{\theta+\frac{n^{-\beta}}{2}u}}{dP_\theta}(\omega)$ i.e. it is a density.

Therefore, applying the Cauchy–Schwartz inequality,

$$E\exp\left(\frac{n^{-\beta}}{4}\int_0^1 u_t\omega_t\,d\omega_t-\frac{1}{8}\int_0^1 [(\underline{\theta}_k+n^{-\beta}u)_t{}^2-(\underline{\theta}_k)_t{}^2]\omega_t{}^2dt\right)$$

$$\leq \left(E\exp\left(\frac{n^{-\beta}}{2}\int_0^1 u_t\omega_t\,dW_t(\omega)-\frac{n^{-2\beta}}{8}\int_0^1 u_t{}^2\omega_t{}^2dt\right)\right)^{\frac{1}{2}}$$

$$\times \left(E\exp\left(-\frac{n^{-2\beta}}{8}\int_0^1 u_t{}^2\omega_t{}^2dt+\frac{n^{-\beta}}{2}\int_0^1 u_t(\theta-\underline{\theta}_k)_t\omega_t{}^2dt\right)\right)^{\frac{1}{2}}$$

$$= 1\,.\,\Big(E\exp\Big(-\frac{1}{8}\int_0^1 [n^{-\beta}u_t - 2(\theta-\underline{\theta}_k)_t]^2\omega_t^{\,2}dt + \frac{1}{2}\int_0^1 (\theta-\underline{\theta}_k)_t^{\,2}\omega_t^{\,2}dt\Big)\Big)^{\frac{1}{2}}.$$

The conclusion of the lemma follows if we apply the Cauchy-Schwartz inequality again.

Lemma 3.5 *For* $u \in U^n{}_k$ *and* k *large so that*

$$1 - 2\,\|\theta-\underline{\theta}_k\|_\theta^2 > 0,$$

$$E\exp\Big(\frac{1}{4}Z^n{}_k(u)\Big) \le \exp\Big(\frac{-n}{16}\,\frac{\|n^{-\beta}u - 2(\theta-\underline{\theta}_k)\|_\theta^2}{1+\|n^{-\beta}u - 2(\theta-\underline{\theta}_k)\|_\theta^2}\Big)$$

$$\times\ \exp\Big(\frac{n\,\|\theta-\underline{\theta}_k\|_\theta^2}{4(1-2\|\theta-\underline{\theta}_k\|_\theta^2)}\Big).$$

Proof Note that $Z_k{}^n$ can be written as

$$Z_k{}^n(u) = \sum_{i=1}^{n} n^{-\beta}\int_0^1 u_t\,\omega^i{}_t\,d\omega^i{}_t - \frac{1}{2}\int_0^1 \big([\underline{\theta}_k + n^{-\beta}u]_t^{\,2} - [\underline{\theta}_k]_t^{\,2}\big)(\omega^i{}_t)^2 dt$$

where $(\omega^i{}_t), 0 \le t \le 1$, i = 1, 2 .. n are i.i.d processes. From this independence and Lemma 3.4 we get,

$$E\exp\Big(\frac{1}{4}Z^n{}_k(u)\Big)$$

$$\le \Big(E\exp\Big(-\frac{1}{4}\int_0^1 [n^{-\beta}u_t - 2(\theta-\underline{\theta}_k)_t]^2\omega_t^{\,2}dt\Big)\Big)^{\frac{n}{4}}\Big(E\exp\Big(\int_0^1 (\theta-\underline{\theta}_k)_t^{\,2}\omega_t^{\,2}dt\Big)\Big)^{\frac{n}{4}}.$$

To obtain the lemma we now apply Lemma 3.2 (ii) to the first term and (i) to the last term of the product in the RHS above.

Let us bound the expression in the last lemma by a slightly simpler expression.

First note that since $\left| \|x\|_\theta - \|y\|_\theta \right| \le \|x-y\|_\theta$,

$$\frac{-n}{16}\frac{\|n^{-\beta}u-2(\theta-\underline{\theta}_k)\|_\theta^2}{1+\|n^{-\beta}u-2(\theta-\underline{\theta}_k)\|_\theta^2} \le \frac{-n}{16}\frac{\left|n^{-\beta}\|u\|_\theta-2\|\theta-\underline{\theta}_k\|_\theta\right|^2}{1+\|n^{-\beta}u-2(\theta-\underline{\theta}_k)\|_\theta^2}.$$

Furthermore since $u \in U^n{}_k = n^\beta(S_k-\underline{\theta}_k)$,

$$\|u\|_\theta \le 2\sqrt{M_\theta}\,k^q\,n^\beta$$

and, for large k, $k > k_\theta$,

$$1+\|n^{-\beta}u-2(\theta-\underline{\theta}_k)\|_\theta^2 \le C_\theta\,k^{2q}$$

(since $\|\theta-\underline{\theta}_k\|^2 \to 0$ as $k \to \infty$). Therefore,

$$\frac{-n}{16}\frac{\|n^{-\beta}u-2(\theta-\underline{\theta}_k)\|_\theta^2}{1+\|n^{-\beta}u-2(\theta-\underline{\theta}_k)\|_\theta^2} \le \frac{-nC_\theta}{k^{2q}}\left|n^{-\beta}\|u\|_\theta-2\|\theta-\underline{\theta}_k\|_\theta\right|^2 \tag{3.28}$$

