C R RAO Advanced Institute of Mathematics, Statistics and Computer Science (AIMSCS)

C R RAO AIMSCS
Lecture Notes Series

Author (s): **B.L.S. PRAKASA RAO**

Title of the Notes: **Inference for Stochastic Processes: An Introduction**

Lecture Notes No.: **LN2013-02**

Date: **July 15, 2013**

Prof. C R Rao Road, University of Hyderabad Campus,
Gachibowli, Hyderabad-500046, INDIA.
www.crraoaimscs.org

Inference for Stochastic Processes: An Introduction

B.L.S. Prakasa Rao

CR Rao AIMSCS
July 15, 2013

Preface

This Lecture notes consist of introductory lectures on "Inference for stochastic processes" delivered by me at the Indian Statistical Institute, Delhi Centre, University of Pune, Indian Institute of Technology, Mumbai, University of Hyderabad and at the "Summer School" arranged by the University of Bocconi,Italy during July 1-20, 2002 and at the University of Bocconi, Milano, Italy during December 3-16, 2006. The earlier books dealing with this topic were *Statistical Inference for Markov Processes* by P. Billingsley, University of Chicago Press,Chicago (1961), *Statistics of Random Processes: General Theory* by R.S. Liptser and A.N. Shiryayev, Springer, New York (1977), *Statistics of Random Processes: Applications* by R.S. Liptser, and A.N. Shiryayev, Springer, New York (1978) , *Statistical Inference for Stochastic processes* by I.V. Basawa and B.L.S. Prakasa Rao, Academic Press, London (1980) and *Abstract Inference* by Ulf Grenander, Wiley, New York (1981). There is a large amount of literature since then dealing with various aspects on inference for stochastic processes. Some of the important review papers and other papers and books are listed at the end of this notes. The journal "Statistical Inference for Stochastic processes" edited by Denis Bosq, published since 1998, deals with parametric, semiparametric and nonparametric inference in discrete and continuous time processes.

B.L.S. Prakasa Rao
CR Rao AIMSCS
Hyderabad, India
July 15. 2013

Lecture 1

Stochastic Processes

Let $(\Omega, \mathcal{F}, P)$ be a Probability space. A stochastic process $\{X_t, t \in \tau\}$ is a family of random variables defined as $(\Omega, \mathcal{F}, P)$. We consider $\tau = [0, \infty)$ or $\tau = \{1, 2, \ldots\}$ in general. Let $t_1, t_2, \ldots, t_k \in \tau$. The joint distribution of $(X(t_1), \ldots, X(t_k))$ is called a finite dimensional distribution of the process. The probability structure of the process will be completely known once we are able to find all the finite dimensional distributions. The finite dimensional distributions of the process form a consistent family.If $\tau = [0, \infty)$, then $\{X_t, t \in \tau\}$ is called a *continuous time* stochastic process. If $\tau = \{1, 2, \ldots\}$, then we will call it a *discrete time* stochastic process.

Discrete time case

Suppose we have observed $X_1, \ldots, X_n$. Is it possible to determine the probability structure of the process as $n \to \infty$?

Continuous case

Suppose the process $\{X(t), 0 \leq t \leq T\}$ is observed. Is it possible to determine the probability structure of the process as $T \to \infty$?

For any n and $t_1, t_2, \ldots, t_n \in \tau$, specify a probability distribution on R^n by the joint distribution function

$$F_{t_1,.,t_n}(x_1, ., x_n).$$

Then F defines a probability measure on the σ-algebra of Borel sets B in R^τ. Let R^τ be the space of all real-valued functions and

$$C = \{x \in R^\tau : (x(t_1), ., x(t_n)) \in B\}$$

where $t_1, ., t_n \in \tau$, $B \subset R^n$, $n \geq 1$. Consider the σ-algebra $\mathcal{B}^\tau$ generated by such cylinder sets.

Kolmogorov's consistency theorem

The family of finite dimensional distribution functions $F_{t_1,.,t_n}(x_1, ., x_n)$, $n \geq 1$, $t_k \in \tau$, $1 \leq k \leq n$ induces a probability measure on $(\mathcal{R}^\tau, \mathcal{B}^\tau)$ if and only if

(a) any $F_{t_1,\ldots,t_n}(x_1, \ldots, x_n)$ is symmetric with respect to any permutation of the vector $(x_1, ., x_n)$ and the *same* permutation of the vector $(t_1, ., t_n)$ and

(b) $\lim_{x \to \infty} F_{t_1,\ldots,t_n}(x_1, \ldots, x_{n-1}, x) = F_{t_1,\ldots,t_{n-1}}(x_1, ., x_{n-1})$.

Let $(\Omega, \mathcal{F})$ be a measurable space and $\{P_\theta, \theta \in \Theta\}$ be a family of probability measures defined on $(\Omega, \mathcal{F})$. Let $\{X_n, n \geq 1\}$ be a stochastic process defined on $(\Omega, \mathcal{F}, P_\theta)$. Suppose we observe the process $\{X_k, 1 \leq k \leq n\}$. The basic problem is to estimate the parameter θ based on the observation $\{X_k, 1 \leq k \leq n\}$.

Let $F_{X_1,.,X_n;\theta}(x_1, ., x_n; \theta)$ be the joint distribution function of $(X_1, ., X_n)$ when *theta* is the true parameter.

We say that the family of probability measures $\{P_\theta\}$ is dominated by the σ-finite measure μ if

$$\mu(A) = 0 \Rightarrow P_\theta(A) = 0 \quad \text{for all} \quad A \in \mathcal{F}. \quad (P_\theta \ll \mu)$$

We can write down the likelihood function for $(X_1, ., X_n)$ in case μ is a Lebesgue measure on R^n or μ is a counting measure on R^n and the problem of estimation of the parameter θ through the maximum likelihood method is well understood.

Suppose $\{X_t, t \geq 0\}$ is a stochastic process defined on $(\Omega, \mathcal{F}, P_\theta), \theta \in \Theta$. Suppose further we observe the process $\{X_t, 0 \leq t \leq T\}$. The problem is to estimate θ. The question arises to what is the joint distribution of $\{X_t, 0 \leq t \leq T\}$. How to define the likelihood function? How to calculate the likelihood function even if it is defined? Let us look at the process as a mapping from Ω to R^T. From the Kolmogorov consistency theorem, there exists a probability measure Q_θ generated by $X^T = \{X_t, 0 \leq t \leq T\}$ on $(R^T, \mathcal{B}_T)$ when θ is the true parameter. If we know that $Q_{\theta_1} \ll Q_{\theta_2}$ and $Q_{\theta_2} \ll Q_{\theta_1}$, for all θ_1, θ_2 in Θ, then we can compute the Radon-Nikodym derivative

$$\frac{dQ_\theta}{dQ_{\theta_0}}$$

with respect to a fixed $\theta_0 \in \Theta$ and try to maximize it to obtain an estimator for θ.

We now discuss some concepts leading to such methodology.

Examples of Stochastic Models

Stochastic models are used in scientific research in a spectrum of disciplines. We now describe a few to indicate their use.

1) Random Walk model for neuron firing (Point process)

The neuron fires when the membrane potential reaches a critical threshold value, say C. Excitatory and inhibitory impulses are the inputs for the neuron: these inputs arrive according to a Poisson process. Each excitatory impulse increases and each inhibitory impulse decreases the membrane potential by a random quantity X with same p.d.f. $f(x)$. After each firing, the membrane potential is reset at zero and the process is repeated.

Let $Y_1, Y_2 \ldots$ denote the times at which the neuron fires. The process of interspike intervals $Y_1, Y_2 - Y_1, Y_3 - Y_2, \ldots$ is of interest to the neurologist.

2) <u>Epidemiology</u> (Greenwood Model)(Markov Chain)

Suppose at the time $t = 0$, there are S_0 suceptibles and I_0 infectives. After a certain latent period of the infection, (say) a unit of time, some of the suceptibles are infected. Thus, at time $t = 1$, the initial S_0 suceptibles split into two groups : those who are infected, I_1 in number say, and the remaining suceptibles say, S_1. The process continues until there are no more suceptibles in the population.

Note that

$$S(t) = S(t+1) + I(t+1), t = 0, 1, 2, \ldots$$

Suppose the probability of a suceptible being infected is p. Then

$$P\left(S(t+1) = s(t+1) | S(t) = s(t)\right) = \left(\begin{array}{c} s(t) \\ s(t) - s(t+1) \end{array}\right) p^{s(t)-s(t+1)}(1-p)^{s(t+1)}.$$

since $s(t) - s(t+1)$ are infected and $s(t+1)$ are suceptible.

The process $\{S(t), t = 0, 1, 2, \ldots\}$ is a Markov chain.

3) <u>Population growth model</u> (Branching process)

Suppose an organism produces a random number, say Y, of offspring with $p_k = P(Y = k), k = 0, 1, 2, \ldots, \sum p_k = 1$. Each offspring in turn produces organism independently according to the same distribution $\{p_k\}$. Suppose $Z(0) = 1$. If $\{Z(t)\}$ denotes the population size at the t-th generation, $t = 0, 1, 2, .,$ then $Z(t)$ is a Markov chain with transition probabilities given by

$$P(Z(t) = j \,| Z(t-1) = i) = P(Y_1 + \cdots + Y_i = j)$$

where $Y_1, Y_2, \ldots$ are i.i.d. with distribution $\{p_k\}$.

4) <u>Population genetics</u> (Diffusion process)

Consider a population of $2N$ genes each of which belongs to one of the two genotypes (say) A and B. Let $X(t)$ denote the proportion of type A genes in the t-th generation. Assuming that the total number of genes remain the same from one generation to next (we are neglecting selection and mutation effects), the genes in the $(t+1)$-th generation may be assumed to be a random sample of size $2N$ of genes from the t-th generation. The sequence $\{X(t), t = 1, 2, \ldots\}$ form a Markov chain. Conditionally on $X(t-1) = x$, $2N\,X(t)$ will be a Bionomial random Variable with $2N$ as the number of trials and x as the probability of success. One can approximate the Markov chain by a continuous time Markov process with a continuous state space $[0, 1]$. Such an approximation is an example of a diffusion process.

5) Storage model

Let $X(t)$ denote the annual random input during the year $(t, t+1)$ and M be the annual non-random release at the end of each year. Let $Z(t)$ denote the content of the dam after the release. Then

$$Z(t+1) = \min\{Z(t) + X(t), K\} - \min\{Z(t) + X(t), M)\}$$

where K is the capacity of the dam and $t = 0, 1, 2, \ldots$ If the inputs $\{X(t)\}$ are assumed to be independent random variables, then the sequence $\{Z(t), t = 0, 1, 2, \ldots\}$ forms a Markov chain.

6) Compound Poisson Model (Insurance)

Suppose an insurance company receives claims from its clients in accordance with a Poisson process with intensity λ. Assume that $Y_k, k = 1, 2, \ldots$ of successive claims are independent random variable with common distribution function $F(.)$. Then the total amount $X(t)$ of claims arising in the time interval $[0, t]$ is given by

$$X(t) = Y_1 + \cdots + Y_{N(t)}$$

where $N(t)$ is a Poisson random variable with mean λt. The process $\{X(t), t \geq 0\}$ is a compound Poisson process.

7) Queuing Model (for telephone calls)

Suppose the calls arrive at a telephone exchange according to a Poisson process. Duration of successive calls may be assumed to be independent exponential random variables. The capacity of the exchange may be limited to (say) K calls at any given time. The expected waiting time for a call to go through and the queue size at any particular time are of interest.

8) Signal processing

Suppose $X(t)$ is a signal satisfying the equation

$$X(t+1) = aX(t) + \xi(t)$$

where a is a fixed parameter and $\xi(t)$ represents error. Suppose the true signal is unobserved but $Y(t)$ is observed where

$$Y(t) = X(t) + Z(t)$$

where $Z(t)$ is noise. The problem is to estimate the signal $X(n+1)$ given $Y(0), \ldots, Y(n)$.

9) Time Series

Let $X(t)$ denote the price of a commodity at time t. Suppose we fit a ARMA model

$$X(t) + \alpha_1 X(t-1) + \cdots + \alpha_p X(t-p) = Z(t) + \beta_1 Z(t-1) + \cdots + \beta_q Z(t-q)$$

where Z 's are i.i.d. unobservable random variables. Given $\{X(0), ., X(n)\}$, we may want to predict $X(n+1)$ and in turn the problem is to estimate α 's and β 's.

Lecture 2

Discrete parameter martingales

Let $(\Omega, \mathcal{F}, P)$ be a probability space. Let $\{\mathcal{F}_n, n \geq 1\}$ be a non-decreasing sequence of sub σ-algebras of $\mathcal{F}$. Suppose $\{Z_n, n \geq 1\}$ is a sequence of random variables defined on $(\Omega, \mathcal{F}, P)$ such that

(i) Z_n is measurable with respect to $\mathcal{F}_n$,

(ii) $E|Z_n| < \infty$,

(iii) $E(Z_n|\mathcal{F}_m) = Z_m$ a.s. for all $1 \leq m < n, n \geq 1$.

Then the sequence $\{Z_n, n \geq 1\}$ is said to be a *martingale* with respect to $\{\mathcal{F}_n, n \geq 1\}$ and we say that $\{Z_n, \mathcal{F}_n, n \geq 1\}$ is a martingale.

Remark. It is clear that $E(Z_n) = E(Z_m)$ for all n and m if $\{Z_n, \mathcal{F}_n, n \geq 1\}$ is a martingale. If (i) and (ii) hold and if (iii) $E(Z_n|\mathcal{F}_m) \geq Z_m$ a.s. for all $1 \leq m \leq n$, then $\{Z_n, \mathcal{F}_n, n \geq 1\}$ is said to be a *submartingale.* A submartingale $\{Z_n, \mathcal{F}_n, n \geq 1\}$ is said to be L^1-bounded if $\sup_n E|Z_n| < \infty$.

Uniform integrability

A sequence of random variable $\{Y_n\}$ is said to be *uniformly integrable* if

$$\lim_{c \to \infty} \sup_n E\{|Y_n| I(|Y_n| > c)\} = 0.$$

Remarks: A sufficient condition for uniform integrability of the sequence $\{Y_n\}$ is that $\sup_n E|Y_n|^{1+\varepsilon} < \infty$ for some $\varepsilon > 0$.

Martingale Convergence Theorem : Let $\{Z_n, \mathcal{F}_n, n \geq 1\}$ be an L^1-bounded submartingale. Then there exists a random variable Z such that $\lim_{n \to \infty} Z_n = Z$ a.s. and

$$E|Z| \leq \liminf_{n \to \infty} E|Z_n| < \infty.$$

Remarks : (i)If the submartingale is uniformly integrable, then $Z_n \to Z$ in L^1 and if $\{Z_n, \mathcal{F}_n\}$ is an L^2-bounded martingale, then $E|Z_n - Z|^2 \to 0$. (ii) Any nonnegative martingale converges a.s.

Examples of martingales

1) Suppose $X_1, X_2, \ldots$ are independent random variable with $E(X_i) < \infty$ for $i \geq 1$. Define

$$S_n = X_1 + \cdots + X_n$$

and let $\mathcal{F}_n$ be σ-algebra generated by $X_1, ., X_n$. Suppose $E(X_i) = 0$ for all i. Then $\{S_n, \mathcal{F}_n, n \geq 1\}$ is a martingale.

2) Suppose $X_1, X_2, \ldots$ are independent random variables with $E(|X_i|) < \infty$ for $i \geq 1$. Let t be a real number and define

$$Z_n = \frac{e^{itS_n}}{E[e^{itS_n}]}, n \geq 1.$$

Then $\{Z_n, \mathcal{F}_n, n \geq 1\}$ is a martingale where $\mathcal{F}_n$ is the σ-algebra generated by $X_1, \ldots, X_n$.

3) Let $\{X_n, n \geq 1\}$ be a stochastic process with $f(x_1, \ldots, x_n; \theta)$ as the joint density of $(X_1, \ldots, X_n)$ when θ is a scalar parameter. Let $L_n(\theta) = f(X_1, ., X_n; \theta)$. Suppose the function $L_n(\theta)$ is differentiable with respect to θ. Let

$$u_n(\theta) = \frac{d}{d\theta}[\log L_n(\theta) - \log L_{n-1}(\theta)]$$

and $\mathcal{F}_n$ be the σ-algebra generated by $X_1, \ldots, X_n$. Then $\{\sum_{i=1}^{n} u_i(\theta), \mathcal{F}_n, n \geq 1\}$ forms a martingale under some regularity conditions.

Sketches of proofs

1.

$$\begin{aligned} & E(S_n|X, \ldots, X_{n-1}) \\ &= E[S_{n-1} + X_n | X_1, ., X_{n-1}] \\ &= S_{n-1} + E[X_n | X_1, \ldots, X_{n-1}] \\ &= S_{n-1} + E(X_n) \\ &= S_{n-1}. \end{aligned}$$

2.

$$\begin{aligned} & E[Z_n | X_1, \ldots, , X_{n-1}] \\ &= E\left[\frac{e^{it(S_{n-1}+X_n)}}{E[e^{itS_{n-1}} e^{itX_n}]} \middle| X_1, \ldots, X_{n-1} \right] \\ &= \frac{e^{itS_{n-1}}}{E[e^{itS_{n-1}}]E(e^{itX_n})} E[e^{itX_n} | X_1, \ldots, X_{n-1}] \\ &= \frac{e^{itS_{n-1}}}{E[e^{itS_{n-1}}]}. \end{aligned}$$

3. $\int_{-\infty}^{\infty} f(x_n|x_1,\ldots,x_{n-1};\theta)\mu(dx_n) = 1$. Suppose we assume that differentiation under the integral sign with respect to the parameter θ is allowed. Then

$$\int_{-\infty}^{\infty} \frac{df(x_n|x,\ldots,x_{n-1};\theta)}{d\theta}\mu(dx_n) = 0$$

which implies that

$$\int_{-\infty}^{\infty} \frac{d}{d\theta}[\log L_n(\theta) - \log L_{n-1}(\theta)]f(x_n|x,\ldots,x_{n-1};\theta)\mu(dx_n) = 0.$$

Hence

$$E[u_n(\theta)|X_1,\ldots,X_{n-1}] = 0.$$

Remarks: Note that $\{\sum_{i=1}^{n}\{Z_i - E(Z_i|Z_1,\ldots,Z_{i-1})\},\mathcal{F}_n, n \geq 1\}$ forms a martingale for any sequence of random variables $\{Z_n\}$ defined on a probability space $(\Omega,\mathcal{F},P)$ with $E|Z_n| < \infty$ where $\mathcal{F}_n$ is the σ -algebra generated by $Z_1,\ldots,Z_n$.

Lecture 3

Weak law of large numbers (WLLN)

Suppose $\{S_n, \mathcal{F}_n, n \geq 1\}$ is a zero mean martingale with $S_n = \sum_{i=1}^{n} X_i$. Further suppose that $E(X_i^2) < \infty$ for $i \geq 1$. Then it follows that

$$\begin{aligned} E(X_i X_j | \mathcal{F}_i) &= X_i E(X_j | \mathcal{F}_i) \quad \text{for } 1 \leq i < j \\ &= 0 \end{aligned}$$

and hence $E(X_i X_j) = 0$ for $1 \leq i < j$. Therefore

$$\mathrm{Var}(S_n) = E(S_n^2) = \sum_{i=1}^{n} \mathrm{Var}(X_i).$$

Hence

$$P(|S_n| \geq \varepsilon) \leq \varepsilon^{-2} E(S_n^2) \quad \text{(by Chebyshev's inequality)}$$

which implies that

$$\frac{S_n}{n} \xrightarrow{p} 0 \text{ if } \frac{1}{n^2} \sum_{j=1}^{n} EX_j^2 \to 0$$

which can be termed as a WLLN for martingales.

Remarks: Weaker condition can be given for the WLLN to hold.

Strong Law of Large Numbers (SLLN)(Feller (1971)), p.242; Loéve (1977), p.250) Suppose $\{S_n\}$ is a zero mean martingale with $E(X_i^2) < \infty$ for $i \geq 1$. Further suppose that there is a sequence $b_n \uparrow \infty$ such that

$$\sum_{n=1}^{\infty} \frac{EX_n^2}{b_n^2} < \infty.$$

Then $\lim\limits_{n\to\infty} \frac{S_n}{b_n} = 0$ a.s.

Remarks: For alternate conditions for the SLLN to hold , see the results stated later in this section.

Central Limit Theorem (CLT)

The following central limit theorem was proved for martingales by Billingsley (1961) and by Ibragimov (1963).

Theorem: Let $\{Z_n, n \geq 1\}$ be a strictly stationary ergodic process such that $E(Z_1^2)$ is finite and $E(Z_n|Z_1, \ldots, Z_{n-1}) = 0$ a.s. for $n > 1$ and $E(Z_1) = 0$. Then

$$n^{-1/2} \sum_{k=1}^{n} Z_k \xrightarrow{\mathcal{L}} N(0, E(Z_1^2)) \text{ as } n \to \infty.$$

<u>CLT</u> (Brown (1971)) Let $\{S_n, \mathcal{F}_n, n \geq 1\}$ denote a zero mean martingale where $S_n = X_1 + \cdots + X_n$. Suppose $E(X_i^2) < \infty, i \geq 1$. Let

$$V_n^2 = \sum_{i-1} E(X_i^2 | \mathcal{F}_{i-1}),$$

and

$$s_n^2 = EV_n^2 = ES_n^2.$$

If $\frac{V_n^2}{s_n^2} \xrightarrow{p} 1$ and

$$\frac{1}{s_n^2} \sum_{i=1}^{n} E(X_i^2 I(|X_i| \geq \varepsilon s_n)) \to 0$$

as $n \to \infty$ for all $\varepsilon > 0$, then

$$\frac{S_n}{s_n} \xrightarrow{\mathcal{L}} N(0,1) \text{ as } n \to \infty.$$

Remark: For more general versions of CLT, see later in this section.

<u>Maximal inequality</u>

If $\{S_n, \mathcal{F}_n, n \geq 1\}$ is a zero mean martingale, then

$$P(\max_{1 \leq k \leq n} |S_n| \geq \varepsilon) \leq \frac{1}{\varepsilon^2} E(S_n^2).$$

<u>Toeplitz Lemma.</u>

If $a_i, i \geq 1$ are positive and $b_n = \sum_{i=1}^{n} a_i \uparrow \infty$, then $x_n \to x$ implies $b_n^{-1} \sum_{i=1}^{n} a_i x_i \to x$.

<u>Kronecker's Lemma</u>

Let $\{x_n\}$ be a real sequence such that $\sum_{n=1}^{\infty} x_n < \infty$. Let $\{b_n\}$ be monotone sequence of positive constants with $b_n \uparrow \infty$. Then

$$\frac{1}{b_n} \sum_{i=1}^{n} b_i x_i \to 0 \text{ as } n \to \infty.$$

The following result involves classical Lindeberg condition for asymptotic normality to hold for partial sums of independent random variables (Loéve, p.280)

Let $X_1, X_2, \ldots$ independent random variables with $E(X_n^2) < \infty$ for all $n \geq 1$ and $E(X_n) = 0$ for all n. Let $S_n = X_1 + \cdots + X_n$, $\sigma_k^2 = \text{Var}(X_k)$, and $s_n^2 = \sum_{i=1}^n \text{Var}(X_i)$. Then

$$\frac{S_n}{s_n} \xrightarrow{\mathcal{L}} N(0,1) \text{ as } n \to \infty$$

and

$$\max_{1 \leq k \leq n} \frac{\sigma_k}{s_n} \to 0 \text{ as } n \to \infty$$

if and only if for every $\varepsilon > 0$,

$$\frac{1}{s_n^2} \sum_{k=1}^n E[X_k^2 I(|X_k| \geq \varepsilon s_n)] \to 0$$

("if" part is due to Lindeberg and the "only if" part is due to Feller).

We now discuss some other versions of the WLLN, SLLN and the CLT for martingales.

WLLN Let $\{S_n = \sum_{i=1}^n X_i, \mathcal{F}_n, n \geq 1\}$ be a martingale and $0 < b_n \uparrow \infty$ as $n \to \infty$. Let $X_{ni} = X_i I(|X_i| \leq b_n)$, $1 \leq i \leq n$. Then

$$\frac{S_n}{b_n} \xrightarrow{p} 0$$

if

(i) $\sum_{i=1}^n P(|X_i| > b_n) \to 0$

(ii) $b_n^{-1} \sum_{i=1}^n E(X_{ni} | \mathcal{F}_{i-1}) \xrightarrow{p} 0$, and

(iii) $b_n^{-2} \sum_{i=1}^n \{EX_{ni}^2 - E[E(X_{ni}|\mathcal{F}_{i-1})]^2\} \to 0$.

Proof See Hall and Heyde (1980), p.30
(See Loéve (1977), p.290 for the independent case)

SLLN: Let $\left\{ S_n = \sum_{i=1}^n X_i, \mathcal{F}_n, n \geq 1 \right\}$ be a zero-mean square integrable martingale and $\{U_n, n \geq 1\}$ be a nondecreasing sequence of positive random variables such that U_n is $\mathcal{F}_{n-1}$ -measurable. Then

$$\lim_{n \to \infty} U_n^{-1} S_n = 0 \quad \text{a.s.}$$

on the set $\{\lim_{n \to \infty} U_n = \infty, \sum_{i=1}^\infty U_i^{-2} E(X_i^2 | \mathcal{F}_{i-1}) < \infty\}$.

Proof See Hall and Heyde (1980), p.35.

CLT Let $\{S_{ni}, \mathcal{F}_{ni}, 1 \leq i \leq k_n\}$ be a zero-mean square-integrable martingale for each $n \geq 1$. Let $X_{ni} = S_{ni} - S_{n,i-1}, 1 \leq i \leq k_n, S_{n0} = 0$. Suppose $k_n \uparrow \infty$ as $n \to \infty$. Then the double sequence $\{S_{ni}, \mathcal{F}_{ni}, 1 \leq i \leq k_n, n \geq 1\}$ is called a martingale array. Let

$$V_{ni}^2 = \sum_{j=1}^{i} E(X_{nj}^2 | \mathcal{F}_{n,j-1})$$

be the conditional variance of S_{ni}.

Special case If $\{S_n = \sum_{i=1}^{n} X_i, \mathcal{F}_n, n \geq 1\}$ is a martingale, then $S_{ni} = \frac{S_i}{s_n}, 1 \leq i \leq n$ where s_n is the standard deviation of S_n, $\mathcal{F}_{ni} = \mathcal{F}_i, k_n = n$ forms a martingale array.

Theorem: Suppose $\{S_{ni}, \mathcal{F}_{ni}, 1 \leq i \leq k_n, n \geq 1\}$ is a zero mean square integrable martingale array. Further suppose that

$$\mathcal{F}_{ni} \subseteq \mathcal{F}_{n+1,i} \text{ for } 1 \leq i \leq k_n, n \geq 1 \quad (\textit{nested condition})$$

and the following conditions hold:

(i) for all $\varepsilon > 0$,

$$\sum_{i=1}^{k_n} E(X_{ni}^2 I(|X_{ni}| > \varepsilon) | \mathcal{F}_{n,i-1}) \xrightarrow{p} 0,$$

(ii) $V_{nk_n}^2 = \sum_{i=1}^{k_n} E(X_{ni}^2 | \mathcal{F}_{n,i-1}) \xrightarrow{p} \eta^2$.

Then

$$S_{nk_n} = \sum_{i=1}^{k_n} X_{ni} \xrightarrow{\mathcal{L}} Z = \eta N(0,1) \quad (\textit{stably})$$

where η and $N(0,1)$ are independent random variables.

Remarks: Note that the random variable Z has the characteristic function $E\left(e^{-\frac{1}{2}\eta^2 t^2}\right)$. In fact $\frac{S_{nk_n}}{V_{n,k_n}} \xrightarrow{\mathcal{L}} N(0,1)$ as $n \to \infty$ provided $P(\eta^2 > 0) = 1$.

(For the definition of stable convergence, see p.13)

Remarks: The nested condition holds automatically in case the martingale array is built out of a single martingale as in the special case discussed above.

Sholomitski (*Theory of Probability and its Applications*, 43 (1999) 434-448) discussed necessary conditions for normal convergence of a martingale.

Let $(\Omega, \mathcal{F}, P)$ be a probability space and $\{X_{jn}, 1 \leq j \leq k_n < \infty\}$ be a double array of random variables defined on $(\Omega, \mathcal{F}, P)$.

Let

$$\mathcal{F}_{j,n} = \sigma(X_{1n}, \ldots, X_{jn}) \text{ with } \mathcal{F}_{0n} = \{\phi, \Omega\}.$$

Suppose

(1) $$E(X_{jn}|\mathcal{F}_{j-1,n}) = 0 \quad \text{a.s}\,[P].$$

Further suppose that X_{jn} are square integrable and

(2) $$\sum_{j=1}^{k_n} E(X_{jn}^2|\mathcal{F}_{j-1,n}) \xrightarrow{p} \sigma^2 > 0.$$

Then the "conditional Lindeberg condition",

(3) $$\wedge_n(\varepsilon) = \sum_{j=1}^{k_n} E(X_{jn}^2 I(|X_{jn}| \geq \varepsilon)|\mathcal{F}_{j-1,n}) \xrightarrow{p} 0$$

as $n \to \infty$ for every $\varepsilon > 0$, implies that

(4) $$\sum_{j=1}^{k_n} X_{jn} \xrightarrow{\mathcal{L}} N(0, \sigma^2) \text{ as } n \to \infty$$

(cf. Brown (1971) Ann. Math. Statist. 42, 59-66.)

Conversely suppose that

(5) $$\max_{1 \leq i \leq k_n} E(X_{jn}^2|\mathcal{F}_{j-1,n}) \xrightarrow{p} 0 \text{ as } n \to \infty$$

and the condition (2) holds. If, as $n \to \infty$.

(6) $$\sum_{j=1}^{k_n} c_{jn} X_{jn} \xrightarrow{\mathcal{L}} N(0, \sigma^2) \text{ as } n \to \infty$$

for any double array c_{jn} of ± 1, then the conditional Lindeberg condition stated in (3) holds. If the conditional distribution of X_{jn} given $\mathcal{F}_{j-1,n}$ is symmetric a.s, then the condition (4) itself implies (6) and hence the conditional Lindeberg condition (3) holds in the presence of the conditions (2) and (5).

Stable Convergence (Renyi(1963))

Let $(\Omega, \mathcal{F}, P)$ be a probability space. Suppose $Y_n \xrightarrow{\mathcal{L}} Y$. Then $Y_n \xrightarrow{\mathcal{L}} Y$ (*stably*) if for all continuity points y of the distribution function of Y and all events $E \in \mathcal{F}$,

$$\lim_{n\to\infty} P\left((Y_n \le y) \cap E\right) = Q_y(E)$$

exists and if $Q_y(E) \to P(E)$ as $y \to \infty$
(note the $Q_y(E)$ is a measure on $(\Omega, \mathcal{F})$ if it exists).

<u>Theorem</u> Suppose that $Y_n \xrightarrow{\mathcal{L}} Y$ where all the Y_n are defined on the same probability space $(\Omega, \mathcal{F}, P)$. Then $Y_n \xrightarrow{\mathcal{L}} Y$ (stably) if and only if there exists a random variable Y' with the same distribution as that of Y (possibly on an extension of $(\Omega, \mathcal{F}, P)$ such that for all real t

$$\exp(itY_n) \to Z(t) = \exp(itY') \text{ weakly in } L^1 \text{ as } n \to \infty$$

and $E[Z(t)I(E)]$ is a continuous function of t for all $E \in \mathcal{F}$.

<u>Remarks</u>: This theorem is a consequence of the continuity theorem for characteristic functions.

<u>Note</u>: A sequence Z_n on $(\Omega, \mathcal{F}, P)$ is said to *converge weakly* in L^1 to an integrable random variable Z on $(\Omega, \mathcal{F}, P)$ if for all $E \in \mathcal{F}$,

$$E(Z_n I(E)) \to E(ZI(E)), \text{ that is, } \int_E Z_n dP \to \int_E Z dP \text{ for all } E \in \mathcal{F}$$

and we write

$$Z_n \to Z \text{ weakly in } L^1.$$

Remarks:(1) Convergence weakly in L^1 is weaker than L^1-convergence. In fact $Z_n \to Z$ (weakly in L^1) implies $E(Z_n X) \to E(ZX)$ for all X which are $\mathcal{F}$-measurable.

Remarks: (2) If for all $E \in \mathcal{F}$ and for all continuity points y of the distribution function of Y,

$$\{P(Y_n \le y) \cap E\} \to P(Y \le y)P(E)$$

Then

$$Y_n \xrightarrow{\mathcal{L}} Y \text{ (mixing)}.$$

In other words Y_n are asymptotically independent of each event $E \in \mathcal{F}$. Mixing convergence is a special case of stable convergence.

Remarks: (3) (Continuation of Theorem on martingales on p.12)
Suppose that $P(\eta^2 > 0) = 1$. Since $S_{nk_n} \to Z$ (stably) where $Z = \eta N(0,1)$, for any

real t, it follows that
$e^{itS_{nk_n}} \to e^{itZ}$ weakly in L^1. Hence $E[e^{itS_{nk_n}}X] \to E[e^{itZ}X]$ for any random variable X which is $\mathcal{F}$-measurable. Let $X = e^{iu\eta+ivI(E)}$ where $-\infty < u, v < \infty$ and $E \in \mathcal{F}$. Then it follows that the joint characteristic function of $(S_{nk_n}, \eta, I(E))$ converges to that of $(\eta N, \eta, I(E))$ where N is a standard normal random variable independent of $(\eta, I(E))$. Therefore

$$(\eta^{-1}S_{nk_n}, I(E)) \xrightarrow{\mathcal{L}} (N, I(E))$$

and hence, if

$$V_{nk_n}^2 = \sum_{i=1}^{n} E(X_{ni}^2|\mathcal{F}_{n,i-1}) \xrightarrow{p} \eta^2,$$

as in the martingale limit theorem on p.12, it follows that

$$(V_{nk_n}^{-1}S_{nk_n}, I(E) \xrightarrow{\mathcal{L}} (N, I(E))$$

which implies that

$$V_{nk_n}^{-1}S_{nk_n} \xrightarrow{\mathcal{L}} N \text{ (stably)}$$

as $n \to \infty$.

Remarks: (4) The notion of stable convergence is helpful in interchanging the random norming and non-random norming for obtaining limit theorems for partial sums of martingale differences in the martingale central limit theory.

Lecture 4

Likelihood ratio

Let $(\Omega, \mathcal{F}, P)$ be a probability space and $\{\mathcal{F}_n\}$ be a sequence of sub σ -algebras of $\mathcal{F}$ such that $\mathcal{F}_n \subseteq \mathcal{F}_{n+1}, n \geq 1$ and $\mathcal{F}_n \uparrow \mathcal{F}$. Let $P^\star$ be another probability measure defined on $(\Omega, \mathcal{F})$. Note that $P^\star$ is absolutely continuous with respect to P $(P^\star \ll P)$ if $P(A) = 0 \Rightarrow P^\star(A) = 0$ for any $A \in \mathcal{F}$. If $P^\star \ll P$, then there exists a random variable

$$Z = \frac{dP^\star}{dP}$$

which is $\mathcal{F}$ - measurable such that

$$P^\star(A) = \int_A Z dP, A \in \mathcal{F}.$$

The random variable Z is the density (Radon-Nikodym derivative) of $P^\star$ with respect to P. If $Z > 0$ a.s. [P], then the measure $P^\star$ and P are equivalent and we write $P \simeq P^\star$.

Let $P_n^\star$ denote the restriction of $P^\star$ to $\mathcal{F}_n$ and P_n be the restriction of P to $\mathcal{F}_n$. If $P_n^\star \ll P_n$ for every n, then we say that $P^\star$ is *locally absolutely continuous* with respect to P and write $P^\star \overset{loc}{\ll} P$. Suppose $P^\star \overset{loc}{\ll} P$. Let

$$Z_n = \frac{dP_n^\star}{dP_n}.$$

Z_n is called the *local density*. For any $A \in \mathcal{F}_n$,

$$\begin{aligned}
\int_A Z_{n+1} dP &= \int_A \frac{dP_{n+1}^\star}{dP_{n+1}} dP \\
&= \int_A dP_{n+1}^\star \\
&= P_{n+1}^\star(A) \\
&= P_n^\star(A) \quad \text{since } A \in \mathcal{F}_n \\
&= \int_A dP_n^\star \\
&= \int_A \frac{dP_n^\star}{dP_n} dP \\
&= \int_A Z_n dP.
\end{aligned}$$

Hence $E(Z_{n+1}|\mathcal{F}_n) = Z_n$, which implies that $\{Z_n, \mathcal{F}_n, n \geq 1\}$ is a martingale and it is a nonnegative martingale. Hence $Z_n \to Z$ a.s. [P]. If $E[Z] = 1$, then $E|Z_n - Z| \to 0$ by Scheffe's theorem and $P^\star \ll P$ and $Z = \frac{dP^\star}{dP}$. In fact $Z_n = E(Z|\mathcal{F}_n)$. In general Z is the density of the absolutely continuous component of $P^\star$ with respect to P.