Also for large k, $1-2\|\theta-\underline{\theta}_k\|_\theta^2 > \frac{1}{2}$ hence

$$\frac{n\|\theta-\underline{\theta}_k\|_\theta^2}{1-2\|\theta-\underline{\theta}_k\|_\theta^2} \le 2n\|\theta-\underline{\theta}_k\|_\theta^2. \tag{3.29}$$

Therefore, from (3.28), (3.29) and Lemma 3.5 we get,

$$E\exp(\frac{1}{4}Z^n{}_k(u)) \le \exp(g^n{}_k(\|u\|_\theta)) \tag{3.30}$$

where

$$g^n{}_k(x) = \frac{-nC_\theta}{k^{2q}}\left|n^{-\beta}x-2\|\theta-\underline{\theta}_k\|_\theta\right|^2 - 2n\|\theta-\underline{\theta}_k\|_\theta^2.$$

The conclusion in (3.30) is stated as the next lemma (which verifies the second condition of Corollary 2.1)

Lemma 3.6 *For $u \in U^n{}_k$ and $k \ge k_\theta$*

$$E\exp(\frac{1}{4}Z^n{}_k(u)) < \exp(g^n{}_k(\|u\|_\theta))$$

where

$$g^n{}_k(x) = \frac{-nC_\theta}{k^{2q}} \,|n^{-\beta}x - 2\,||\theta - \underline{\theta}_k\,||_\theta\,|^2 - 2n\,||\theta - \underline{\theta}_k\,||_\theta^2.$$

It is easy to check the remaining conditions of Corollary 2.1. Note that, here,

1) $m = \alpha = 2k$, $p = 0$
2) $B^n{}_k \leq n^{2k(1-\beta)+1}\, k^{k(3+2q)}\, C_\theta{}^{2k}$ and
3) $g^n{}_k(x) = C_\theta n^\delta k^{-\mu}\,|x - a(n,k)|^\nu - t(n,k)$.

where

$\delta = 1-2\beta$, $\mu = 2q$, $\nu = 2$,

$a(n,k) = 2n^\beta\,||\theta - \underline{\theta}_k\,||_\theta$ and

$t(n,k) = 2n\,||\theta - \underline{\theta}_k\,||_\theta^2$.

Also recall that, by assumption, $\theta \in \Theta_\varepsilon$ and $1-2\beta-2\varepsilon-2\varepsilon q > 0$. Therefore if we take $s = \varepsilon > 0$ then

$$\delta - \mu s - 2s = 1-2\beta-2\varepsilon-2\varepsilon q > 0\,.$$

Now let us see if we can show

i) $a(n,[n^\varepsilon]) = 2\,n^\beta\,||\theta - \underline{\theta}_{n^\varepsilon}||_\theta \;\to\; 0$ and

ii) $\dfrac{t(n,[n^\varepsilon])}{[n^\varepsilon]} = 4n^{1-\varepsilon}\,||\theta - \underline{\theta}_{n^\varepsilon}||_\theta^2 \;\to\; 0$ as $n \;\to\; \infty$.

(ii) is easy to see because $\theta \in \Theta_\varepsilon$ implies that

$$n^{1-\varepsilon}\,||\theta - P_{n^\varepsilon}\theta\,||^2 \;\to\; 0$$

which implies that

$$n^{1-\varepsilon}\,||\theta - \underline{\theta}_{n^\varepsilon}||_\theta^2 \;\to\; 0;$$

which is (ii). It also shows (i) since $1-\varepsilon > 2\beta$. This concludes the proof of Theorem 3.1.

Now we show that Theorems 3.2 - 3.5 follow from Theorem 3.1.

Proof of Theorem 3.2 Obvious from Theorem3.1, the Borel-Cantelli lemma and the following:

$$n^{\beta} \, || \, \hat{\theta}_{n^{\varepsilon}}{}^{n} - \theta \, || \leq n^{\beta} \, || \, \hat{\theta}_{n^{\varepsilon}}{}^{n} - \underline{\theta}_{n^{\varepsilon}} \, || + n^{\beta} \, || \, \theta - \underline{\theta}_{n^{\varepsilon}} \, ||.$$

Before starting the proof of Theorem 3.3 let us review some standard facts from Fourier theory. These are stated as propositions. From now on, $\{ f_j \}$ will always denote the CONS defined in (3.9) i.e. the one with the trigonometric functions. Also recall that, for $m \geq 0$ and $0 \leq \alpha \leq 1$, $\Theta_{m,\alpha}$ is the space of m-times continuously differentiable functions with m-th derivative Lipschitz continuous of order α.

Proposition 3.1 *Let* $\theta \in \Theta_{m,\alpha}$. *Then there exists* $M = M(m,\alpha,\theta)$ *such that*

$$|| \, \theta - P_n \, \theta \, ||^2 = \int_0^1 [\, \theta_t - (P_n \, \theta)_t \,]^2 \, dt \leq M \, n^{-2(m+\alpha)}.$$

The above inequality is a simple consequence of a result called Jackson's inequality [see Zygmund (1949), Theorem 13.6, pp 115, vol I.] and the fact that $P_n \, \theta$ is the projection of θ on the linear span of $\{ f_j \}$, $1 \leq j \leq n$;

Corollary 3.5

$$\Theta_{m,\alpha} \subset \Theta_{\varepsilon} \text{ for } \textit{all } \varepsilon > \frac{1}{1+2(m+\alpha)}$$

Proof We have to show that $\theta \in \Theta_{m,\alpha}$, $\varepsilon > \dfrac{1}{1+2(m+\alpha)}$, implies that

$$n^{1-\varepsilon} \, || \, \theta - P_{n^{\varepsilon}} \theta \, ||^2 \to 0 \text{ as n } \to \infty$$

From the last proposition,

$$n^{1-\varepsilon} \, || \, \theta - P_{n^\varepsilon} \theta \, ||^2 \leq M \; n^{1-\varepsilon-2\varepsilon(m+\alpha)}.$$

The RHS above tends to zero if $1-\varepsilon-2\varepsilon(m+\alpha) < 0$ i.e. if $\varepsilon > \dfrac{1}{1+2(m+\alpha)}$, which is the required conclusion.

The next proposition shows that the Fourier coefficients of functions in $\Theta_{m,\alpha}$ enjoy nice convergence properties (see pp 240, Zygmund (1949) vol II and Theorem 5, pp71, vol I).

Proposition 3.2

i) If $\theta \in \Theta_{m,\alpha}$, $m + \alpha > \dfrac{1}{2}$ *then*

$$\sup_{0 \leq t \leq 1} | \theta(t) | < \sqrt{2} \sum_{1}^{\infty} | \theta_j | < \infty .$$

where θ_j*'s are the Fourier coefficients of* θ.

ii) $\theta \in \Theta_{m,\alpha}$, *implies that*

$$| \theta_j | - O \, (j^{-m-\alpha})$$

Proof of Theorem 3.3 Follows from Theoerem 3.2 and Corollary 3.5 (because for every $\beta < \dfrac{(m+\alpha)-\frac{1}{2}}{2(m+\alpha)+1}$ we can find an ε and $q > 0$ such that $\Theta_{m,\alpha} \subset \Theta_\varepsilon$ and $1-2\beta-2\varepsilon-2\varepsilon q > 0$).

Proof of Theorem 3.4 : This result follows easily from Theorem 3.2, Corollary 3.5 and Proposition 3.2. To see this first note that

$$\sup_{0 \leq t \leq 1} | \hat{\theta}_{n^\varepsilon}{}^n (t) - \theta_t | \leq \sqrt{2} \, \{ \sum_{j=1}^{n^\varepsilon} | \hat{\theta}_j - \theta_j | + \sum_{n^\varepsilon+1}^{\infty} | \theta_j | \}.$$

Hence the result is proved if we show that

a) $n^\beta \displaystyle\sum_{j=1}^{n^\varepsilon} | \hat{\theta}_j{}^n - \theta_j | \to 0$ and

b) $n^{\beta} \sum_{n^{\varepsilon}+1}^{\infty} |\theta_j| \to 0$ as $n \to \infty$.

First let us show (a). Using the Cauchy–Schwartz inequality,

$$n^{\beta} \sum_{j=1}^{n^{\varepsilon}} |\hat{\theta}_j^{\,n} - \theta_j| \le n^{\beta} n^{\frac{\varepsilon}{2}} \left(\sum_{j=1}^{n^{\varepsilon}} |\hat{\theta}_j^{\,n} - \theta_j|^2 \right)^{\frac{1}{2}}$$

Therefore (a) is proved if we show that

$$n^{\beta + \frac{\varepsilon}{2}} \, || \hat{\theta}_{n^{\varepsilon}}^{\,n} - \underline{\theta}_{n^{\varepsilon}} || \to 0 \ a.s. \text{ as } n \to \infty .$$

But this follows from Theorem 3.2 because for $\beta < \dfrac{(m+\alpha)-1}{2(m+\alpha)+1}$ we can find an ε and q such that $\Theta_{m,\alpha} \subset \Theta_{\varepsilon}$ and $1 - 2(\beta + \dfrac{\varepsilon}{2}) - 2\varepsilon - 2\varepsilon q > 0$. Now consider (b). From Proposition 3.2 (ii), there exists a C such that for large n,

$$n^{\beta} \sum_{n^{\varepsilon}+1}^{\infty} |\theta_j| \le C \, n^{\beta} \sum_{n^{\varepsilon}+1}^{\infty} j^{-m-\alpha}.$$

Hence

$$n^{\beta} \sum_{n^{\varepsilon}+1}^{\infty} |\theta_j| \le C \, n^{\beta} (n^{\varepsilon})^{1-m-\alpha} \le C \, n^{\beta + \varepsilon(1-m-\alpha)}.$$

The RHS above tends to zero if $\beta + \varepsilon(1-m-\alpha) < 0$. Again, for $\beta < \dfrac{(m+\alpha)-1}{2(m+\alpha)+1}$ one can find an ε such that $\Theta_{m,\alpha} \subset \Theta_{\varepsilon}$ and $\beta + \varepsilon(1-m-\alpha) < 0$; and thus (b) is proved.

IV Intensity estimation of Poisson process:

Let $(N_t), 0 \le t \le 1$ be a Poisson process with intensity function $\theta = (\theta_t), 0 \le t \le 1$; i.e., $M_t = N_t - \int_0^t \theta_s \, ds$ is a square integrable martingale with square variation process

$$< M >_t = \int_0^t \theta_s \, ds. \tag{4.1}$$

The deterministic function θ is assumed to be positive, bounded and bounded away from zero. The aim is to estimate θ using a sample of n independent trajectories of $(N_t), 0 \leq t \leq 1$. This problem is treated as one of the examples in a paper by A.F. Karr (1987). There he proposes a sieve estimator (Histogram sieve estimator) which is shown to be strongly consistent in the L_1 norm. The rate of convergence of this estimator is not discussed. As in the case of Section III, for this example, the procedure suggested in Section II yields sieve estimators possessing good consistency properties. A detailed discussion of the rate of convergence of these estimators is given after developing the necessary notation. All the proofs are given at the end.