As an application of the above idea, one can obtain the following result which gives a method for calculating the Radon-Nikodym derivative (Gikhman and Skorokhood (1974) *Theory of Stochastic Processes*).

Theorem. Let $(\Omega, \mathcal{F}, P)$ be a probability space and Q be another probability measure on $(\Omega, \mathcal{F})$ absolutely continuous with respect to P. Let $\{A_{nk}, k \geq 1\}$ be a measurable partition of Ω for each $n \geq 1$. Suppose the sequence of partitions is nested. Let

$$g_n(\omega) = \frac{Q(A_{n,k(\omega)})}{P(A_{n,k(\omega)})}$$

if $P(A_{n,k(\omega)}) > 0$ where $A_{n,k(\omega)}$ is that set of the sequence $\{A_{nk,}k \geq 1\}$ which contains ω. If $P(A_{n,k(\omega)}) = 0$, let $g_n(\omega) = 0$. Then the sequence $\{g_n, \mathcal{F}_n, n \geq 1\}$ is a martingale where $\mathcal{F}_n = \sigma(A_{n1}, A_{n2}, \ldots)$. Suppose $\mathcal{F}_n \uparrow \mathcal{F}$ as $n \to \infty$. Then there exists a limiting function $g(\omega)$ such that

$$g_n(\omega) \to g(\omega) \text{ a.s as } n \to \infty$$

independent of the sequence of partitions $\{A_{nk}, k \geq 1; n \geq 1\}$, and for arbitrary $B \in \mathcal{F}$,

$$Q(B) = \int_B g(\omega) P(d\omega).$$

Estimation by the maximum likelihood method

Consider a stochastic process $\{X_{n_1} n \geq 1\}$ such that the finite dimensional distributions of the process are known but for a scalar parameter θ. Suppose $\theta \in \Theta$ open. Let us suppose that the process is observed up to time "n''.

Let $L_n(\theta)$ be the likelihood function associated with the observation $(X_1, \ldots, X_n)$. Let $p_n(x_1, \ldots, x_n; \theta) = L_n(\theta)$ be the joint probability (density) function of $(X_1, \ldots, X_n)$. Note that

$$\begin{aligned} p_n(x_1, \ldots, x_n; \theta) &= p_1(x_1; \theta) \frac{p_2(x_1, x_2, ; \theta)}{p_1(x_1; \theta)} \cdots \frac{p_n(x_1, \ldots, x_n; \theta)}{p_{n-1}(x_1, \ldots, x_{n-1}; \theta)} \\ &= p_1(x_1, \theta) p_2(x_2; \theta | x_1) \cdots p_n(x_n; \theta | x_1, \ldots, x_{n-1}). \end{aligned}$$

Hence

$$\begin{aligned} \log p_n(x_1, \ldots, x_n, \theta) &= \log p_1(x_1; \theta) \\ &\quad + \log p_2(x_1, x_2; \theta) - \log p_1(x_1; \theta) \\ &\quad + \log p_n(x_1 \ldots, x_n; \theta) - \log p_{n-1}(x_1, \ldots, x_{n-1}; \theta). \end{aligned}$$

In other words

$$\log L_n(\theta) = \log L_1(\theta) \\ +\log L_2(\theta) - \log L_1(\theta) + \cdots + \log L_n(\theta) - \log L_{n-1}(\theta).$$

For convenience, let us define $L_0(\theta) \equiv 1$. Then

$$\log L_n(\theta) = \sum_{i=1}^{n} \left\{\log L_i(\theta) - \log L_{i-1}(\theta)\right\}.$$

Assume that

$$p_n(x_n;\theta|x_1,\ldots,x_{n-1}) = \frac{p_n(x_1,\ldots,x_n;\theta)}{p_{n-1}(x_1,\ldots,x_{n-1};\theta)} = \frac{L_n(\theta)}{L_{n-1}(\theta)}$$

is differentiable twice with respect to θ under the (summation) integral sign. and

$$E_\theta\left(\frac{d\log L_n(\theta)}{d\theta}\right)^2 < \infty, \quad \theta \in \Theta.$$

Note that

$$\begin{aligned}\frac{d\log L_n(\theta)}{d\theta} &= \sum_{i=1}^{n} \frac{d}{d\theta}[\log L_i(\theta) - \log L_{i-1}(\theta)] \\ &= \sum_{i=1}^{n} u_i(\theta) \quad \text{(say)}.\end{aligned}$$

Then

$$\begin{aligned}E_\theta(u_i(\theta)|\mathcal{F}_{i-1}) &= E_\theta\left(\frac{d}{d\theta}\log p_i(X_i;\theta|x_1,\ldots,x_{i-1})|\mathcal{F}_{i-1}\right) \\ &= 0 \quad \text{a.s.}\end{aligned} \tag{7}$$

and

$$E_\theta\left(u_i^2(\theta)|\mathcal{F}_{i-1}\right) = -E_\theta\left(\frac{du_i(\theta)}{d\theta}|\mathcal{F}_{i-1}\right) \tag{8}$$

in view of the assumption made above. Let

$$I_n(\theta) = \sum_{i=1}^{n} E_\theta(u_i^2(\theta)|\mathcal{F}_{i-1}). \tag{9}$$

Observe that $I_n(\theta)$ is the partial sum of the conditional information in X_i given $X_1,\ldots,X_{i-1}$ summed over $1 \le i \le n$. Let

$$J_n(\theta) = \sum_{i=1}^{n} v_i(\theta) \quad \text{where } v_i(\theta) = \frac{du_i(\theta)}{d\theta}. \tag{10}$$

In view of (7),

$$\left\{\frac{d\log L_n(\theta)}{d\theta}, \mathcal{F}_n, n \geq 1\right\} \tag{11}$$

is a martingale. Furthermore

$$E_\theta(u_i^2(\theta) + v_i(\theta)|\mathcal{F}_{i-1}) = 0 \text{ a.s.} \tag{12}$$

Existence of a consistent solution of the likelihood equation

Observe that

$$\begin{aligned}
\left.\frac{d\log L_n(\theta)}{d\theta}\right|_{\theta=\theta'} &= \sum_{i=1}^n u_i(\theta') \\
&= \sum_{i=1}^n u_i(\theta) + (\theta' - \theta)\sum_{i=1}^n \left.\frac{du_i(\theta)}{d\theta}\right|_{\theta^\star} \\
&= \sum_{i=1}^n u_i(\theta) + (\theta' - \theta)J_n(\theta^\star) \\
&= \sum_{i=1}^n u_i(\theta) - (\theta' - \theta)I_n(\theta) + (\theta' - \theta)(J_n(\theta^\star) + I_n(\theta))
\end{aligned} \tag{13}$$

where $\theta^\star = \theta + \gamma(\theta' - \theta)$ with $|\gamma| < 1$. Let $X_i = u_i(\theta)$ and $U_n = I_n(\theta)$. Applying the SLLN stated on p.11, it follows that

$$\frac{\sum_{i=1}^n u_i(\theta)}{I_n(\theta)} \to 0 \quad \text{a.s as } n \to \infty \tag{14}$$

provided

$$I_n(\theta) \to \infty \text{ a.s as } n \to \infty \tag{15}$$

and

$$\sum_1^\infty I_i^{-2}(\theta)E(u_i^2(\theta)|\mathcal{F}_{i-1}) < \infty \text{ a.s.} \tag{16}$$

Let $\{a_n\}$ be any sequence of positive numbers and $b_n = \sum_{j=1}^{n} a_j$. Then

$$\begin{aligned}
&\sum_{n=1}^{\infty} \left(\sum_{1}^{n} a_j \right)^{-2} a_n \\
&= \sum_{n=1}^{\infty} b_n^{-2}(b_n - b_{n-1}) \quad (b_0 \equiv 0) \\
&= \sum_{n=1}^{\infty} b_n \left(b_n^{-2} - b_{n+1}^{-2} \right) \\
&= \sum_{n=1}^{\infty} b_n (b_n^{-1} - b_{n+1}^{-1})(b_n^{-1} + b_{n+1}^{-1}) \\
&\leq 2 \sum_{n=1}^{\infty} (b_n^{-1} - b_{n+1}^{-1}) \quad \text{since } b_n \leq b_{n+1} \\
&\leq \frac{c}{b_1} < \infty.
\end{aligned}$$

It can now be checked that the condition (15) implies the condition (16) (see Hall and Heyde, p.158) since $I_n(\theta) = \sum_{j=1}^{n} E(u_j^2(\theta)|f_{j-1})$. Equation (13) implies that

$$\begin{aligned}
\frac{1}{I_n(\theta)} \left. \frac{d\log L_n(\theta)}{d\theta} \right|_{\theta=\theta'} &= \frac{1}{I_n(\theta)} \sum_{i=1}^{n} u_i(\theta) - (\theta' - \theta) \\
&\quad + (\theta' - \theta) \frac{J_n(\theta^\star) + I_n(\theta)}{I_n(\theta)}
\end{aligned} \tag{17}$$

Relation (17) implies that the likelihood equation

$$\frac{d\log L_n(\theta)}{d\theta} = 0$$

has a solution in $[\theta - \delta, \theta + \delta]$ a.s. if

(C1) $I_n(\theta) \overset{\text{a.s}}{\to} \infty$ as $n \to \infty$, and

(C2) $\overline{\lim\limits_{n\to\infty}} \frac{|I_n(\theta) + J_n(\theta^\star)|}{I_n(\theta)} < 1$ a.s.

Remarks: Another set of sufficient conditions for the existence of a strongly consistent root are

(C1) $I_n(\theta) \overset{\text{a.s}}{\to} \infty$ as $n \to \infty$, and

(C3) for any $\delta > 0$ such that $(\theta - \delta, \ \theta + \delta) \subset \Theta$, there exists $K(\delta) > 0$ and $h(\delta) \downarrow 0$ such that

$$\liminf_{n\to\infty} P_\theta \left\{ \sup_{|\theta' - \theta| \geq \delta} \frac{1}{I_n(\theta)} [\log L_n(\theta') - \log L_n(\theta)] < -k(\delta) \right\} \geq 1 - h(\delta).$$

Asymptotic Normality

Let us now consider the equation (13), viz.,

$$\left.\frac{d\log L_n(\theta)}{d\theta}\right|_{\theta=\theta'} = \sum_{i=1}^{n} u_i(\theta) + (\theta' - \theta)J_n(\theta^\star).$$

Let $\theta' = \hat{\theta}_n$ be a MLE. Then

$$\left.\frac{d\log L_n(\theta)}{d\theta}\right|_{\theta=\hat{\theta}_n} = 0$$

and hence

$$\sum_{i=1}^{n} u_i(\theta) = (\theta - \hat{\theta}_n)J_n(\theta')$$

where $|\theta^\star - \theta| \le |\theta - \hat{\theta}_n|$. Divide both sides by $[I_n(\theta)]^{\frac{1}{2}}$, we have

$$\frac{1}{(I_n(\theta))^{\frac{1}{2}}} \sum_{i=1}^{n} u_i(\theta) = (I_n(\theta))^{\frac{1}{2}}(\hat{\theta}_n - \theta)\left\{\frac{-J_n(\theta^\star)}{I_n(\theta)}\right\}$$

Under some conditions

$$\frac{1}{I_n^{\frac{1}{2}}(\theta)} \sum_{i=1}^{n} u_i(\theta) \xrightarrow{\mathcal{L}} N(0,1) \text{ as } n \to \infty.$$

If $\frac{J_n(\theta^\star)}{I_n(\theta)} \xrightarrow{p} -1 \text{ as } n \to \infty$, then it follows that

$$(I_n(\theta))^{\frac{1}{2}}(\hat{\theta}_n - \theta) \xrightarrow{\mathcal{L}} N(0,1) \text{ as } n \to \infty.$$

Theorem. Suppose the following conditions hold:

(C1) (i) $I_n(\theta) \xrightarrow{a.s} \infty \text{ as } n \to \infty$,
(ii) $\frac{J_n(\theta)}{EI_n(\theta)} \xrightarrow{p} \eta^2(\theta) > 0$ a.s. for some random variable $\eta(\theta)$
(iii) $\frac{J_n(\theta)}{I_n(\theta)} \xrightarrow{p} -1 \text{ as } n \to \infty$
uniformly on compact subsets of θ.

(C2) For $\delta > 0$, suppose $|\theta_n - \theta| \le \delta/(E_\theta I_n(\theta))^{\frac{1}{2}}$. Then
(i) $E_{\theta_n}\{I(\theta_n)\} = E_\theta\{I(\theta)\}(1 + o(1)) \text{ as } n \to \infty$
(ii) $I_n(\theta_n) = I_n(\theta)(1 + o(1))\text{a.s as } n \to \infty$
(iii) $J_n(\theta_n) = J_n(\theta) + o(I_n(\theta))$ a.s. as $n \to \infty$.

Then

$$\left((E_\theta I_n(\theta))^{-\frac{1}{2}}\frac{d\log L_n(\theta)}{d\theta}, \frac{I_n(\theta)}{E_\theta I_n(\theta)}\right) \xrightarrow{\mathcal{L}} (\eta(\theta)N(0,1), \eta^2(\theta))$$

where $\eta(\theta)$ and N are independent. Further more

$$\hat{\theta}_n \overset{\text{a.s}}{\to} \theta \text{ a.s } n \to \infty$$

and

$$I_n^{\frac{1}{2}}(\theta)(\hat{\theta}_n - \theta) \overset{\mathcal{L}}{\to} N(0,1) \text{ as } n \to \infty.$$

Proof: Fix $c > 0$. Let $\theta_n = \theta + c(E_\theta I_n(\theta))^{-\frac{1}{2}}$. Let

$$\wedge_n = \log\frac{L_n(\theta_n)}{L_n(\theta)}.$$

Apply Taylor's expansion:

$$\wedge_n = (\theta_n - \theta)\sum_{i=1}^{n} u_i(\theta) + \frac{1}{2}(\theta_n - \theta)^2 J_n(\theta_n^\star) \quad \text{where } |\theta_n^\star - \theta_n| \le |\theta - \theta_n|.$$

Let

$$W_n(\theta) = (E_\theta I_n(\theta))^{-\frac{1}{2}}\sum_{i=1}^{n} u_i(\theta) = \frac{(\theta_n - \theta)}{c}\sum_{i=1}^{n} u_i(\theta)$$

and

$$V_n(\theta) = -(E_\theta I_n(\theta))^{-1} J_n(\theta_n^\star) = -\frac{(\theta_n - \theta)^2}{c^2} J_n(\theta_n^\star).$$

Note that

$$\frac{L_n(\theta_n)}{L_n(\theta)} = e^{\wedge_n}$$

where $\wedge_n = cW_n(\theta) - \frac{1}{2}c^2 V_n^2(\theta)$.

In other words

$$e^{cW_n(\theta)} L_n(\theta) = e^{\frac{c^2}{2}V_n^2(\theta)} \tag{$\star$}$$

Let x_0 be a continuity point of the distribution function of $\eta(\theta)$. Assumptions (C1) and (C2) imply that

$$V_n(\theta) \overset{p}{\to} \eta^2(\theta)$$

under P_{θ_n} and

$$P_{\theta_n}(|V_n(\theta)| \le x_0) \to P_\theta(\eta^2(\theta) \le x_0).$$

Let f be a bounded continuous function on $(-\infty, \infty)$ with $f(x) = 0$ for $|x| > x_0$. Then

$$\begin{aligned}
& E_\theta\left[f(V_n(\theta))e^{cW_n(\theta)}\middle|\, |V_n(\theta)| \le x_0\right] \\
= \;& E_{\theta_n}\left[f(V_n(\theta))e^{c^2 V_n(\theta)/2}\middle|\, |V_n(\theta)| \le x_0\right] \qquad \text{from (*)} \\
\to \;& E_\theta\left[f(\eta^2(\theta))e^{c^2\eta^2(\theta)/2}\middle|\, \eta^2(\theta) \le x_0\right] \\
= \;& E_\theta\left[f(\eta_{x_0}^2(\theta))c^{c^2\eta_{x_0}^2(\theta)/2}\right]
\end{aligned}$$

where $\eta_{x_0}(\theta)$ has the distribution of $\eta(\theta)$ conditional on $\eta^2(\theta) \leq x_0$. But

$$E_\theta \left[f(\eta^2_{x_0}(\theta) e^{c^2 \eta^2_{x_0}(\theta)/2} \right] = E_\theta \left[f(\eta^2_{x_0}(\theta)) e^{c\eta_{x_0}(\theta) N(0,1)} \right]$$

where η_{x_0} is independent of N. Hence the joint distribution of $(W_n(\theta), V_n(\theta))$, conditional on $|V_n(\theta)| \leq x_0$, converges to that of $(\eta_{x_0}(\theta)N(0,1), \eta^2_{x_0}(\theta))$. Let $x_0 \to \infty$. We obtain that

$$g\left((E_\theta I_n(\theta))^{-\frac{1}{2}} \frac{d\log L_n(\theta)}{d\theta}, \frac{I_n(\theta)}{E_\theta(I_n(\theta))} \right) \xrightarrow{\mathcal{L}} g\left(\eta(\theta)N(0,1), \eta^2(\theta)\right)$$

by the continuous mapping theorem.

Remarks: If

$$\frac{I_n(\theta)}{E_\theta I_n(\theta)} \xrightarrow{p} \eta^2(\theta) > 0 \text{ as } n \to \infty,$$

then one can replace the random norming $I_n(\theta)$ by the non-random norming $E_\theta I_n(\theta)$ and we obtain that

$$(E_\theta I_n(\theta))^{\frac{1}{2}} (\hat{\theta}_n - \theta) \xrightarrow{\mathcal{L}} \eta(\theta) N(0,1)$$

where $\eta(\theta)$ and N are independent.