Let $D = D\,[0, 1]$ be the space of right continuous, left limit functions on $[0, 1]$, endowed with the Skorokhod topology; $\mathbf{D}$ will denote the σ-field of Borel sets in D.

Let

$$\Theta = \{ f \in L^2([0,1], dt) : f \geq 0, \sup_{0 \leq t \leq 1} f(t) < \infty \text{ and } \inf_{0 \leq t \leq 1} f(t) > 0 \}. \tag{4.2}$$

Θ will serve as the parameter space for our problem.

Suppose, for $\theta \in \Theta$, P_θ is the measure induced on $(D, \mathbf{D})$ by the paths of a Poisson process with intensity θ; then if $\theta, \mu \in \Theta$, it is well-known that (see Karr (1987), pp 475), P_θ and P_μ are mutually absolutely continuous and, for $\omega \in D$,

$$\frac{dP_\theta}{dP_\mu}(\omega) = \exp\left(\int_0^1 (\mu_t - \theta_t)\, dt + \int_0^1 \ln\left(\frac{\theta_t}{\mu_t}\right) d\omega_t \right). \tag{4.3}$$

In what follows, $(D^n, \mathbf{D}^n)$ will denote the n-fold product of $(D, \mathbf{D})$, ω^n a typical point in D^n and $P^n{}_\theta$ the n-fold product of P_θ.

Following the method from Sec II, a sieve estimator $\hat{\theta}^n{}_k$ is proposed to estimate θ. The loglikelihood function, from (4.3), can be written as,

$$l\,(\theta, \omega^n) = \ln \frac{dP^n{}_\theta}{dP^n{}_1}(\omega^n)$$

$$= n \int_0^1 (1 - \theta_t)\, dt + \sum_{i=1}^{n} \int_0^1 \ln(\theta_t)\, d\omega^i{}_t . \qquad (4.4)$$

(μ is taken to be identically equal to 1 and ω^i are components of ω^n; under $P^n{}_\theta$ they are i.i.d. Poisson processes with intensity θ.)

Two different types of sieves S_k are used.

I) For $k \geq 1$ and $q > 0$ let

$$S_k = \{ \theta \in \Theta : \theta = \sum_{j=1}^{k} \theta_j f_j,\ k^{-q} \leq \theta_t \leq k^q \text{ for all t } \} \qquad (4.5)$$

where for $1 \leq j \leq k$, $f_j(t) = 1_{[\frac{j-1}{k}, \frac{j}{k})}(t)$.

(This sieve is some times referred to as the *Histogram sieve*.)

II)

$$S_k = \{ \theta \in \Theta : \theta = \sum_{j=1}^{k} \theta_j f_j,\ k^{-q} \leq \theta_t \leq k^q \text{ for all t } \}$$

where for $j \geq 1$, f_j's are defined as follows :

Take $f_1 \equiv 1$ and for $j \geq 2$

$$f_j\,(t) = \sqrt{2} \cos(\pi j t) \quad \text{for even j's and}$$

$$= \sqrt{2} \sin[\pi (j+1) t] \quad \text{for odd j's.} \qquad (4.6)$$

The results in this section are almost identical to the results in the last section. To obtain consistency results, using the method from Section II, we need some additional "smoothness" assumptions on the Fourier coeficients of θ. Indeed, in order for the the main consistency result, Theorem 4.1, of this section to be applicable, we require that $\theta \in W_{\varepsilon,q}$ for some $\varepsilon > 0$ and $q > 0$ such that

$1-2\varepsilon q-2\varepsilon > 0$, the class $W_{\varepsilon,q}$ being defined as follows:

$$W_{\varepsilon,q} = \{ \theta \in \Theta : n^{1+2\varepsilon q-\varepsilon} \, || P_{n^\varepsilon}\theta - \theta || \to 0 \} \tag{4.7}$$

where $P_m\theta$ is the projection of θ on the span of $\{f_j, 1 \le j \le m\}$, f_j's are as in (2.4) or (2.5) and $||.||$ denotes the L^2-norm.

The class $W_{\varepsilon,q}$ corresponds to the class Θ_ε of the last section and in this case also if the f_j's used in defining $W_{\varepsilon,q}$ are the trigonometric functions given in (4.6), it is easy to show that,

$$\Theta_{m,\alpha} \subset W_{\varepsilon,q} \quad \text{for} \quad \varepsilon(1+(m+\alpha)-2q) > 1. \tag{4.8}$$

Recall that $\Theta_{m,\alpha}$ is the class of m-times continuously differentiable functions with m-th derivative Lipschitz continuous of order α.

We state the results of this section without any further discussion, the relations among different results here being exactly identical to the relations among the corresponding results in the last section.

For the next two results $\hat{\theta}_k{}^n$ can correspond to either of the two sieves given in (4.5) and (4.6).

Theorem 4.1 *Assume that the true parameter* θ *belongs to* $W_{\varepsilon,q}$ *for some* $\varepsilon > 0$, *and* $q > 0$ *where* $1-2\varepsilon-2\varepsilon q > 0$. *Choose* $\beta \ge 0$ *such that*

$$b = 1 - 2\beta - \varepsilon - 2\varepsilon q > \varepsilon > 0.$$

Then, for all n *sufficiently large and* $h > 0$,

$$P^n{}_\theta \{ n^\beta \, || \hat{\theta}_{n^\varepsilon}{}^n - \underline{\theta}_{n^\varepsilon} || > h \} \le \exp(-C\, n^b h^2)$$

where $C > 0$, *depends on* θ *alone.*

Theorem 4.2 *Assume that* $\theta \in W_{\varepsilon,q}$ *for some* ε, *and* $q > 0$ *where* $1-2\varepsilon-2\varepsilon q > 0$. *Choose* $\beta \ge 0$ *such that*

$$1 - 2\beta - \varepsilon - 2\varepsilon q > \varepsilon > 0.$$

Then,

$$n^{\beta} \, || \, \hat{\theta}_{n^{\varepsilon}}{}^{n} - \theta \, || \to 0 \ a.s. \ as \ n \to \infty .$$

For the following results it is assumed that the CONS $\{ f_j \}$ used in defining the sieve S_k is the one given in (4.6) (i.e. consisting of trigonometric functions).

Theorem 4.3 *Let* $\theta \in \Theta_{m,\alpha}$, $m + \alpha > 1$. *Then, for every* $\beta < \dfrac{(m+\alpha)-1}{2(m+\alpha)+2}$, *there exist* $\varepsilon > 0$ *and* $q > 0$ *such that*

$$n^{\beta} \, || \, \hat{\theta}_{n^{\varepsilon}}{}^{n} - \theta \, || \to 0 \ \text{a.s. as} \ n \to \infty .$$

Theorem 4.4 *Let* $\theta \in \Theta_{m,\alpha}$, $m + \alpha > 2$. *Then, for every* $\beta < \dfrac{(m+\alpha)-2}{2(m+\alpha)+2}$, *there exist* $\varepsilon > 0$ *and* $q > 0$ *such that*

$$n^{\beta} \sup_{0 \le t \le 1} | \, \hat{\theta}_{n^{\varepsilon}}{}^{n}(t) - \theta(t) \, | \to 0 \ \text{a.s. as} \ n \to \infty .$$

Corollary 4.1 *Let* $\theta \in \Theta_{m,\alpha}$, *for every* m; *then,*

for every $\beta < \dfrac{1}{2}$ *there exist* $\varepsilon > 0$ *and* $q > 0$ *such that*

$$n^{\beta} \sup_{0 \le t \le 1} | \, \hat{\theta}_{n^{\varepsilon}}{}^{n}(t) - \theta(t) \, | \to 0 \ a.s., \ as \ n \to \infty .$$

Proof of Theorem 4.1 : This theorem will be proved using Corollary 2.1. The conditions of this corollary are verified using several lemmas. The first lemma, Lemma 4.1, is a simple result from point process theory. It can be easily proved first by verifying it for simple functions and then extending it to more general functions using the standard techniques.

Lemma 4.1 *Suppose (N_t) , $0 \leq t \leq 1$ is a Poisson process with bounded intensity (θ_t) then, for all f , bounded and deterministic,*

$$E \exp \left(\int_0^1 f_s \, dN_s \right) = \exp \left(- \int_0^1 (1 - \exp (f_s)) \, \theta_s \, ds \right).$$

Let us recall the definition of the sieve S_k. For $k \geq 1$ and $q > 0$,

$$S_k = \{ \theta \in \Theta : \theta = \sum_{j=1}^{k} \theta_j f_j, \ k^{-q} \leq \theta(t) \leq k^q, \ t \in [0,1] \}$$

where $\{ f_j \}$, $1 \leq j \leq k$; is a set of independent vectors given in either (4.5) or in (4.6).

For $\theta \in \Theta$, let $\underline{\theta}_k$ denote its projection in S_k. Recall from (4.4) that, the loglikelihood function

$$l(\theta, \omega^n) = n \int_0^1 (1 - \theta_t) \, dt + \sum_{i=1}^{n} \int_0^1 \ln(\theta_t) \, d\omega_t^i \tag{4.9}$$

where under $P^n{}_\theta$, (ω_t^i), $0 \leq t \leq 1$ are i.i.d. Poisson processes with common intensity (θ_t), $0 \leq t \leq 1$.

As usual, let us define random functions $Z^n{}_k(u)$ with domains $U_k{}^n$ as follows :

For $n \geq 1$ and $k \geq 1$, let

$$U_k{}^n = n^\beta (S_k - \underline{\theta}_k) \tag{4.10}$$

and for $u \in U_k{}^n$,

$$Z_k{}^n(u) = l(\underline{\theta}_k + n^{-\beta} u) - l(\underline{\theta}_k). \tag{4.11}$$

Let $g_k{}^n : [0,\infty) \rightarrow R^1$ be given by,

$$g_k{}^n(x) = C \, n^{1-2\beta} k^{-2q} x^2 - n \, k^{2q} \, || \theta - \underline{\theta}_k \, || \tag{4.12}$$

where $C > 0$ depends on θ alone.