Definition: An estimator T_n of θ is said to be *asymptotically first order efficient* if

$$I_n^{\frac{1}{2}}(\theta) \left[T_n - \theta - r(\theta) I_n^{-1}(\theta) \frac{d\log L_n(\theta)}{d\theta} \right] \xrightarrow{p} 0$$

as $n \to \infty$ for some $r(\theta)$ not depending on n or the observations.

Remarks: It can be checked that the MLE is asymptotically first order efficient in the above sense under the conditions stated above.

Lecture 5

Note that

$$\left.\frac{d\log L_n(\theta)}{d\theta}\right|_{\theta=\theta'} = \sum_{i=1}^{n} u_i(\theta) + (\theta'-\theta)J_n(\theta^*)$$
$$= \sum_{i=1}^{n} u_i(\theta) - (\theta'-\theta)I_n(\theta) + (\theta'-\theta)(J_n(\theta^*)+I_n(\theta)).$$

Suppose that $J_n(\theta^*)+I_n(\theta)=0$ a.s. Then

$$\left.\frac{d\log L_n(\theta)}{d\theta}\right|_{\theta=\theta'} = \sum_{i=1}^{n} u_i(\theta) - (\theta'-\theta)I_n(\theta). \qquad (\alpha)$$

Substituting $\theta'=\hat{\theta}_n$, we have

$$0 = \left.\frac{d\log L_n(\theta)}{d\theta}\right|_{\theta=\hat{\theta}_n} = \sum_{i=1}^{n} u_i(\theta) - (\hat{\theta}_n-\theta)I_n(\theta) \qquad (\beta)$$

Subtracting (β) from (α), we have

$$\left.\frac{d\log L_n(\theta)}{d\theta}\right|_{\theta'} = (\hat{\theta}_n-\theta')I_n(\theta)$$

and in general

$$\frac{d\log L_n(\theta)}{d\theta} = (\hat{\theta}_n-\theta)I_n(\theta).$$

Special case (Conditional exponential family:)

Suppose

(18) $$\frac{d\log L_n(\theta)}{d\theta} = I_n(\theta)(\hat{\theta}_n-\theta), \theta\in\Theta, n\geq 1.$$

Then $\hat{\theta}_n$ is the MLE. Differentiating with respect to θ on both sides of the equation (18), we obtain that

(19) $$\frac{d^2\log L_n(\theta)}{d\theta^2} = I_n'(\theta)(\hat{\theta}_n-\theta) - I_n(\theta)$$

and

$$E_\theta\left(\left.\frac{d^2\log L_n(\theta)}{d\theta^2}\right|\mathcal{F}_{n-1}\right) = I_n'(\theta)E(\hat{\theta}_n-\theta\,|\,\mathcal{F}_{n-1}) - I_n(\theta)$$
$$= I_n'(\theta)E_\theta\left(\left.\frac{d\log L_n(\theta)}{d\theta}\frac{1}{I_n(\theta)}\right|\mathcal{F}_{n-1}\right) - I_n(\theta)$$

(20) (from (18))

$$= \frac{I_n'(\theta)}{I_n(\theta)}\frac{d\log L_{n-1}(\theta)}{d\theta} - I_n(\theta)$$

(21) (by the martingale property)

But

$$\begin{aligned} & E_\theta\left(\left.\frac{d^2\log L_n(\theta)}{d\theta^2}\right| \mathcal{F}_{n-1}\right) \\ = \ & E_\theta\left(\left.\frac{d^2\log L_{n-1}(\theta)}{d\theta^2} + \frac{d^2\log L_n(\theta)}{d\theta^2} - \frac{d^2\log L_{n-1}(\theta)}{d\theta^2}\right| \mathcal{F}_{n-1}\right) \\ = \ & \frac{d^2\log L_{n-1}(\theta)}{d\theta^2} + E_\theta\left(\left.\frac{d^2}{d\theta^2}\{\log L_n(\theta) - \log L_{n-1}(\theta)\}\right| \mathcal{F}_{n-1}\right) \\ = \ & \frac{d^2\log L_{n-1}(\theta)}{d\theta^2} + E_\theta\left(v_n(\theta) | \mathcal{F}_{n-1}\right) \\ = \ & \frac{d^2\log L_{n-1}(\theta)}{d\theta^2} - E_\theta\left(u_n^2(\theta) | \mathcal{F}_{n-1}\right) \\ = \ & \frac{d^2\log L_{n-1}(\theta)}{d\theta^2} - (I_n(\theta) - I_{n-1}(\theta)). \end{aligned} \tag{22}$$

Relations (20) and (21) imply that

$$\frac{I_n'(\theta)}{I_n(\theta)}\frac{d\log L_{n-1}(\theta)}{d\theta} - I_n(\theta) = \frac{d^2\log L_{n-1}(\theta)}{d\theta^2} - (I_n(\theta) - I_{n-1}(\theta)).$$

Hence

$$\frac{I_n'(\theta)}{I_n(\theta)} = \frac{\frac{d^2\log L_{n-1}(\theta)}{d\theta^2} + I_{n-1}(\theta)}{\frac{d\log L_{n-1}(\theta)}{d\theta}}. \tag{23}$$

Relations (18) and (19) imply that

$$\frac{d^2\log L_n(\theta)}{d\theta^2} = \frac{I_n'(\theta)}{I_n(\theta)}\frac{d\log L_n(\theta)}{d\theta} - I_n(\theta)$$

and hence

$$I_n(\theta)\frac{d^2\log L_n(\theta)}{d\theta^2} = I_n'(\theta)\frac{d\log L_n(\theta)}{d\theta} - I_n^2(\theta)$$

which implies that

$$I_n(\theta)\{\frac{d^2\log L_n(\theta)}{d\theta^2} + I_n(\theta)\} = I_n'(\theta)\frac{d\log L_n(\theta)}{d\theta}.$$

Therefore

$$\frac{I_n'(\theta)}{I_n(\theta)} = \frac{\frac{d^2\log L_n(\theta)}{d\theta^2} + I_n(\theta)}{\frac{d\log L_n(\theta)}{d\theta}}. \tag{24}$$

Comparing (22) and (23), we obtain that

$$\frac{I_n'(\theta)}{I_n(\theta)} = C(\theta) \qquad \text{for all } n$$

for some $C(\theta)$. This implies that

$$I_n(\theta) = \phi(\theta) H_n(X_1, \ldots, X_{n-1})$$

since $I_n(\theta)$ is $\mathcal{F}_{n-1}$-measurable for some function $\phi(\theta)$. Therefore, from the equation (18), it follows that

$$\frac{d\log L_n(\theta)}{d\theta} = \phi(\theta)(\hat{\theta}_n - \theta) H_n(X_1, \ldots, X_{n-1}) \tag{25}$$

which implies that

$$\log L_n(\theta) = H_n(X_1, \ldots, X_{n-1}) \left\{ \int \phi(\theta) d\theta \, \hat{\theta}_n - \int \phi(\theta) \theta d\theta \right\} + K_n(X_1, \ldots, X_n).$$

By the factorization theorem, it follows that $(\hat{\theta}_n, H_n(X_1, \ldots, X_{n-1}))$ is a sufficient statistic for θ. Furthermore

$$L_n(\theta) = \exp\{H_n(X_1, \ldots, X_{n-1})(r_1(\theta)\hat{\theta}_n + r_2(\theta)) + K_n(X_1, \ldots, X_n)\}.$$

Special case of Markov process:

Suppose $\{X_n, n \geq 0\}$ is a time-homogenous Markov process. Suppose the conditional probability (density) function of X_n given X_{n-1} is $f(x_n | x_{n-1}, \theta)$. Then the likelihood function of $(x_1, \ldots, x_n)$ is given by

$$L_n^\star(x_1, \ldots, x_n; \theta) \equiv L_n^\star(\theta) = g(x_0) \prod_{i=1}^n f(x_i | x_{i-1}, \theta)$$

(we assume that the initial density $g(.)$ of X_0 does not depend on θ). Since X_0 does not have information about the parameter θ, let us consider the likelihood function to be

$$L_n(\theta) \equiv \prod_{i=1}^n f(x_i | x_{i-1}, \theta).$$

Hence

$$\begin{aligned} \frac{d\log L_n(\theta)}{d\theta} &= \sum_{i=1}^n \frac{d}{d\theta} \log f(x_i | x_{i-1}, \theta) \\ &= \sum_{i=1}^n u_i(\theta), \qquad u_i(\theta) = \frac{d}{d\theta} \log f(x_i | x_{i-1}, \theta). \end{aligned}$$

Suppose that

$$\frac{d\log L_n(\theta)}{d\theta} = I_n(\theta)(\hat{\theta}_n - \theta). \tag{26}$$

In particular, for $n = 1$, we have from (25),

$$\frac{d}{d\theta}\text{log}f(X_1|X_o,\theta) = I_1(\theta)(\hat{\theta}_1 - \theta)$$
$$= \phi(\theta)H(X_0)(\hat{\theta}_1 - \theta).$$

Note that $\hat{\theta}_1$ depends on X_0 and X_1 and it is a solution of the equation

$$\frac{d}{d\theta}\text{log}f(X_1|X_0,\theta) = 0.$$

Let $\hat{\theta}_1 = m(x,y)$ be the solution of the equation

$$\frac{d}{d\theta}\text{log}f(y|x,\theta) = 0. \quad (25a)$$

Then

$$\frac{d}{d\theta}\text{log}f(y|x,\theta) = \phi(\theta)H(x)(m(x,y) - \theta) \quad (25b)$$

and hence

$$\text{log}f(y|x,\theta) = H(x)m(x,y)\int\phi(\theta)d\theta - H(x)\int\theta\phi(\theta)d\theta + K(x,y)$$

or equivalently

(27) $$f(y|x,\theta) = [\exp\{H(x)m(x,y)J_1(\theta) - H(x)J_2(\theta)\}]K^\star(x,y).$$

Such a family of distributions is called a *Conditional exponential family.* Relation (25b) implies that

(28) $$\frac{d\text{log}L_n(\theta)}{d\theta} = \phi(\theta)\sum_{i=1}^{n} H(X_{i-1})[m(X_{i-1},X_i) - \theta]$$

and

$$u_i(\theta) = \frac{d}{d\theta}\text{log}f(X_i|X_{i-1},\theta) = \phi(\theta)H(X_{i-1})[m(X_{i-1},X_i) - \theta].$$

Hence

$$E_\theta[u_i(\theta)|\mathcal{F}_{i-1}] = \phi(\theta)H(X_{i-1})[E_\theta(m(X_{i-1},X_i)|\mathcal{F}_{i-1}) - \theta]$$
$$= 0 \text{ a.s}$$

by earlier remarks.Therefore

(29) $$E_\theta[m(X_{i-1},X_i)|\mathcal{F}_{i-1}] - \theta = 0 \text{ a.s.}$$

It also follows from (27) that

(30) $$\hat{\theta}_n = [\sum_{i=1}^{n} H(X_{i-1})]^{-1}[\sum_{i-=1}^{n} H(X_{i-1})m(X_{i-1},X_i)].$$

Note that

$$E_\theta(u_i^2(\theta)|\mathcal{F}_{i-1}) = -E_\theta\left(\frac{d^2}{d\theta^2}\log f(X_i\,|X_{i-1},\theta)|\,\mathcal{F}_{i-1}\right).$$

But

$$\frac{d^2\log f(X_i|X_{i-1},\theta)}{d\theta^2} = \phi'(\theta)H(X_{i-1})[m(X_{i-1},X_i)-\theta] + \phi(\theta)H(X_{i-1})(-1).$$

Hence

$$-E_\theta\left(\frac{d^2\log f(X_i|X_{i-1};\theta)}{d\theta^2}\bigg|\,\mathcal{F}_{i-1}\right) = \phi(\theta)H(X_{i-1}) \tag{31}$$

from (28). Hence

$$\begin{aligned} I_n(\theta) &= \sum_{i=1}^n E_\theta(u_i^2(\theta)|\mathcal{F}_{i-1}) \\ &= \sum_{i=1}^n \phi(\theta)H(X_{i-1}) = \{\sum_{i=1}^n H(X_{i-1})\}\phi(\theta). \end{aligned} \tag{32}$$

Using the equations (29) and (31), it can be checked that

$$\frac{d\log L_n(\theta)}{d\theta} = I_n(\theta)(\hat{\theta}_n - \theta) \tag{33}$$

from (27). In other words, the relation (32) is a necessary and sufficient condition for the transition probability (density) function to belong to a conditional exponential family.

Lecture 6

Bienayme - Galton - Watson Branching process (Guttorp (1991))

Let $\{X_{ij}, i \geq 1, j \geq 1\}$ be i.i.d. random variables taking values in the nonnegative integers with the probability generating function (p.g.f.),

$$g(s) = \sum_{k=0}^{\infty} s^k p_k, -1 < s \leq 1.$$

Let X be a random variable with $P(X = k) = p_k, \quad k = 0, 1, 2, \ldots$ The distribution of X is called the offspring distribution. We define the branching process $\{Z_k, k \geq 0\}$ with the offspring distribution $\{p_k, k \geq 0\}$ recursively by

$$Z_0 = z_0, \quad Z_k = \sum_{i=1}^{Z_{k-1}} X_{ik}. \tag{34}$$

We will assume that $z_0 = 1$ and $p_0 + p_1 < 1$ in the following discussion.

The conditional p.g.f. of Z_k given $Z_{k-1} = z$ is given by

$$E[s^{Z_k} | Z_{k-1} = z] = \{g(s)\}^z$$

due to the independence of the random variables $\{X_{ij}\}$ and hence the p.g.f. of Z_k is

$$g_k(s) = E[g(s)^{Z_{k-1}}] = g_{k-1}(g(s)).$$

This implies that $E(Z_k) = \theta^k$ provided $\theta = E(X_{ij}) < \infty$, and

$$V(Z_k) = \sigma^2 \theta^{k-1} \sum_{j=0}^{k-1} \theta^j \text{ provided } \sigma^2 = V(X_{ij}) < \infty.$$

Remarks: Note that once a generation is extinct, all the following generations will be extinct as well. Extinction will occur in or before the k-th generation in one of the following ways : the ancestor has
(i) 0 children
(ii) one child whose family becomes extinct in or before $(k-1)$-th generation
(iii) two children both of whose families become extinct in or before $(k-1)$-th generation and so on.

The probability q_k of extinction after k generations is

$$\begin{aligned} q_k &= P(Z_k = 0) = \sum_{j=0}^{\infty} p_j g_{k-1}^j(0) \\ &= g(g_{k-1}(0)) = g(q_{k-1}). \end{aligned} \tag{35}$$

Assume that $p_0 > 0$ and that $p_1 > 0$. Then the function $g(.)$ is a strictly increasing function on the interval $[0,1]$. Hence the sequence $\{q_k\} = \{g(q_{k-1})\}$ forms a strictly increasing sequence of positive numbers bounded by one. Hence $\{q_k\}$ has a limit, say q, with $p_0 \leq q \leq 1$. Taking limits in (34) we get that

$$q = g(q). \tag{36}$$

Hence

$$\frac{g(q) - g(q_k)}{q - q_k} = \frac{q - q_{k+1}}{q - q_k} < 1.$$

Let $k \to \infty$. We observe that $g'(q) \leq 1$. Note that $g'(s)$ is a strictly increasing function on $(0,1)$ and hence $g(s)$ is convex.

Proposition : The extinction probability q is the smallest nonnegative root of the equation $g(s) = s$. If $\theta > 1$, then $0 \leq q < 1$ with equality occurring if and only if $p_0 = 0$. If $\theta \leq 1$, then $q = 1$ unless $p_1 = 1$ when $q = 0$.

Example: Suppose the offspring distribution is geometric. Then

$$p_k = p(1-p)^k, k = 0, 1, \ldots$$

and hence

$$g(s) = \frac{p}{1 - (1-p)s}.$$

Therefore $\theta = g'(1) = \frac{1-p}{p}$. The process becomes extinct with probability one if $p > \frac{1}{2}$. If $p < \frac{1}{2}$, then the extinction probability q is a solution of the equation

$$\frac{p}{1 - (1-p)s} = s$$

with solutions 1 and $\frac{1}{\theta}$. Hence $q = \frac{1}{\theta}$.

Proposition: If $p_1 \neq 1$, then $Z_n \overset{a.s.}{\to} \infty$ as $n \to \infty$ with probability $1 - q$.

Proof: If $q = 1$, there is nothing to prove. Suppose $q < 1$. Note that

$$g_k(s) = g_{k-1}(g(s))$$

and hence

$$g_k'(s) = g_{k-1}'(g(s))g'(s)$$

If $s = q$, then $g(s) = s$ and hence

$$g_k'(q) = g_{k-1}'(q)g'(q) \quad \text{for all } k \geq 1.$$

Hence

$$g_k'(q) = [g'(q)]^k \quad \text{for all } k \geq 1.$$

Furthermore

$$\begin{aligned} P(1 \leq Z_n \leq k) &\leq \sum_{j=1}^{k} P(Z_n = j) \\ &\leq \sum_{j=1}^{k} P(Z_n = j) \frac{jq^{j-1}}{q^k} \quad \text{since } q < 1 \\ &\leq \frac{g_n'(q)}{q^k} = \frac{[g'(q)]^n}{q^k} \end{aligned}$$

Therefore

$$\sum_{n=1}^{\infty} P(1 \leq Z_n \leq k) \leq \frac{1}{q^k} \sum_{n=1}^{\infty} [g'(q)]^n < \infty$$

provided $g'(q) < 1$. Hence, by the Borel-Cantelli lemma, it follows

$$P(1 \leq Z_n \leq k \text{ infinitely often}) = 0 \text{ for any } k$$

that is

$$Z_n \to \infty \text{ a.s as } n \to \infty.$$

(Ref: Guttorp "Statistical Inference for Branching process").

As pointed out earlier, we assume that $Z_0 \equiv 1$ in the following discussion.