Lemma 4.2 *For $u \in U_k{}^n$,*

$$E \exp\left(\frac{1}{2} Z_k{}^n (u) \right) \leq \exp\left(- g_k{}^n (\|u\|) \right).$$

Proof : In view of (4.9) and (4.11),

$$Z_k{}^n (u) = -n^{1-\beta} \int_0^1 u_t \, dt + \sum_1^n \ln \left(\frac{\underline{\theta}_k + n^{-\beta} u}{\underline{\theta}_k} \right)_t d\omega^i{}_t .$$

Recall that, under $P^n{}_\theta$, $(\omega_t{}^i)$, $0 \leq t \leq 1$ are i.i.d. Poisson processes with common intensity (θ_t), $0 \leq t \leq 1$. Hence,

$$E \exp\left(\frac{1}{2} Z_k{}^n (u) \right) = \exp\left(- \frac{n^{1-\beta}}{2} \int_0^1 u_t \, dt \right)$$

$$\cdot \prod_1^n E \exp\left(\frac{1}{2} \ln \left(\frac{\underline{\theta}_k + n^{-\beta} u}{\underline{\theta}_k} \right)_t d\omega^i{}_t \right).$$

Setting $f_s = \frac{1}{2} \ln \left(\frac{\underline{\theta}_k + n^{-\beta} u}{\underline{\theta}_k} \right)_s$ and then using Lemma 4.1 in the second factor in the above product we have

$$E \exp\left(\frac{1}{2} Z_k{}^n (u) \right) = \exp\left(- \frac{n^{1-\beta}}{2} \int_0^1 u_t \, dt \right)$$

$$. \exp\left(-n \int_0^1 \left[1 - \left(\frac{\underline{\theta}_k + n^{-\beta} u}{\underline{\theta}_k} \right)_t^{\frac{1}{2}} \right] \theta_t \, dt \right). \tag{4.13}$$

Let $F(x) = 1 - \left(\frac{a + x}{a} \right)^{\frac{1}{2}}$ where $a > 0$ and $x \in (-a, a)$. Then by Taylor's formula we have

$$F(x) = \frac{-x}{2a} + \frac{x^2}{4\sqrt{a}} (a + x^*)^{\frac{-3}{2}}$$

for some x^*, $x^* \in (0, x)$ (or $(x, 0)$ as the case may be).

Taking $a = \underline{\theta}_k\ (t)$ and $x = n^{-\beta}\, u_t$, the RHS of (4.13) can be written as

$$\exp\Big(-\frac{n^{1-\beta}}{2}\int_0^1 u_t\, dt + \frac{n^{1-\beta}}{2}\int_0^1 u_t\, \frac{\theta_t}{\underline{\theta}_k\ (t)}\, dt$$

$$- \frac{n^{1-2\beta}}{4}\int_0^1 \frac{u_t^{\,2}}{\sqrt{\underline{\theta}_k\ (t)}}\,(\,\underline{\theta}_k + n^{-\beta}\, u^*\,)_t^{\frac{-3}{2}}\ \theta_t\, dt\ \Big)$$

Writing $\dfrac{\theta_t}{\underline{\theta}_k\ (t)} = 1 + \dfrac{(\,\theta - \underline{\theta}_k\,)_t}{\underline{\theta}_k\ (t)}$ in the second term, the above quantity takes the form

$$\exp\Big(\frac{n^{1-\beta}}{2}\int_0^1 u_t\, \frac{(\,\theta - \underline{\theta}_k\,)_t}{\underline{\theta}_k\ (t)}\, dt$$

$$- \frac{n^{1-2\beta}}{4}\int_0^1 \frac{u_t^{\,2}}{\sqrt{\underline{\theta}_k\ (t)}}\,(\,\underline{\theta}_k + n^{-\beta}\, u^*\,)_t^{\frac{-3}{2}}\ \theta_t\, dt\ \Big). \qquad (4.14)$$

Note that, since it is a member of S_k, $\underline{\theta}_k\ (t) > k^{-q} > 0$ for *every* t. Let us consider the first term in the exponent above. From the Cauchy–Schwartz inequality,

$$\frac{n^{1-\beta}}{2}\int_0^1 \Big|\, u_t\, \frac{(\,\theta - \underline{\theta}_k\,)_t}{\underline{\theta}_k\ (t)}\,\Big|\, dt \le \frac{n^{1-\beta}}{2}\ \Big\|\, \frac{u}{\underline{\theta}_k}\,\Big\|\,.\ \|\,\theta - \underline{\theta}_k\,\|.$$

However, note that,

$$\Big|\, \frac{u_t}{\underline{\theta}_k\ (t)}\,\Big| \le\ 2\, n^{\beta}\, k^{2q};$$

because, (i) $\underline{\theta}_k \in S_k$ implies that $k^{-q} \le \underline{\theta}_k\ (t) \le k^{q}$ for all t and (ii) $u \in U_k{}^n$ implies that $|\, u_t\,| \le\ 2\, n^{\beta}\, k^{q}$; therefore,

$$\Big|\, \frac{u_t}{\underline{\theta}_k\ (t)}\,\Big| \le\ \frac{2\, n^{\beta}\, k^{q}}{k^{-q}} = 2\, n^{\beta}\, k^{2q}\ .$$

Hence $\Big\|\, \dfrac{u}{\underline{\theta}_k}\,\Big\| \le\ 2\, n^{\beta}\, k^{2q}$ and therefore,

$$\frac{n^{1-\beta}}{2}\int_0^1 \Big|\, u_t\, \frac{(\,\theta - \underline{\theta}_k\,)_t}{\underline{\theta}_k\ (t)}\,\Big|\, dt \le \frac{n^{1-\beta}}{2}\ 2\, n^{\beta}\, k^{2q}\ .\ \|\,\theta - \underline{\theta}_k\,\|. \qquad (4.15)$$

Now consider the second term in the exponent of (4.14). First, since $\theta \in \Theta$ there exist m_θ and M_θ such that, $0 < m_\theta < \theta_t < M_\theta < \infty$. Furthermore,

$$\frac{(\underline{\theta}_k + n^{-\beta} u^*)_t^{\frac{-3}{2}}}{\sqrt{\underline{\theta}_k (t)}} \geq k^{-2q} \quad \text{for all t.}$$

This is so because, $u^*_t \in (0, u_t)$ (i.e. $| u^*_t | \leq | u_t |$), implies that $\underline{\theta}_k + n^{-\beta} u^*$ is a member of S_k (so is $\underline{\theta}_k$) and therefore,

$$k^{-q} \leq (\underline{\theta}_k + n^{-\beta} u^*)_t \leq k^q$$

and

$$k^{-q} \leq (\underline{\theta}_k)_t \leq k^q \quad \text{for all t.}$$

Hence,

$$\frac{(\underline{\theta}_k + n^{-\beta} u^*)_t^{\frac{-3}{2}}}{\sqrt{\underline{\theta}_k (t)}} = \{ (\underline{\theta}_k)_t (\underline{\theta}_k + n^{-\beta} u^*)_t^3 \}^{-\frac{1}{2}} \geq k^{-2q}.$$

Therefore,

$$- \frac{n^{1-2\beta}}{4} \int_0^1 \frac{u_t^2}{\sqrt{\underline{\theta}_k (t)}} (\underline{\theta}_k + n^{-\beta} u^*)_t^{\frac{-3}{2}} \theta_t \, dt$$

$$\leq - \frac{n^{1-2\beta}}{4} m_\theta k^{-2q} \| u \|^2. \tag{4.16}$$

From (4.15) and (4.16) we then have

$$\exp \Big(\frac{n^{1-\beta}}{2} \int_0^1 u_t \frac{(\theta - \underline{\theta}_k)_t}{\underline{\theta}_k (t)} dt$$

$$- \frac{n^{1-2\beta}}{4} \int_0^1 \frac{u_t^2}{\sqrt{\underline{\theta}_k (t)}} (\underline{\theta}_k + n^{-\beta} u^*)_t^{-\frac{3}{2}} \theta_t \, dt \Big)$$

$$\leq \exp \Big(\frac{n^{1-\beta}}{2} 2 n^\beta k^{2q} . \| \theta - \underline{\theta}_k \| - \frac{n^{1-2\beta}}{4} m_\theta k^{-2q} \| u \|^2 \Big). \tag{4.17}$$

Therefore, in view of (4.14) and (4.17), we have shown that

$$E \exp\left(\frac{1}{2} Z_k{}^n(u)\right) \leq \exp\left(-C_\theta n^{1-2\beta} k^{-2q} \|u\|^2 + n k^{2q} \|\theta - \underline{\theta}_k\|\right)$$

$$= \exp\left(-g^n{}_k(\|u\|)\right);$$

which completes the proof.

Lemma 4.3 For $u, v \in U_k{}^n$,

$$E \mid Z_k{}^n(u) - Z_k{}^n(v) \mid^{2k} \leq B^n{}_k \|u - v\|^{2k}$$

where

$$B^n{}_k = 2^{3k+1} n^{2k(1-\beta)+1} k^{2k(1+q)} M_\theta{}^{2k}.$$

The proof of this lemma is very similar to the proof of Lemma 3.3 and we omit it. Lemmas 4.2 & 4.3 verify all the conditions of Corollary 2.1 (the verification being quite routine) and thus Theorem 4.1 is proved. The deduction of Theorems 4.2 - 4.4 from Theorem 4.1 is also straightforward.

Bibliography

Dellacherie, C. and Meyer, P., (1982), *Probabilty and Potential B*, North - Holland.

Geman, S., (1981), *Sieves for nonparametric estimation of densities and regressions*, Reports in Pattern analysis, No. 99, D.A.M.. Brown University.

Geman, S, and Hwang, C.R., (1982) *Nonparametric maximum likelihood estimation by the method of sieves*, Ann. Stat., 10, 401 - 414.

Grenander, U., (1981), *Abstract inference* John Wiley & Sons, New York.

Ibragimov, I.A. and Hasminski, R.Z., (1981), *Statistical estimation: Asymptotic theory*, Springer - Verlag, New York.

Karr, A.F., (1987), *Maximum likelihood estimation in multiplicative intensity model*, Ann. Stat. 15, No. 2, 473 - 490.

Kutoyants, Yu.A., (1984), *Statistics and control of stochastic processes*, Steclov Seminar.

Liptser, R.S. and Shiryayev, A.N., (1977), *Statistics of random processes I*, Springer - Verlag, Berlin.

Liptser, R.S. and Shiryayev, A.N., (1978), *Statistics of random processes II*, Springer - Verlag, Berlin.

Nguyen, H.T., and Pham, D.P., (1982), *Indentification of the Non-stationary Diffusion Model by the method of sieves*, SIAM J. of Control and Optimization, 20 (5).

Rao, C.R., (1973), *Linear statistical inference and its applications*, John Wiley & Sons, New York.

Selukar, R.S., (1989) *On estimation of Hilbert space valued parameters*, Ph.D. thesis, University of North Carolina at Chapel Hill.

Wald, A., (1949), *Note on the consistency of the maximum likelihood estimate*, Ann. Math. Stat. 20. 595 - 601.