Theorem: Suppose that $\sigma^2 = \text{Var} Z_1 < \infty$ and $EZ_1 = \theta$. Let $W_n = \frac{Z_n}{\theta^n}$ and $\mathcal{F}_n = \sigma(Z_0, \ldots, Z_n)$. Then

(i) $\{W_n, \mathcal{F}_n, n \geq 0\}$ is a martingale;

(ii) $W_n \to W$ a.s. where $P(W = 0) = q$ where q is the smallest nonnegative root of the equation $g(s) = s$ (here $g(.)$ is the probability generating function of the offspring distribution, that is, $g(s) = \sum_{j=0}^{\infty} s^j p_j$);

(iii) $\{W > 0\} = \{Z_n \to \infty\}$ a.s;

(iv) if $\theta > 1$, then $EW = 1$ and $V(W) = \frac{\sigma^2}{\theta(\theta-1)}$;

(v) the m.g.f. $\phi(s) = E[e^{-sW}]$ satisfies the equation $\phi(\theta s) = g(\phi(s))$, and

(vi) if $\theta > 1$ and $\sigma^2 < \infty$, then the distribution of W is absolutely continuous except for a jump of size q at 0.

Special case : Suppose a random variable X has the offspring distribution

$$P(X=j)=p_j=\frac{a_j\lambda^j}{f(\lambda)}, j\geq 0.$$

Then

$$E(X)=\sum_{j=0}^{\infty} jp_j=\sum_{j=0}^{\infty} j\frac{a_j\lambda^j}{f(\lambda)}=\sum_{j=1}^{\infty} j\frac{a_j\lambda^j}{f(\lambda)}=\lambda\sum_{j=1}^{\infty}\frac{ja_j\lambda^{j-1}}{f(\lambda)}=\frac{\lambda f'(\lambda)}{f(\lambda)}.$$

Note that

$$f(\lambda)=\sum_{j=0}^{\infty} a_j\lambda^j,$$

and

$$f'(\lambda)=\sum_{j=1}^{\infty} a_j j\lambda^{j-1}, \qquad f''(\lambda)=\sum_{j=2}^{\infty} a_j j(j-1)\lambda^{j-2}.$$

Hence

$E(X)=\frac{\lambda f'(\lambda)}{f(\lambda)}=\theta$ and $\frac{d\theta}{d\lambda}=\frac{f(\lambda)[\lambda f''(\lambda)+f'(\lambda)-\lambda f'^2(\lambda)]}{f^2(\lambda)}.$

Further more

$$\sigma^2=\mathrm{Var}(X)=E(X^2)-(E(X))^2=E(X(X-1))+EX-(EX)^2$$

Now

$$\begin{aligned}E(X(X-1)) &= \sum_{j=0}^{\infty} j(j-1)p_j=\sum_{j=2}^{\infty} j(j-1)\frac{a_j\lambda^j}{f(\lambda)}\\ &= \frac{\lambda^2 f''(\lambda)}{f(\lambda)}.\end{aligned}$$

Hence

$$\begin{aligned}\sigma^2 &= \frac{\lambda^2 f''(\lambda)}{f(\lambda)}+\frac{\lambda f'(\lambda)}{f(\lambda)}-\frac{\lambda^2 f'^2(\lambda)}{f^2(\lambda)}=\frac{\lambda^2 f''(\lambda)f(\lambda)+\lambda f'(\lambda)f(\lambda)-\lambda^2(f'(\lambda))^2}{f^2(\lambda)}\\ &= \lambda\left\{\frac{\lambda f''(\lambda)f(\lambda)+f'(\lambda)f(\lambda)-\lambda f'^2(\lambda)}{f^2(\lambda)}\right\}=\lambda\frac{d\theta}{d\lambda}.\end{aligned}$$

Example :(Branching process)(Bienayme-Galton-Watson process)

Let $Z_0, Z_1, \ldots, Z_n, \ldots$ be the consecutive generation sizes with $Z_0=1$.
Let $\theta=E(Z_1)$. Suppose that $1<\theta<\infty$. Assume that $\sigma^2=\mathrm{Var}(Z_1)<\infty$. Let

$$p_j=P(Z_1=j), \quad =0,1,2,\ldots$$

Assume that $\{p_j\}$ belongs to a family of power series distributions as discussed above. Then

$$p_j = \frac{a_j\lambda^j}{f(\lambda)}, j = 0, 1, 2, \ldots \text{ where } \lambda > 0 \text{ is fixed constant,}$$

$a_j \geq 0$ and $f(\lambda) = \sum_{j=0}^{\infty} a_j\lambda^j$. We have noted that

$$\theta = \frac{\lambda f'(\lambda)}{f(\lambda)} \quad \sigma^2 = \lambda\frac{d\theta}{d\lambda}.$$

We assume that $\theta > 1$. Then $Z_n \to \infty$ with probability $1-q$ where q is the probability of extinction (*Biometrika 62,* 49-59 (1975)). Let $p(x|y, \lambda)$ be the transition probability function of the process $\{Z_k, k \geq 1\}$. It can be checked that

$$\frac{d}{d\theta}\log p(x|y, \lambda) = \sigma^{-2}(x - \theta y)$$

and

$$I_n(\theta) = \sigma^{-2}\sum_{i=0}^{n-1} Z_i.$$

Note that $\{Z_0 = 1, \ Z_1, \ldots, Z_n\}$ is a realization of a Markov chain with the transition probabilities

$$p(Z_k|Z_{k-1}, \lambda) \propto \lambda^{Z_k}\{f(\lambda)\}^{-Z_{k-1}}$$

and hence the likelihood function

$$L_n(\lambda) = \Pi_{k=1}^n p(Z_k|Z_{k-1}, \lambda) \propto \{f(\lambda)\}^{-\sum_{k=1}^n Z_{k-1}}\lambda^{\sum_{k=1}^n Z_k}.$$

Therefore

$$\begin{aligned}
\frac{d\text{log}L_n(\lambda)}{d\lambda} &= \frac{d}{d\lambda}\left\{(-\sum_{k=1}^n Z_{k-1})\text{log}f(\lambda) + (\sum_{k=1}^n Z_k)\text{log}\lambda\right\} \\
&= (-\sum_{k=1}^n Z_{k-1})\frac{f'(\lambda)}{f(\lambda)} + (\sum_{k=1}^n Z_k)\frac{1}{\lambda} \\
&= \frac{1}{\lambda}\left[(-\sum_{k=1}^n Z_{k-1})\frac{\lambda f'(\lambda)}{f(\lambda)} + (\sum_{k=1}^n Z_k)\right].
\end{aligned}$$

Hence

$$\frac{d\text{log}L_n(\lambda)}{d\lambda} = \frac{1}{\lambda}\left[(\sum_{k=1}^n Z_k) - \theta(\sum_{k=1}^n Z_{k-1})\right] = 0$$

provided

$$\hat{\theta} = \frac{\sum_{k=1}^n Z_k}{\sum_{k=1}^n Z_{k-1}}.$$

Note that $\sigma^2 = \lambda \frac{d\theta}{d\lambda}$ and hence $\frac{d\theta}{d\lambda} = \frac{\sigma^2}{\lambda} > 0$. Therefore $\theta(.)$ is a strictly increasing function of λ and we can reparametrize the problem through θ. Observe that

$$\begin{aligned}
\frac{d\log L_n(\lambda)}{d\theta} &= \frac{d\log L_n(\lambda)}{d\lambda}\frac{d\lambda}{d\theta} = \frac{d\log L_n(\lambda)}{d\lambda}\frac{\lambda}{\sigma^2} \\
&= \frac{1}{\sigma^2}\left[(\sum_{k=1}^{n} Z_k) - \theta(\sum_{k=1}^{n} Z_{k-1})\right] = \frac{1}{\sigma^2}\sum_{k=1}^{n}(Z_k - \theta Z_{k-1}) \\
&= \sum_{k=1}^{n} u_k(\theta), \quad u_k(\theta) = \frac{Z_k - \theta Z_{k-1}}{\sigma^2}.
\end{aligned}$$

Note that $E[u_k(\theta)|Z_{k-1}] = 0$ since $E(Z_k|Z_{k-1}) = \theta Z_{k-1}$ and the conditional information is given by

$$\begin{aligned}
I_n(\theta) &= \sum_{k=1}^{n} E[u_k^2(\theta)|\mathcal{F}_{k-1}] = -\sum_{k=1}^{n} E[\frac{du_k(\theta)}{d\theta}|\mathcal{F}_{k-1}] \\
&= \sum_{k=1}^{n}\frac{Z_{k-1}}{\sigma^2} = \frac{1}{\sigma^2}\sum_{k=1}^{n} Z_{k-1}.
\end{aligned}$$

Note that $\zeta_n(\theta) \equiv E(I_n(\theta)) = \frac{1}{\sigma^2}\sum_{k=1}^{n}\theta^{k-1} = \frac{1}{\sigma^2}\frac{\theta^n - 1}{\theta - 1}$. It is easy to see that $\{\frac{Z_n}{\theta^n}, \mathcal{F}_n, n \geq 0\}$ is a martingale since

$$E\left(\frac{Z_n}{\theta^n}\middle| \mathcal{F}_{n-1}\right) = \frac{Z_{n-1}}{\theta^{n-1}}$$

and it is a nonnegative martingale. Hence

$$W_n \equiv \frac{Z_n}{\theta^n} \stackrel{\text{a.s}}{\to} W \text{ (say) } n \to \infty$$

where $W \geq 0$ a.s. We now apply the Toeplitz Lemma (Loéve (1963)), viz,

$$x_n \to x \Rightarrow \frac{1}{\sigma_n}\sum_{k=0}^{n} a_k x_k \to x \text{ if } \sigma_n = \sum_{k=0}^{n} a_k \uparrow \infty \text{ as } n \to \infty.$$

Note that

$$\begin{aligned}
\hat{\theta_n} &= \frac{\sum_{k=1}^{n} Z_k}{\sum_{k=1}^{n} Z_{k-1}} \\
&= \frac{\sum_{k=0}^{n} Z_k - Z_0}{\sum_{k=0}^{n-1} Z_k} = \frac{\sum_{k=0}^{n} Z_k}{\sum_{k=0}^{n-1} Z_k} - \frac{1}{\sum_{k=0}^{n-1} Z_k} \\
&\simeq \frac{\sum_{k=0}^{n} Z_k}{\sum_{k=0}^{n-1} Z_k} \text{ on } [W > 0]
\end{aligned}$$

and hence

$$\begin{aligned}
\hat{\theta_n} = \frac{\sum_{k=0}^{n} Z_k}{\sum_{k=0}^{n-1} Z_k} &= \frac{\sum_{k=0}^{n} Z_k}{\sum_{k=0}^{n}\theta^k}.\frac{\sum_{k=0}^{n}\theta^k}{\sum_{k=0}^{n-1}\theta^k}.\frac{\sum_{k=0}^{n-1}\theta^k}{\sum_{k=0}^{n-1} Z_k} \\
&\to W \lim_{n\to\infty}\frac{\theta^{n+1} - 1}{\theta^n - 1}.\frac{1}{W} = \theta \text{ whenever } W > 0.
\end{aligned}$$

Therefore

$$\hat{\theta}_n = \frac{\sum_{k=1}^n Z_k}{\sum_{k=1}^n Z_{k-1}} \to \theta \quad \text{a.s. on the set } [W > 0]$$

This proves that strong consistency of the estimator on the set $[W > 0]$. The strong consistency might not hold on the set $[W = 0]$ which might have positive probability. Furthermore

$$\begin{aligned} \frac{I_n(\theta)}{\zeta_n(\theta)} &= \frac{\sum_{k=1}^n Z_{k-1}}{\sum_{k=1}^n \theta^{k-1}} \\ &= \left\{ \sum_{k=1}^n \theta^{k-1} . \frac{Z_{k-1}}{\theta^{k-1}} \right\} \frac{1}{\sum_{k=1}^n \theta^{k-1}} \\ &\to W \text{ a.s.} \quad \text{as } n \to \infty. \end{aligned}$$

Note that

$$\begin{aligned} (I_n(\theta))^{-\frac{1}{2}} \frac{d \log L_n(\theta)}{d\theta} &= \left(\frac{1}{\sigma^2} \sum_{k=1}^n Z_{k-1} \right)^{-\frac{1}{2}} \sum_{k=1}^n \frac{Z_k - \theta Z_{k-1}}{\sigma^2} \\ &= \frac{\sum_{k=1}^n u_k(\theta)}{\{\sum_{k=1}^n E(u_k^2(\theta)|\mathcal{F}_{k-1})\}^{\frac{1}{2}}} \\ &\xrightarrow{\mathcal{L}} Z \sim N(0,1) \text{ as } n \to \infty \end{aligned}$$

by the martingale CLT. However

$$\frac{I_n(\theta)}{\zeta_n(\theta)} \to W \text{ a.s. as } n \to \infty.$$

Note that

$$\begin{aligned} & (I_n(\theta))^{-\frac{1}{2}} \frac{d \log L_n(\theta)}{d\theta} \\ &= \left(\frac{1}{\sigma^2} \sum_{k=1}^n Z_{k-1} \right)^{\frac{1}{2}} \frac{\sum_{k=1}^n (Z_k - \theta Z_{k-1})}{\sum_{k=1}^n Z_{k-1}} \\ &= (I_n(\theta))^{\frac{1}{2}} (\hat{\theta}_n - \theta) \end{aligned}$$

and

$$(I_n(\theta))^{-\frac{1}{2}} \frac{d \log L_n(\theta)}{d\theta} = (I_n(\theta))^{\frac{1}{2}} (\hat{\theta}_n - \theta).$$

Hence

$$(I_n(\theta))^{\frac{1}{2}} (\hat{\theta}_n - \theta) \xrightarrow{\mathcal{L}} N(0,1) \text{ as } n \to \infty \qquad \textit{(Random norming)}$$

and

$$(\zeta_n(\theta))^{\frac{1}{2}} (\hat{\theta}_n - \theta) \xrightarrow{\mathcal{L}} W^{\frac{1}{2}} N(0,1) \text{ as } n \to \infty \qquad \textit{(Nonrandom norming)}$$

where $\zeta_n(\theta) = E[I_n(\theta)]$, and W and N are independent. In other words the asymptotic distribution of the maximum likelihood estimator is not normal. Such models are called *non-ergodic models.*

Special case: Suppose that $\theta > 1$ and

$$(\star) \qquad p_j = P(Z_1 = j|Z_0 = 1) = \frac{1}{\theta}(1 - \frac{1}{\theta})^{j-1}, j = 1, 2, \ldots;$$

that is the off-spring distribution is geometric. Then $E(X_1) = \theta$ and the probability of extinction $q = 0$. Furthermore

$$\frac{I_n(\theta)}{\zeta_n(\theta)} = \frac{\sigma^{-2}\sum_{i=0}^{n-1} Z_i}{\sigma^{-2}\sum_{i=0}^{n-1}\theta^i} = \frac{\sum_{i=0}^{n-1} Z_i}{\sum_{i=0}^{n-1}\theta^i} \overset{\text{a.s.}}{\to} W \text{ as } n \to \infty$$

where W is standard exponential. In fact $\phi(s) = E[e^{-sW}]$ satisfies the equation

$$(\star\star) \qquad \phi(\theta s) = \frac{\frac{1}{\theta}\phi(s)}{1 - (1 - \frac{1}{\theta})\phi(s)}.$$

A solution of the equation ($\star\star$) is $\phi(s) = \frac{\lambda}{\lambda+s}$ where $\lambda > 0$. Since $E[W] = 1$, it follows that $\lambda = 1$ and W is exponential with mean 1.

Bayesian Estimation: Suppose the off-spring distribution is Poisson with mean θ. Further assume that the parameter θ has a prior density which is Gamma with the parameters α and β, that is,

$$\begin{aligned} p(\theta) &= \frac{e^{-\theta\beta}\theta^{\alpha-1}\beta^{\alpha}}{\Gamma(\alpha)}, \quad 0 < \theta < \infty \\ &= 0 \ \text{ otherwise} \end{aligned}$$

where $\alpha > 0$ and $\beta > 0$ are known. We have seen that the likelihood function $L_n(\theta)$ is proportional to

$$\exp(-\theta\sum_{k=1}^{n} Z_{k-1})\theta^{\sum_{k=1}^{n} Z_k}.$$

Hence the posterior density of θ , given $(Z_0, \ldots, Z_n)$, is proportional to

$$\exp(-\theta(\beta + \sum_{k=1}^{n} Z_{k-1}))\theta^{\alpha+\sum_{k=1}^{n} Z_k - 1}.$$

Therefore the posterior density of the parameter θ is again Gamma with the parameters $\alpha+\sum_{k=1}^{n} Z_k$ and $\beta+\sum_{k=1}^{n} Z_{k-1}$. The mean of the posterior density is the Bayes estimator of the parameter θ under the quadratic loss function. It is given by

$$\tilde{\theta}_n = \frac{\alpha + \sum_{k=1}^{n} Z_k}{\beta + \sum_{k=1}^{n} Z_{k-1}}.$$

It can be checked that $\tilde{\theta}_n$ is asymptotically equivalent to the MLE

$$\hat{\theta}_n = \frac{\sum_{k=1}^{n} Z_k}{\sum_{k=1}^{n} Z_{k-1}}$$

on the set of non-extinction, that is, on the set $[W > 0]$.

Least squares approach: Let us again consider the BGW branching process process as discussed earlier. We have seen that

$$E(Z_{n+1}|Z_n) = Z_n\theta$$

and

$$Var(Z_{n+1}|Z_n) = Z_n\sigma^2.$$

Let U_{n+1} be defined by the relation

$$(*)Z_{n+1} = \theta Z_n + Z_n^{1/2}U_{n+1}, n \geq 0, Z_0 = 1.$$

Check that (i) $E(U_k) = 0, k \geq 1$ (ii) $Var(U_k) = \sigma^2, k \geq 1$,and (iii) $E(U_kU_j) = 0, 1 \leq j \leq k-1, E(U_kZ_{k-1}) = 0, k \geq 1.$

The relation (*) is an autoregressive type model for the process $\{Z_k, k \geq 0\}$. Since the error terms in (*) satisfy classical assumptions in the theory of least squares, we may consider the least squares approach for the estimation of the parameter θ. This is done by minimizing the error sum of square $\sum_{k=1}^{n} U_k^2$ with respect to θ. This gives the estimator

$$\theta_n^* = \frac{\sum_{k=1}^{n} Z_k}{\sum_{k=1}^{n} Z_{k-1}}$$

which is the same as the MLE if the off-spring distribution is the power series distribution. The variance σ^2 can be estimated by the residual sum of squares, namely,

$$\sigma^{2*} = \frac{1}{n}\sum_{k=1}^{n}\{\frac{(Z_k - \theta_n^* Z_{k-1})^2}{Z_{k-1}}\}.$$

Lecture 7

Estimation by conditional least squares

Let $\{X_n, n \geq 1\}$ be a stochastic process defined on a probability space $(\Omega, \mathcal{F}, P_\theta)$, $\theta = (\theta_1, \ldots, \theta_p) \in \Theta \subset R^p$, Θ open. Consider

$$Q_n(\theta) = \sum_{k=1}^{n} \left[X_k - E_\theta(X_k|\mathcal{F}_{k-1})\right]^2.$$

We estimate θ by minimizing $Q_n(\theta)$ over Θ. We assume that $Q_n(\theta)$ has partial derivatives with respect to $\theta_i, 1 \leq i \leq p$.