Zygmund, A., (1968), *Trigonometric Series, Vol. I & II*, University Press, Cambridge.

238. G. Kallianpur and I. Mitoma, A Langevin-type stochastic differential equation on a space of generalized functionals, Aug. 88.

239. C. Houdré, Harmonizability, V-boundedness, (2,P)-boundedness of stochastic processes, Aug. 88. *Prob. Th. Rel. Fields*, to appear.

240. C.W. Johnson and G. Kallianpur, Some remarks on Hu and Meyer's paper and infinite dimensional calculus on finitely additive canonical Hilbert space, Sept. 88. *Th. Prob. Appl.*, to appear.

241. L. de Haan, A Brownian bridge connected with extreme values, Sept. 88. *Sankhya*, 1989, to appear.

242. O. Kallenberg, Exchangeable random measures in the plane, Sept. 88, *J. Theor. Probab.*, to appear.

243. E. Masry and S. Cambanis, Trapezoidal Monte Carlo integration, Sept. 88. *SIAM J. Numer. Anal.*, 1989, to appear.

244. L. Pitt, On a problem of H.P. McKean, Sept. 88. *Ann. Probability*, 17, 1989, to appear.

245. C. Houdré, On the linear prediction of multivariate (2,P)-bounded processes, Sept. 88.

246. C. Houdré, Stochastic processes as Fourier integrals and dilation of vector measures, Sept. 88, *Bull. Amer. Math. Soc.*, to appear.

247. J. Mijnheer, On the rate of convergence in Strassen's functional law of the iterated logarithm, Sept. 88, *Probab. Theor. Rel. Fields*, to appear.

248. G. Kallianpur and V. Perez-Abreu, Weak convergence of solutions of stochastic evolution equations on nuclear spaces, Oct. 88, Proc. Trento Conf. on Infinite Dimensional Stochastic Differential Equations, 1989, to appear.

249. R.L. Smith, Bias and variance approximations for estimators of extreme quantiles, Nov. 88.

250. H. Hurd, Spectral coherence of nonstationary and transient stochastic processes, Nov. 88. 4th Annual ASSP Workshop on Spectrum Estimation and Modeling, Minneapolis, 1988, 387-390.

251. J. Leskow, Maximum likelihood estimator for almost periodic stochastic processes models, Dec. 88.

252. M.R. Leadbetter and T. Hsing, Limit theorems for strongly mixing stationary random measures, Jan. 89. *Stochastic Proc. Appl.*, to appear.

253. M.R. Leadbetter, I. Weissman, L. de Haan, H. Rootzén, On clustering of high values in statistically stationary series, Jan. 89. *Proc. 4th Int. Meeting on Stat. Climatology*, to appear.

254. J. Leskow, Least squares estimation in almost periodic point processes models, Feb. 89.

255. N.N. Vakhania, Orthogonal random vectors and the Hurwitz-Radon-Eckmann theorem, Apr. 89.

256. E. Mayer-Wolf, A central limit theorem in nonlinear filtering, Apr. 89.

257. C. Houdré, Factorization algorithms and non-stationary Wiener filtering, Apr. 89.

258. C. Houdré, Linear Fourier and stochastic analysis, Apr. 89.

259. G. Kallianpur, A line grid method in areal sampling and its connection with some early work of H. Robbins, Apr. 89. *Amer. J. Math. Manag. Sci.*, 1989, to appear.

260. G. Kallianpur, A.G. Miamee and H. Niemi, On the prediction theory of two-parameter stationary random fields, Apr. 89. *J. Multivariate Anal.*, to appear.

261. I. Herbst and L. Pitt, Diffusion equation techniques in stochastic monotonicity and positive correlations, Apr. 89.

262. R. Selukar, On estimation of Hilbert space valued parameters, Apr. 89. (*Dissertation*)

263. E. Mayer-Wolf, The noncontinuity of the inverse Radon transform with an application to probability laws, Apr. 89.

264. D. Monrad and W. Philipp, Approximation theorems for weakly dependent random vectors and Hilbert space valued martingales, Apr. 89.

265. K. Benhenni and S. Cambanis, Sampling designs for estimating integrals of stochastic processes, Apr. 89.

266. S. Evans, Association and random measures, May 89.

267. H. Hurd, Correlation theory of almost periodically correlated processes, June 89.

268. O. Kallenberg, Random time change and an integral representation for marked stopping times, June 89.

269. O. Kallenberg, Some uses of point processes in multiple stochastic integration, Aug. 89.

270. W. Wu and S. Cambanis, Conditional variance of symmetric stable variables, Sept. 89.

271. J. Mijnheer, U-statistics and double stable integrals, Sept. 89.

272. O. Kallenberg, On an independence criterion for multiple Wiener integrals, Sept. 89.

273. G. Kallianpur, Infinite dimensional stochastic differential equations with applications, Sept. 89.

274. G.W. Johnson and G. Kallianpur, Homogeneous chaos, p-forms, scaling and the Feynman integral, Sept. 89.

275. T. Hida, A white noise theory of infinite dimensional calculus, Oct. 89.

276. K. Benhenni, Sample designs for estimating integrals of stochastic processes, Oct. 89. (*Dissertation*)

277. I. Rychlik, The two-barrier problem for continuously differentiable processes, Oct. 89.

278. G. Kallianpur and R. Selukar, Estimation of Hilbert space valued parameters by the method of sieves, Oct. 89.

279. G. Kallianpur and R. Selukar, Parameter estimation in linear filtering, Oct. 89.