Assume that $E_\theta(X_n|\mathcal{F}_{n-1})$ is a.s. twice continuously differentiable with respect to θ in some neighbourhood S of the true parameter (say) $\theta^\circ = (\theta_1^\circ, \ldots, \theta_p^\circ) \in \Theta$ open. Applying the Taylor series expansion, we have

$$\begin{aligned} Q_n(\theta) = Q_n(\theta^\circ) \quad &+ \quad (\theta - \theta^\circ)' \left.\frac{\partial Q_n(\theta)}{\partial \theta}\right|_{\theta=\theta^\circ} \\ &+ \quad \frac{1}{2}(\theta - \theta^\circ)' \left.\frac{\partial^2 Q_n(\theta)}{\partial \theta^2}\right|_{\theta=\theta^\star} (\theta - \theta^\circ) \qquad (\star) \end{aligned}$$

where $\|\theta - \theta^\star| \leq \|\theta - \theta^\circ\|$. Hence

$$\begin{aligned} Q_n(\theta) = Q_n(\theta^\circ) \quad &+ \quad (\theta - \theta^\circ)' \left.\frac{\partial Q_n(\theta)}{\partial \theta}\right|_{\theta=\theta^\circ} + \frac{1}{2}(\theta - \theta^\circ)' V_n (\theta - \theta^\circ) \\ &+ \quad \frac{1}{2}(\theta - \theta^\circ)' T_n(\theta^\star)(\theta - \theta^\circ) \end{aligned} \tag{37}$$

where

$$T_n = \left.\frac{\partial^2 Q_n(\theta)}{\partial \theta^2}\right|_{\theta=\theta^\star} - V_n \text{ and } V_n = \left.\frac{\partial^2 Q_n(\theta)}{\partial \theta^2}\right|_{\theta=\theta^\circ}. \tag{38}$$

Theorem 1:(Klimko and Nelson (1978)) (Consistency) Suppose that

(i)

$$\lim_{n\to\infty} \lim_{\delta \downarrow 0} \sup_{||\theta^* - \theta|| \leq \delta} \frac{1}{n\delta} |T_n(\theta^\star)_{ij}| < \infty. \quad 1 \leq i,\ j \leq p, \tag{39}$$

(ii)

$$(2n)^{-1} V_n \overset{\text{a.s}}{\to} V \tag{40}$$

where V is a positive definite (symmetric) $p \times p$ matrix of constants and

(iii)

$$\frac{1}{n}\left.\frac{\partial Q_n(\theta)}{\partial \theta_i}\right|_{\theta=\theta^\circ} \overset{\text{a.s}}{\to} 0, \quad 1 \leq i \leq p. \tag{41}$$

Then there exits a sequence of estimators $\hat{\theta}_n$ such that

$$\hat{\theta}_n \overset{\text{a.s}}{\to} \theta^\circ$$

and for any $\varepsilon > 0$ there exists an event E with $P(E) > 1-\varepsilon$ and an n_0 such that on E, for $n > n_\circ$, $\hat{\theta}_n$ satisfies the equation

$$\left.\frac{\partial Q_n(\theta)}{\partial \theta}\right|_{\theta=\hat{\theta}_n} = 0 \tag{42}$$

and Q_n attains local minimum at $\hat{\theta}_n$.

Proof: Given $\varepsilon > 0$ and condition (38)-(40), applying Egorov's theorem, we can find an event E with $P(E) > 1-\varepsilon$, $0 < \delta^\star < \delta$, $M > 0$ and an n_0 such that on E, for any $n > n_0$ and $\theta \in N_{\delta^\star}$ (an open sphere with center at θ° and radius $\delta^\star$) such that

(a) $\left|(\theta-\theta^\circ)'\frac{\partial Q_n(\theta)}{\partial \theta}\right|_{\theta=\theta^\circ}| < n\delta^3$,

(b) the minimum eigenvalue of $(2n)^{-1}V_n$ is greater than some $\Delta > 0$ and

(c) $\frac{1}{2}(\theta-\theta^\circ)'T_n(\theta^\star)(\theta-\theta^\circ) < nM\delta^3$.

Using ($\star$) for θ on the boundary $N_{\delta^\star}$, we have

$$\begin{aligned} Q_n(\theta) &\geq Q_n(\theta^\circ) + n(-\delta^3 + \delta^2\Delta - M\delta^3) \\ &= Q_n(\theta^\circ) + n\delta^2(-\delta + \Delta - M\delta). \end{aligned}$$

Since $\Delta - \delta - M\delta$ can be made positive by choosing δ sufficiently small, $Q_n(\theta)$ must attain a minimum at some $\hat{\theta}_n = (\hat{\theta}_{n1}, \ldots, \hat{\theta}_{np})$ in $N_{\theta^\star}$ at which point the least squares equation (37) must be satisfied on E for any $n > n_0$.

Replace ε by $\varepsilon_k = 2^{-k}$ and δ by $\delta_k = \frac{1}{k}$, $k \geq 1$ to determine sequence of events $\{E_k\}$ and sequence of sets $\{E_k\}$ and an increasing sequence $\{n_k\}$ such that the equation (41) has a solution on E_k for any $n > n_k$. For $n_k < n \leq n_{k+1}$, define $\hat{\theta}_n$ on E_k to be a solution of (41) within δ_k of θ° and at which Q_n attains a relative minimum and define $\hat{\theta}_n$ to be zero off E_k. Then

$$\hat{\theta}_n \to \theta^\circ \quad \text{on} \quad \liminf_{k\to\infty} E_k,$$

but

$$1-P\left(\liminf_{k\to\infty} E_k\right) = P\left(\limsup_{k\to\infty} E_k^c\right)$$
$$= \lim_{k\to\infty} P\left(\cup_{j=k}^{\infty} E_j^c\right) \le \lim_{k\to\infty} \sum_{j=k}^{\infty} 2^{-j} = 0.$$

Asymptotic normality

Asymptotic normality of the estimator $\hat{\theta}_n$ can be obtained if the linear term in (36) has asymptotically multivariate normal distribution. This can be verified by the Cramer-Wold technique and appropriate Central limit theorem for martingales. Note that

$$n^{-\frac{1}{2}} \boldsymbol{\lambda}' \left.\frac{\partial Q_n(\theta)}{\partial \theta}\right|_{\theta=\theta^\circ}$$
$$= -2n^{-\frac{1}{2}} \sum_{k=1}^{n} \left[\sum_{i=1}^{p} \lambda_i \left\{ \left.\frac{\partial E_\theta(X_k|\mathcal{F}_{k-1})}{\partial \theta_i}\right|_{\theta=\theta^\circ} \right\} \right] (X_k - E_\theta(X_k|\mathcal{F}_{k-1})) \tag{43}$$

where $\boldsymbol{\lambda} = (\lambda_1, \ldots, \lambda_p) \in R^p$ is arbitrary non zero vector. Furthermore

$$\sum_{k=1}^{p} \lambda_i \left.\frac{\partial E_\theta(X_k|\mathcal{F}_{k-1})}{\partial \theta_i}\right|_{\theta=\theta^\circ}$$

is an $\mathcal{F}_{k-1}$ - measurable function. Hence, from (42), it follows that

$$\left\{ n^{-\frac{1}{2}} \boldsymbol{\lambda}' \left.\frac{\partial Q_n(\theta)}{\partial \theta}\right|_{\theta=\theta^\circ}, \mathcal{F}_n, n \ge 1 \right\}$$

is a martingale. If

$$\frac{1}{2} n^{-\frac{1}{2}} \boldsymbol{\lambda}' \left.\frac{\partial Q_n(\theta)}{\partial \theta}\right|_{\theta=\theta^\circ} \xrightarrow{\mathcal{L}} N(0, \boldsymbol{\lambda}' W \boldsymbol{\lambda})$$

for any $\boldsymbol{\lambda} \neq 0$ where W is a $p \times p$ covariance matrix, then

$$\frac{1}{2} n^{-\frac{1}{2}} \left.\frac{\partial Q_n(\theta)}{\partial \theta}\right|_{\theta=\theta^\circ} \xrightarrow{\mathcal{L}} N(0, W) \text{ as } n \to \infty. \tag{44}$$

Theorem 2: Suppose the conditions of Theorem 1 hold. In addition, suppose that

$$\lim_{n\to\infty} \lim_{\delta\downarrow 0} \sup_{||\theta^* - \theta|| \le \delta} \frac{1}{n\delta} |T_n(\theta^\star)_{ij}| = 0, \quad \le i, \; j \le p,$$

and

$$\frac{1}{2} n^{-\frac{1}{2}} \left.\frac{\partial Q_n(\theta)}{\partial \theta}\right|_{\theta=\theta^\circ} \xrightarrow{\mathcal{L}} N(0, W) \tag{45}$$

as $n \to \infty$ where V is as defined by (39). Then

$$n^{1/2}(\hat{\theta}_n - \theta^\circ) \stackrel{\mathcal{L}}{\to} N(0, V^{-1}WV^{-1})$$

as $n \to \infty$.

Proof: Let $\hat{\theta}_n$ be as given by Theorem 1. Note that $\hat{\theta}_n$ satisfies (41). Expanding

$$n^{-\frac{1}{2}} \left. \frac{\partial Q_n(\theta)}{\partial \theta} \right|_{\theta=\hat{\theta}_n}$$

in a Taylor's expansion about θ°, we have, by (37),

$$\begin{aligned} 0 &= n^{-\frac{1}{2}} \left. \frac{\partial Q_n(\theta)}{\partial \theta} \right|_{\theta=\hat{\theta}_n} \\ &= n^{-\frac{1}{2}} \left. \frac{\partial Q_n(\theta)}{\partial \theta} \right|_{\theta=\theta^\circ} + n^{-1} \left(V_n + T_n(\theta^\star)\right) n^{\frac{1}{2}} (\hat{\theta}_n - \theta^\circ). \end{aligned}$$

Since $n^{-1}(V_n + T_n(\theta^\star)) \stackrel{\text{a.s}}{\to} 2V$ as $n \to \infty$ by (38) and (39), it follows that

$$n^{\frac{1}{2}} (\hat{\theta}_n - \theta^\circ)$$

has the same asymptotic distribution as that of

$$-(2V)^{-1} n^{-\frac{1}{2}} \left. \frac{\partial Q_n}{\partial \theta} \right|_{\theta=\theta^\circ}.$$

This proves that

$$n^{\frac{1}{2}} (\hat{\theta}_n - \theta^\circ) \stackrel{\mathcal{L}}{\to} N(0, V^{-1}WV^{-1})$$

in view of (44).

Example : (BGW process with immigration): Consider a subcritical BGW process $\{Z_n, n = 0, 1, \ldots\}$ with immigration. Suppose that the process has an initial distribution for Z_0 with $EZ_0^2 < \infty$. Let m and λ be the means of the offspring distribution and immigration distribution respectively. Assume that these distribution have finite variances. The problem is to estimate $\theta = (m, \lambda)$ based on $Z_0, \ldots, Z_n$. Note that the $(n+1)$ -th generation is obtained from the independent reproduction of each of the individuals in the n -th generation plus an independent immigration input with immigration distribution.

Thus,

$$E_\theta(Z_{n+1} | \mathcal{F}_n) = mZ_n + \lambda.$$

Let

$$\begin{aligned} Q_n(\theta) &= \sum_{k=1}^{n} [Z_k - E_\theta(Z_k | \mathcal{F}_{k-1})]^2 \\ &= \sum_{k=1}^{n} [Z_k - mZ_{k-1} - \lambda]^2. \end{aligned}$$

Then

$$\frac{\partial Q_n(\theta)}{\partial m} = 2\sum_{k-1}^{n}[Z_k - mZ_{k-1} - \lambda](-Z_{k-1})$$

and

$$\frac{\partial Q_n(\theta)}{\partial \lambda} = 2\sum_{k=1}^{n}[Z_k - mZ_{k-1} - \lambda](-1).$$

Equating the above to zero, we obtain that

$$\hat{m}_n = \frac{n\sum_{i=1}^{n} Z_{i-1}Z_i - (\sum_{i=1}^{n} Z_{i-1})(\sum_{i=1}^{n} Z_i)}{n\sum_{i=1}^{n} Z_{i-1}^2 - (\sum_{i=1}^{n} Z_{i-1})^2}$$

and

$$\hat{\lambda}_n = \frac{1}{n}\left\{\sum_{i=1}^{n} Z_i - \hat{m}_n \sum_{i=1}^{n} Z_{i-1}\right\}.$$

It can be shown that the process $\{Z_n\}$ is a Markov process with a stationary distribution. If Z_0 has this stationary distribution, then the process $\{Z_n\}$ is stationary and ergodic and the erogodic theorem can be applied and we have

$$\begin{aligned} \frac{1}{n}\sum_{i=1}^{n} Z_i &\overset{\text{a.s}}{\to} E(Z_0) = \lambda(1-m)^{-1} \equiv r_1 \text{ (say)} \\ \frac{1}{n}\sum_{i=1}^{n} Z_i^2 &\overset{\text{a.s}}{\to} E(Z_0^2) = c^2(1-m^2)^{-1} + r_1^2 = r_2 \text{ (say)} \end{aligned}$$

and

$$\frac{1}{n}\sum_{i=1}^{n} Z_i Z_{i-1} \overset{\text{a.s}}{\to} mc^2(1-m^2)^{-1} + r_1^2$$

where $$c^2 = b^2 + \sigma^2\lambda(1-m)^{-1}$$

and σ^2 and b^2 are the variance of the offspring and immigration distribution. Note the r_1 and r_2 are the first and second moments of Z_0.

In fact, the above results hold even if the initial distribution is a general distribution($\star$) and we get that

$$\hat{m}_n \overset{\text{a.s}}{\to} m, \hat{\lambda}_n \overset{\text{a.s}}{\to} \lambda$$

($\star$: Follows from Billingsley (1961): Statistical Inference for Markov processes; Revesz (1968): Law of large numbers).

Suppose the offspring and the immigration distribution have finite third moments. Then $r_3 = E(Z_0^3) < \infty$. It can be shown that

$$n^{\frac{1}{2}}\begin{pmatrix} \hat{m}_n - m \\ \hat{\lambda}_n - \lambda \end{pmatrix} \overset{\mathcal{L}}{\to} N(0, V^{-1}WV^{-1})$$

where

$$V^{-1} = c^{-2}(1-m^2)\begin{pmatrix} 1 & -r_1 \\ -r_1 & r_2 \end{pmatrix}$$

and

$$W = \begin{pmatrix} \sigma_1^2 r_3 + \sigma_2^2 r_2 & \sigma_1^2 r_2 + \sigma_2^2 r_1 \\ \sigma_1^2 r_2 + \sigma_2^2 r_1 & \sigma_1^2 r_1 + \sigma_2^2 \end{pmatrix}$$

Here σ_1^2 and σ_2^2 are defined by the relation

$$\text{Var}(Z_1|Z_0) = \sigma_1^2 Z_0 + \sigma_2^2.$$

(Ref. : Hall and Heyde (1980), p. 180-181).

Method of moments

This method does not generally lead to an estimator with any "optimal property" but it is easy to implement. We illustrate the method through two examples.

Example 1 : Let $Z_0 = 1, Z_1, Z_2, \ldots$ be a super critical BGW branching process. Let $\theta = EZ_1 > 1$ and $0 < \text{Var}Z_1 = \sigma^2 < \infty$. Suppose the problem is to estimate θ and σ^2 on the basis of a single realization $\{Z_k, 0 \le k \le n+1\}$.

Since

$$Z_{k+1} = X_{k1} + \cdots + X_{kZ_k}$$

where, conditional on Z_k, the $X_{ki}, 1 \le i \le Z_k$ are i.i.d. random variables each with the distribution of Z_1, we have

$$E(Z_{k+1}|Z_k) = \theta Z_k \text{ a.s} \quad \text{i.e., } E(\frac{Z_{k+1}}{Z_k}|Z_k) = \theta$$

and

$$E((Z_{k+1} - \theta Z_k)^2|Z_k) = \sigma^2 Z_k \text{ a.s i.e., } E\left(\left.\frac{(Z_{k+1} - \theta Z_k)^2}{Z_k}\right| Z_k\right) = \sigma^2 \text{ a.s.}$$

Suppose that $P(Z_1 = 0) = 0$. Note that

$$\left\{\frac{Z_n}{\theta^n}, \mathcal{F}_n, n \ge 0\right\}$$

is a nonnegative martingale and

$$\frac{Z_n}{\theta^n} \overset{\text{a.s}}{\to} W \quad \text{(say) as } n \to \infty.$$

It is known that W is non-degenerate and positive a.s. (Harris (1963)) and hence

$$\hat{\theta}_n = \frac{Z_{n+1}}{Z_n} \to \theta \text{ a.s as } n \to \infty.$$

Let $\overset{\sim}{\theta}_n = \frac{1}{n}\sum_{j=0}^{n} Z_{j+1}Z_j^{-1}$. Then $\overset{\sim}{\theta}_n$ can be considered as a moment estimator for m. In fact

$$\overset{\sim}{\theta}_n \to \theta \text{ a.s as } n \to \infty.$$

However $\hat{\theta}_n$ is a better estimator than $\tilde{\theta}_n$ as far as the rate of convergence to θ is concerned (Heyde and Leslie (1971) *Bull. Austral. Math. Soc. 5,* 145-155). An estimator by the method of moments for σ^2 is

$$\hat{\sigma}_n^2 = \frac{1}{n}\sum_{k=0}^{n} \frac{(Z_{k+1} - \hat{\theta}_n Z_k)^2}{Z_k}.$$

It is clear that

$$E\left(\frac{(Z_{k+1} - \theta Z_k)^2}{Z_k} - \sigma^2 \middle| Z_0, Z_1, \ldots Z_k\right) = 0$$

and hence

$$\left\{\sum_{k=0}^{n} \frac{(Z_{k+1} - \theta Z_k)^2}{Z_k} - \sigma^2, \mathcal{F}_k, k \geq 0\right\}$$

forms a martingale where $\mathcal{F}_k = \sigma(Z_0, Z_1, \ldots, Z_k)$. An application of the SLLN proves that

$$\lim_{n\to\infty} \frac{1}{n}\sum_{k=1}^{n}\left\{\frac{(Z_{k+1} - \theta Z_k)^2}{Z_k} - \sigma^2\right\} = 0 \text{ a.s.}$$

One can prove that " θ " in the above equation can be replaced by " $\hat{\theta}_n$ " by applying the result

$$\hat{\theta}_n - \theta = \sigma\zeta(n)(2Z_n^{-1}\log n)^{\frac{1}{2}}$$

where $\limsup_n \zeta(n) = 1\,\text{a.s}$ and $\liminf_n \zeta(n) = -1\,\text{a.s}$ (Heyde (1974) *Advances in Appl. Prob.* **3**, 421-433.)

Example 2 : Consider a stochastic process $\{X_n\}$ governed by the model

$$X_n = \varepsilon_n + \alpha X_{n-1}\varepsilon_{n-1}, n \geq 1$$

where ε_i, $i \geq 1$ are i.i.d. with $E\varepsilon_0 = 0$, $\sigma^2 = E\varepsilon_0^2 < \infty$, $E\varepsilon_0^3 = 0$.

We assume that $\alpha^2\sigma^2 < 1$. Note that, for k large,

$$a = EX_k = \alpha\sigma^2$$

and

$$b = E(X_k X_{k-1}) = \alpha[E\varepsilon_0^3 + 2\alpha\sigma^4] = 2\alpha^2\sigma^4.$$

Let

$$\hat{a}_n = \frac{1}{n}\sum_{k=1}^{n} X_k \text{ and } \hat{b}_n = \frac{1}{n}\sum_{k=1}^{n} X_k X_{k-1}.$$

It can be shown that

$$(\star) \qquad \begin{aligned} \hat{a}_n &\to a \text{ a.s} \\ \hat{b}_n &\to b \text{ a.s} \end{aligned}$$

and one can estimate α and σ^2. In fact

$$n^{\frac{1}{2}}(\hat{a}_n - a) \xrightarrow{\mathcal{L}} N(0, \operatorname{Var}(X_0) + 2\operatorname{cov}(X_0, X_1))$$

and if $E\varepsilon_0^6 < \infty$, then

$$n^{\frac{1}{2}}(\hat{b}_n - b) \xrightarrow{\mathcal{L}} N(0, \operatorname{Var}(X_0X_1) + 2\operatorname{cov}(X_0X_1, X_1X_2) + 2\operatorname{cov}(X_0X_1, X_2X_3).$$

Remarks : If the process $\{X_k, -\infty < k < \infty\}$ is stationary, then

$$X_n = \varepsilon_n + \alpha\varepsilon_{n-1}^2 + \sum_{k=2}^{\infty} \alpha^k \varepsilon_{n-k}^2 \prod_{j=0}^{k-1} \varepsilon_{n-j}$$

and the results given in ($\star$) can be proved.

Lecture 8

Likelihood ratios in abstract space

Let $\mathcal{X}$ be the sample space and $\Theta = \{0, 1\}$ be the parameter space. Let P_0 and P_1 be probability measures defined on a measurable space $(\mathcal{X}, \mathcal{B})$. A fundamental problem is to test the null hypothesis

$$H_0 : \theta = 0 \quad \text{against } H_1 : \theta = 1.$$

In the Neyman-Pearson formulation, we choose a critical region $W \subset \mathcal{X}$ such that if the observed $x \in W$, we reject H_0 and if $x \notin W$, we do not reject H_0 (accept H_0). The performance of the test is determined by

$$\alpha = \text{ level of significance of the test } = P_0(W)$$

and the power $\gamma = P_1(W)$. Neyman-Pearson lemma gives a method to find a test which maximizes γ for a given α. It gives a test based on the likelihood ratio.

In a general abstract space, there is no measure equivalent to the Lebesgue measure on R^n and hence the concept of likelihood is not possible to formulate. We let the Radon-Nikodym derivative play the role of the likelihood ratio. The basic problem is therefore to find a method for calculating the Radon-Nikodym derivative whenever it exists.

Given P_0 and P_1, there exists a measurable set H contained in $\mathcal{X}$ with $P_0(H) = 0$ and a non -negative function f integrable with respect to P_0 such that for any measurable set $E \subset \mathcal{X}$,

$$P_1(E) = \int_E f(x) P_0(dx) + P_1(E \cap H).(\star)$$

This result is known as the Lebesgue-decomposition. The function f is the Radon-Nikodym derivative and it will be denoted by $\frac{dP_1}{dP_0}(x)$. Note that $\frac{dP_1}{dP_0}(x)$ is the Radon-Nikodym derivative of the absolutely continuous component of P_1 with respect to P_0. Recall that if the set H has $P_1(H) = 1$, then the measures P_0 and P_1 are *singular* with respect to each other. If this happens, then the critical region $W = H$ will allow perfect (probability one) discrimination between H_0 and H_1. The test will give the correct result with the first and second kind of errors zero.

If P_0 and P_1 are both absolutely continuous with respect to each other, then they are said to be equivalent. In such a case $P_1(H) = 0$ and $f \neq 0$ with P_0 -probability one.

It is possible that P_0 and P_1 are neither singular nor equivalent. However if P_0 and P_1 are Gaussian measures, then they are either equivalent or singular with respect to each other (Hajek (1958), Feldman (1958)).

Let $\mathcal{X} = R^\infty$ and $\boldsymbol{X}_n = (X_1, \ldots, X_n)$ and $g_{ni}(\boldsymbol{X}_n)$ be the joint p.d.f. with respect to the Lebesgue measure on R^n under P_i for $i = 0$ and 1. Let $\mathcal{B}$ be the σ-algebra generated by all the cylinder sets with finite dimensional base and $\mathcal{B}_n$ be the Borel σ-algebra in R^n. Let $\boldsymbol{x}_n = (x_1, \ldots, x_n)$ and

$$f_n(\boldsymbol{x}_0) = \frac{g_{n1}(\boldsymbol{x}_n)}{g_{n0}(\boldsymbol{x}_n)}, \; \boldsymbol{x}_0 = (x_1, x_2, \ldots,) \in \mathcal{X} = R^\infty.$$

Suppose $f_n(\boldsymbol{x}_0)$ is defined a.e. P_0. Let H be the set as defined above in ($\star$). Then

(i) $f_n(\boldsymbol{x}) \stackrel{\text{a.s}}{\to} f(\boldsymbol{x})$ (P_0)

(ii) $f_n(\boldsymbol{x}) \stackrel{p}{\to} f(\boldsymbol{x})$ (P_1) in H^c, and

(iii) $f_n(\boldsymbol{x}) \to +\infty$ (P_1) in H.

(Ref.: Grenander (1950)).

Proof: (i) Suppose H is such that $P_1(H) = 0$. Hence P_0 and P_1 are absolutely continuous with respect to each other. Then the sequence $\{f_n\}$ forms a martingale and by the martingale convergence theorem and

$$f_n(\boldsymbol{x}) \stackrel{\text{a.s}}{\to} f(\boldsymbol{x}) \; (P_0) \text{ as } n \to \infty$$

where $f(\boldsymbol{x}) = \frac{dP_1}{dP_0}(\boldsymbol{x})$ as in ($\star$). Suppose H is such that $0 < P_1(H) \leq 1$. For proofs of (ii) and (iii), see Grenander (1950) p. 108-110.

Neyman-Pearson Lemma : Suppose P_1 is absolutely continuous with respect to P_0. Let $f(x) = \frac{dP_1}{dP_0}(x)$. If the critical region W is of the form

$$W = \{x | f(x) \geq c\} \subset \mathcal{X},$$

then W is the "best" critical region of given size. In other words, no other critical region at the same level of significance has greater power.

Proof: Let $V \subset \mathcal{X}$ such that $P_0(V) = P_0(W)$. Note that

$$\begin{aligned} P_0(W \cap V^c) &= P_0(W) - P_0(W \cap V) \\ &= P_0(V) - P_0(W \cap V) = P_0(V \cap W^c) \end{aligned} \tag{46}$$

Hence

$$\begin{aligned} P_1(W \cap V^c) &= \int_{W \cap V^c} f(\boldsymbol{x}) P_0(d\boldsymbol{x}) \\ &\geq c \int_{W \cap V^c} P_0(d\boldsymbol{x}) \quad \text{since } W = [\boldsymbol{x} : f(\boldsymbol{x}) \geq c] \\ &= c\, P_0(W \cap V^c) \\ &= c\, P_0(V \cap W^c) \qquad \text{(by (46))}. \end{aligned} \tag{47}$$

However

$$
\begin{aligned}
P_1(V \cap W^c) &= \int_{V \cap W^c} f(\boldsymbol{x}) P_0(d\boldsymbol{x}) \\
&\le c \int_{V \cap W^c} P_0(d\boldsymbol{x}) \qquad \text{since } W^c = [\boldsymbol{x} : f(\boldsymbol{x}) < c] \\
&= c\, P_0(V \cap W^c). \qquad (48)
\end{aligned}
$$

Combining (47) and (48), we get that

$$P_1(W \cap V^c) \ge P_1(V \cap W^c). \qquad (49)$$

Adding $P_1(W \cap V)$ on both sides, we get that

$$P_1(W) \ge P_1(V)$$

which completes the proof of the result.

We have the following theorem for best Bayesian test.

Theorem: If P_1 is absolutely continuous with respect to P_0 and if the apriori probabilities of the two hypotheses H_0 and H_1 are π_0 and π_1 respectively, then the "best" test, in the sense of minimizing the probability of an error, is given by the critical region

$$W = [\, \boldsymbol{x} : f(\, \boldsymbol{x}\,) > \frac{\pi_0}{\pi_1}].$$

Proof: The probability of the test leading to the wrong result is

$$
\begin{aligned}
\alpha &= \pi_0\, P_0(\text{Reject } H_0) + \pi_1\, P_1(\text{Reject } H_1) \\
&= \pi_0 P_0(W) + \pi_1 P_1(W^c)
\end{aligned}
$$

when W is the critical region. Hence

$$
\begin{aligned}
\alpha &= \pi_0 \int_W P_0(d\boldsymbol{x}) + \pi_1 \int_{W^c} P_1(d\boldsymbol{x}) = \pi_0 \int_W P_0(d\boldsymbol{x}) + \pi_1 \int_{W^c} f(\boldsymbol{x}) P_0(d\boldsymbol{x}) \\
&= \pi_1 + \int_W (\pi_0 - \pi_1 f(\boldsymbol{x})) P_0(d\boldsymbol{x})
\end{aligned}
$$

since

$$\int_{W^c} f(\, \boldsymbol{x}\,) P_0(d\, \boldsymbol{x}\,) = 1 - \int_W f(\, \boldsymbol{x}\,) P_0(d\, \boldsymbol{x}\,).$$

To minimize α, we should choose W in such a way that the integral is as small as possible. This can be done by choosing W as the set where the integrand is negative. Note that the integrand is negative when

$$\pi_0 - \pi_1 f(\, \boldsymbol{x}\,) < 0, \text{ that is, } f(\, \boldsymbol{x}\,) > \frac{\pi_0}{\pi_1}.$$

Remarks: The best critical region need not be unique.

Lecture 9

Representation of a second order stochastic process

Let $\{X(t), t \in T\}$ be a stochastic process with $E[X(t)]^2 < \infty$ for all $t \in T$. Let $m(t) = E[X(t)]$ and $r(s,t) = \text{cov}(X(s), X(t))$. The fundamental problem is how to represent a stochastic process, possibly with a complicated dependence structure, as a linear combination of "simple" elements. Here "simple" means orthogonal (uncorrelated).

Mercers' Theorem : Consider a symmetric non-negative definite continuous function $r(s,t)$ on $[a,b] \times [a,b]$ and the integral equation

$$\lambda\phi(t) = \int_a^b r(s,t)\phi(s)ds.(\star)$$

The eigenvalues $\lambda_1, \lambda_2, \ldots$ and the associated normalized eigenfunctions $\phi_1, \phi_2, \ldots$ satisfy

$$r(s,t) = \sum_{i=1}^{\infty} \lambda_i \phi_i(s)\phi_i(t)$$

in L_2-sense as well as with absolute and uniform convergence. Note that $\{\phi_v\}$ are orthogonal.

Remarks : Note that $r(.,t) \in L_2([a,b])$. Hence, for fixed t,

$$r(.,t) = \sum_{v=1}^{\infty} \rho_v(t)\phi_v(.)$$

where

$$\begin{aligned} \rho_v(t) &= \int_v^b r(s,t)\phi_v(s)ds \\ &= \lambda_v \phi_v(t). \end{aligned}$$

Hence $r(s,t) = \sum_{i=1}^{\infty} \lambda_i \phi_i(s)\phi_i(t)$.

Karhunen-Loéve expansion :(Karhunen (1947), Loeve (1946)) Let $\{X(t), t \in T = [a,b]\}$ be a second order process continuous in the mean on $[a,b]$, that is

$$E|X(t+h) - X(t)|^2 \to 0 \text{ as } h \to 0.$$

Define λ_v and ϕ_v for $v \geq 1$ as above through the covariance function $r(s,t)$ of the process $\{X(t), t \in T\}$. Introduce the variables

$$Z_v = \int_a^b X(t)\phi_v(t)dt.$$

Note that $\{Z_v\}$ are uncorrelated and form the expansion

$$Z(t) = \sum_{v=1}^{\infty} \phi_v(t) Z_v.$$

Then the expansion holds in L_2 -mean and $P(Z(t) = X(t)) = 1, t \in T.$

Applications:

Example : (*Test for the mean value function of a Gaussian process* $\{X(t), t \in [0,1]\}$ *with known covariance function* $r(s,t)$)

We want to test the hypothesis

$$H_0; E[X(t)] = m_0(t) \text{ against the alternative } H_1 : E[X(t)] = m_1(t).$$

We take the coordinates of the process

$$Z_v = \int_0^1 X(t)\phi_v(t)dt$$

as observables where $\phi_v(t)$ are as defined by ($\star$). Note that the random variables $\{Z_v\}$ are independent normal random variables. In fact, under H_i, $Z_v \sim N(a_{iv}, \lambda_v)$ where

$$a_{iv} = \int_0^1 m_i(t)\phi_v(t)dt.$$

It is clear that $E_i(Z_v) = a_{iv}$. Let

$$Y(t) = X(t) - m_i(t), 0 \le t \le 1.$$

Note that $E[Y(t)Y(s)] = r(t,s)$ and

$$\begin{aligned}
\mathrm{Var}(Z_v) &= \int_0^1 \int_0^1 E[(Y(t)Y(s)]\phi_v(t)\phi_v(s)dt\, ds \\
&= \int_0^1 \int_0^1 r(t,s)\phi_v(t)\phi_v(s)dt\, ds \\
&= \int_0^1 \lambda_v\phi_v(t)\phi_v(t)dt \quad (\text{by } (\star)) \\
&= \lambda_v \int_0^1 \phi_v^2(t)dt = \lambda_v.
\end{aligned}$$

Suppose that $\lambda_v \neq 0$ for all v (this holds when $r(s,t)$ is positive definte). In other words, assume that the covariance function $r(s,t)$ is positive definite. Then

$$\begin{aligned}
p_n(\boldsymbol{z}) &= \frac{\pi_{v=1}^n (2\pi\lambda_v)^{-\frac{1}{2}} \exp\{-\frac{1}{2\lambda_v}(Z_v - a_{1v})^2\}}{\pi_{v=1}^n (2\pi\lambda_v)^{-\frac{1}{2}} \exp\{-\frac{1}{2\lambda_v}(Z_v - a_{0v})^2\}} \\
&= \exp\{q_n(\boldsymbol{z})\}
\end{aligned}$$

where

$$q_n(\boldsymbol{z}) = \sum_{v=1}^{n} \left\{ Z_v(\frac{a_{1v} - a_{0v}}{\lambda_v}) - (\frac{a_{1v}^2 - a_{0v}^2}{2\lambda_v}) \right\} = \sum_{v=1}^{n} \zeta_v \quad \text{(say)}.$$

Suppose that

$$\sum_{v=1}^{\infty} \frac{(a_{1v} - a_{0v})^2}{\lambda_v} < \infty.$$

Then

$$\begin{aligned} E_0(\zeta_v) &= -\frac{(a_{1v} - a_{0v})^2}{2\lambda_v} \\ E_1(\zeta_v) &= \frac{(a_{1v} - a_{0v})^2}{2\lambda_v} \end{aligned}$$

and

$$\text{Var}(\zeta_v) = \frac{(a_{1v} - a_{0v})^2}{\lambda_v}.$$

Note that $\sum_{v=1}^{\infty} E_i(\zeta_v) < \infty$ and $\sum_{v=1}^{\infty} \text{Var}(\zeta_v) < \infty$ and $\zeta_v, v \geq 1$ are independent random variables. Hence the series $\sum_{v=1}^{\infty} \zeta_v$ converges a.s. under P_0 and P_1 and the Radon-Nikodym derivative p of P_0 with respect to P_1 is the limit of p_n. The most powerful test for testing P_0 versus P_1 is given by

$$\{ \boldsymbol{z} : p(\boldsymbol{z}) \geq c \},$$

or equivalently, by

$$\{ \boldsymbol{z} : q(\boldsymbol{z}) \geq c^\star \}.$$

Let

$$f_n(t) = \sum_{v=1}^{n} (\frac{(a_{1v} - a_{0v})}{\lambda_v}) \phi_v(t).$$

Then,

$$q_n(\boldsymbol{Z}) = \int_0^1 f_n(t) \{ X(t) - \frac{m_0(t) + m_1(t)}{2} \} dt.$$

Under the additional condition $\sum_{v=1}^{\infty} (\frac{a_{1v} - a_{0v}}{\lambda_v})^2 < \infty$, it can be shown that $f_n \to f$ in L_2 - mean and the test can be written in the form $\{ \boldsymbol{z} : q(\boldsymbol{z}) \geq c^\star \}$ where

$$q(\boldsymbol{z}) = \int_0^1 f(t) \{ X(t) - \frac{m_0(t) + m_1(t)}{2} \} dt \text{ and } \int_0^1 r(s,t) f(t) dt = m_1(s) - m_0(s).$$

Example : Consider a nonhomogenous Poisson process $N(t)$ on $[0,1]$ with positive and continuous intensity $\lambda(t)$. Then

$$P(N(t) = k) = \frac{(\int_0^t \lambda(u) du)^k e^{-\int_0^t \lambda(u) du}}{k!}, k = 0, 1, 2, \ldots$$

Note that

$$P(N(t) - N(s) = k) = \frac{(\int_s^t \lambda(u)du)^k e^{-\int_s^t \lambda(u)du}}{k!}, k = 0, 1, 2, \ldots.$$

As the observables here, we take the number of events in the interval

$$I_v = [\frac{v}{n}, \frac{v+1}{n}), v = 0, 1, ., n-1.$$

Let us test $H_0 : \lambda(t) = \lambda_0(t)$ against $H_1 : \lambda(t) = \lambda_1(t)$. Then the likelihood ratio is given by

$$p_n = \pi_{v=0}^{n-1} \frac{(\int_{\frac{v}{n}}^{\frac{v+1}{n}} \lambda_1(u)du)^{f_v} \exp(-\int_{\frac{v}{n}}^{\frac{v+1}{n}} \lambda_1(u)du)}{(\int_{\frac{v}{n}}^{\frac{v+1}{n}} \lambda_0(u)du)^{f_v} \exp(-\int_{\frac{v}{n}}^{\frac{v+1}{n}} \lambda_0(u)du)}$$

where f_v is the number of events in the interval $[\frac{v}{n}, \frac{v+1}{n})$. This follows from the fact that the Poisson process has independent increments. As $n \to \infty$, the sequence $\{p_n\}$ converges a.s. $[P_0]$ and $[P_1]$ to the Radon-Nikodym derivative of P_1 with respect to P_0 namely

$$p = [\Pi_{k=1}^{N} \frac{\lambda_1(t_k)}{\lambda_0(t_k)}] \exp\{-\int_0^1 (\lambda_1(t) - \lambda_0(t))dt\}$$

where N is the number of events that occurred in the interval $[0, 1]$ and $t_0, t_1, \ldots, t_n$ are the corresponding time points of occurrence and the most powerful test for testing H_0 versus H_1 is given by the critical region

$$\left[\Pi_{k=1}^{n} \frac{\lambda_1(t_k)}{\lambda_0(t_k)} > c\right].$$

Remarks: In both the examples described above, the observables are independent random variables.

Lecture 10

The following theorem due to Kakutani (1948) gives a necessary and sufficient condition for the equivalence of two product measures.

Theorem: Consider two product measures $P_0 = P_0^{(1)} \times P_0^{(2)} \times \ldots$ and $P_1 = P_1^{(1)} \times P_1^{(2)} \times \ldots$ defined on some product space $\mathcal{X} = \mathcal{X}_1 \times \mathcal{X}_2 \times \ldots$ with the associated product σ -algebra. Suppose that the probability measure $P_0^{(n)}$ is equivalent to $P_1^{(n)}$ for all $n \geq 0$. Let

$$\rho(P_0^{(n)}, P_1^{(n)}) = \int_{\mathcal{X}_n} \sqrt{f_n(x_n)} P_0(dx_n) \quad \text{where } f_n(x_n) = \frac{dP_1^{(n)}}{dP_0^{(n)}}(x_n).$$

(Note that $0 \leq \rho(P_0^{(n)}, P_1^{(n)}) \leq 1$ and $\rho = 1$ if and only if $f_n = 1$ a.s. which holds if $P_1^{(n)} = P_0^{(n)}$). Then P_0 and P_1 are equivalent if and only if

$$\Pi_{n=1}^{\infty} \rho(P_0^{(n)}, P_1^{(n)}) > 0.$$

Remarks: Note the Hellinger distance between the probability measures $P_0^{(n)}$ and $P_1^{(n)}$ is defined by

$$\begin{aligned} &\rho(P_0^{(n)}, P_1^{(n)}) \\ &= \int_{\mathcal{X}_n} \sqrt{\frac{dP_1^{(n)}(x_n)}{dP_0^{(n)}(x_n)}} dP_0^{(n)}(x_n) \\ &= \int_{\mathcal{X}_n} \sqrt{dP_1^{(n)}(x_n) dP_0^{(n)}(x_n)}. \end{aligned}$$

Proof: Let

$$f_n(x_n) = \frac{dP_1^{(n)}(x_n)}{dP_0^{(n)}(x_n)} \text{ and } \rho_n = \rho(P_0^{(n)}, P_1^{(n)}).$$

Then

$$\Pi_{n=1}^{N} \rho_n = \int_{\mathcal{X}} \sqrt{R_N(x)} P_0(dx) \text{ where } R_N(x) = \Pi_{n=1}^{N} f_n(x_n)$$

since P_0 is a product measure on the space $\mathcal{X}$. By the martingale convergence theorem,

$$R_N(x) \stackrel{a.s.}{\to} f(x) \text{ (say) as } N \to \infty$$

with respect to the probability measure P_0. Hence, by the Fatou's Lemma,

$$0 \leq E_{P_0}(\lim_{N\to\infty} \sqrt{R_N(x)}) \leq \liminf_{N\to\infty} E_{P_0}(\sqrt{R_n(x)}) = \lim_{N\to\infty} inf \Pi_{n=1}^{N} \rho_n$$

which implies that

$$0 \leq E_{P_0}(\sqrt{f(x)}) \leq \Pi_{n=1}^{\infty} \rho_n.$$

If this infinite product is zero, then $f(x) = 0$ a.s. $[P_0]$ so that P_1 and P_0 are singular with respect to each other. Suppose the infinite product is positive. Let $M > N$ and consider

$$\{E_{P_0}[R_N - R_M]\}^2 \leq E_{P_0}[|R_N^{\frac{1}{2}} - R_M^{\frac{1}{2}}|]^2 E_{P_0}[|R_N^{\frac{1}{2}} + R_M^{\frac{1}{2}}|]^2$$

by the Cauchy-Schwartz inequality. But

$$\begin{aligned} E_{P_0}[|R_N^{\frac{1}{2}} - R_M^{\frac{1}{2}}|^2] &= E_{P_0}[(1 - \Pi_{n=N+1}^{M}\sqrt{f_n})^2\, R_N] \\ &= E_{P_0}[R_N + R_M - 2\Pi_{n=N+1}^{M}\sqrt{f_n}\, R_N] \\ &= 2(1 - \Pi_{n=N+1}^{M}\rho_n) \to 0 \end{aligned}$$

as M and $N \to \infty$. Further more

$$E_{P_0}[|R_N^{\frac{1}{2}} + R_M^{\frac{1}{2}}|^2] \leq 2E_{P_0}(R_N + R_M) = 4$$

since $(x+y)^2 \leq 2(x^2 + y^2)$. Hence $\{R_N, n \geq 1\}$ is a Cauchy sequence in $L_2(\Omega, \mathcal{F}, P)$. But the L_2 -space is complete. Hence R_N converges in L_2, which implies that

$$\int_{\mathcal{X}} f(x) P_0(dx) = 1.$$

Hence P_1 is absolutely continuous with respect to P_0.

Fatou's Lemma: Let $(\Omega, \mathcal{F}, P)$ be a probability space. Suppose $f_n \geq g$ and $f_n \to f$ a.s as $n \to \infty$. Further suppose that $E(f_n) < \infty$ and Eg is finite. Then $E(f) \leq \liminf_{n\to\infty} E(f_n)$.

Remarks: Note that the lemma holds if $f_n \geq 0$ and $f_n \to f$ a.s. as $n \to \infty$.

Example : (Grenander, p.269). Let $\{X(t), t \in T\}$, $T = [a,b]$, $-\infty < a < b < \infty$ be a Gaussian process with continuous covariance function $r(s,t)$ and continuous mean function $m_i(t)$, $i = 0, 1$ under the probability measures P_0 and P_1 respectively. Let $\{\lambda_v\}$ be eigenvalues with corresponding eigenfunctions $\{\phi_v\}$ such that

$$\lambda_v \phi_v(s) = \int_a^b r(s,t)\phi_v(t)dt.$$

Here we choose ϕ_v to be orthogonal and orthonormal. Let

$$Z_v = \int_0^1 X(t)\phi_v(t)dt.$$

Then $Z_v, v \geq 1$ are independent random variables. Let $P_i^{(v)}$ be the probability measure of Z_v under P_i. Then $Z_v \sim N(a_{iv}, \lambda_v)$ under P_i. Let

$$\begin{aligned}\rho(P_0^{(v)}, P_1^{(v)}) &= \int_{-\infty}^{\infty} \sqrt{P_0^{(v)}(dx) P_1^{(v)}(dx)} = \int_{-\infty}^{\infty} \sqrt{\frac{dP_1^{(v)}(x)}{dP_0^{(v)}(x)}} dP_0^{(v)}(x) \\ &= \frac{1}{\sqrt{2\pi\lambda_v}} \int_{-\infty}^{\infty} \exp\{-\frac{1}{4}(\frac{(x-a_{0v})^2}{\lambda_v} + \frac{(x-a_{1v})^2}{\lambda_v})\} dx \\ &= \exp\{-\frac{1}{8\lambda_v}(a_{0v} - a_{1v})^2\}.\end{aligned}$$

Hence, by Kakutani's theorem, P_0 and P_1 are equivalent if and only if

$$\prod_{v=1}^{\infty} \rho(P_0^v, P_1^v) > 0$$

or equivalently

$$\sum_{v=1}^{\infty} \frac{(a_{0v} - a_{1v})^2}{\lambda_v} < \infty.$$

Hence we have the following result.

Theorem : Let P_i be the probability measure generated by a Gaussian process $\{X(t), t \in T\}, T = [a,b], -\infty < a < b < \infty$ with a continuous covariance function $r(s,t)$ and a continuous mean function $m_i(t)$ for $i = 0, 1$. Then the Gaussian probability measures P_0 and P_1 with continuous mean functions $m_0(.)$ and $m_1(.)$ and the common continuous covariance function $r(.,.)$ are equivalent if and only if

$$\sum_{v=1}^{\infty} \frac{(a_{0v} - a_{1v})^2}{\lambda_v} < \infty.$$

Remarks : If the Gaussian measures P_0 and P_1 are equivalent, then the Radon-Nikodym derivative is given by

$$\begin{aligned}\lim_{n\to\infty} f_n(x) &= \lim_{n\to\infty} \frac{\pi_{v=1}^n \frac{1}{\sqrt{2\Pi\lambda_v}} \exp\{-\frac{1}{2\lambda_v}(z_v - a_{1v})^2\}}{\pi_{v=1}^n \frac{1}{\sqrt{2\pi\lambda_v}} \exp\{-\frac{1}{2\lambda_v}(z_v - a_{0v})^2\}} \\ &= \exp\{\sum_{v=1}^{\infty} \frac{(a_{1v} - a_{ov})}{\lambda_v}(z_v - c_v)\}\end{aligned}$$

where $c_v = \frac{a_{1v} + a_{ov}}{2}$.

Example : Test for covariance function of a gaussian process (Sagdar (1974))

Let $\{X(t), t \in [a,b]\}$ be a mean zero Gaussian process under the probability measures P_1 and P_2 with the covariance functions $r_1(s,t)$ and $r_2(s,t)$ respectively. We want to test the hypothesis

$$H_0 : r(s,t) = r_1(s,t) \quad \text{against} \quad H_1 : r(s,t) = r_2(s,t).$$

Consider the integral equation

$$\lambda \int_a^b r_2(s,t)\phi(t)dt = \int_a^b r_1(s,t)\phi(t)dt.$$

Let $\{\lambda_k\}$ and $\{\phi_k\}$ be the sequence of nonzero eigenvalues and the corresponding eigenfunctions respectively satisfying the above integral equation. Consider the integral equation

$$r_2(s,t) - r_1(s,t) = \int_a^b r_1(s,u)c(u,t)du.$$

Then

$$\begin{aligned}
\int_a^b \{r_2(s,t) - r_1(s,t)\}\phi_k(s)ds &= \int_a^b \phi_k(s)\left\{\int_a^b r_1(s,u)c(u,t)du\right\}ds \\
&= \int_a^b \left\{\int_a^b r_1(s,u)\phi_k(s)ds\right\}c(u,t)du \\
&= \lambda_k \int_a^b \left\{\int_a^b r_2(s,u)\phi_k(s)ds\right\}c(u,t)du.
\end{aligned}$$

Let

$$g_k(u) = \int_a^b r_2(s,u)\phi_k(s)ds.$$

Hence

$$(1-\lambda_k)\int_a^b r_2(s,t)\phi_k(s)ds = \lambda_k \int_a^b g_k(u)c(u,t)du$$

which implies that

$$(1-\lambda_k)g_k(t) = \lambda_k \int_a^b g_k(u)c(u,t)du.$$

Hence $g_k(t)$ is an eigenfunction of the kernel $c(u,t)$. Now

$$\begin{aligned}
\int_a^b \phi_k(s)g_k(s)ds &= \int_a^b \phi_k(s)\left\{\int_a^b r_2(s,t)\phi_k(t)dt\right\}ds \\
&= \int_a^b \int_a^b \phi_k(s)\phi_k(t)r_2(s,t)ds\,dt = a_k \neq 0
\end{aligned}$$

for some constant a_k and, for $k \neq j$,

$$\begin{aligned}
\int_a^b \phi_j(s)g_k(s)ds &= \int_a^b \phi_j(s)\left\{\int_a^b r_2(s,t)\phi_k(t)dt\right\}ds \\
&= \int_a^b \int_a^b r_2(s,t)\phi_j(s)\phi_k(t)dt\,ds \\
&= 0.
\end{aligned}$$

Let us normalize the bi-orthogonal system $\{\phi_k\}$ and $\{g_k\}$ so that $a_k = 1, k \geq 1$. Then

$$c(s,t) = \sum_{k=1}^{\infty} \frac{1-\lambda_k}{\lambda_k} \phi_k(s) g_k(t).$$

Define

$$Z_k = \int_a^b X(t)\phi_k(t)dt.$$

Then $Z_k \sim N(0, \lambda_k), k \geq 1$ are i.i.d. under P_1 and $Z_k \sim N(0,1)$ i.i.d. under P_2. Hence the likelihood ratio, given $Z_1, \ldots, Z_n$, is

$$L_n = (\lambda_1 \ldots \lambda_n)^{\frac{1}{2}} e^{\frac{1}{2}\sum_{k=1}^{n} \frac{(1-\lambda_k)}{\lambda_k} Z_k^2}.$$

Suppose that

$$\sum_{k=1}^{\infty} \frac{(1-\lambda_k)^2}{\lambda_k} < \infty.$$

Then the probability measures P_1 and P_2 are equivalent and the best test for H_0 against H_1 is of the form

$$\sum_{k=1}^{\infty} \frac{1-\lambda_k}{\lambda_k} Z_k^2 \geq u.$$

Let

$$\eta(t) = \int_a^b c(s,t)X(s)ds.$$

Suppose there exists a solution $\zeta(s)$ such that

$$\int_a^b r_2(s,t)\zeta(s)ds = \eta(t).$$

Then

$$\int_a^b X(s)\zeta(s)ds = \sum_{k=1}^{\infty} \frac{1-\lambda_k}{\lambda_k} Z_k^2$$

and the best critical region for testing the hypothesis H_0 against H_1 is given by

$$\int_a^b X(s)\zeta(s)ds \geq u.$$

The following result due to Baxter (1956) gives sufficient conditions for checking singularity of two gaussian measures.

Theorem : Let $\{X(t), t \in [0,1]\}$ be a Gaussian process with mean function $m(t)$ with bounded derivatives. Let the covariance function $r(s,t)$ of the process be continuous with uniformly bounded second partial derivatives for $s \neq t$. Let

$$f(t) = D^-(t) - D^+(t)$$

where

$$D^-(t) = \lim_{s \uparrow t} \frac{r(t,t) - r(s,t)}{t-s}$$

and

$$D^+(t) = \lim_{s \downarrow t} \frac{r(t,t) - r(s,t)}{t-s}.$$

Then

$$\lim_{n\to\infty} \sum_{k=1}^{n} \left[X(\frac{k}{2^n}) - X(\frac{k-1}{2^n}) \right]^2 = \int_0^1 f(t)dt \text{ a.s.}$$

As a consequence of the above theorem, we have the following result.

Corollary : Let $\{X(t), t \in [0,1]\}$ be a Gaussian process under the probability measures P_0 and P_1 for which the condition of the theorem given above hold. Define f_0 and f_1 as before. If

$$\int_0^1 f_0(t)dt \neq \int_0^1 f_1(t)dt,$$

then the probability measures P_0 and P_1 are singular with respect to each other.

Lecture 11

Stochastic Integrals and Stochastic Differential Equations

Let $\{W(t), t \geq 0\}$ be the standard Wiener process, that is, the process $\{W(t), t \geq 0\}$ is a Gaussian process with (i) $W(0) = 0$ (ii) $W(t) - W(s) \sim N(0, |t-s|)$ and (iii) the increments $W(t_1) - W(t_2)$ and $W(t_4) - W(t_3)$ are independent if $0 \leq t_1 < t_2 \leq t_3 < t_4 < \infty$.

Remarks : (i) A Wiener process has a version which has continuous sample paths almost surely.
(ii) A Wiener process has unbounded variation on any finite interval almost surely.
(iii) The sample paths of a Wiener process are nowhere differentiable almost surely.

Let $C[0,T]$ be the space of continuous functions on $[0,T]$ with the associated topology generated by the uniform metric. The Wiener process $\{W(t), 0 \leq t \leq T\}$ generates a probability measure on $C[0,T]$. Let us denote it by P_W^T.

Theorem : Doob(1953). Let $\{\xi(t), t \geq 0\}$ be a stochastic process defined on a probability spacew $(\Omega, \mathcal{F}, P)$ with continuous sample paths almost surely and $\{\mathcal{F}_t\}$ be a family of sub- σ -algebras of $\mathcal{F}$ such that $\mathcal{F}_t \subset \mathcal{F}_s$ if $t \leq s$. Suppose that
(i) for all $t \geq 0$, $\xi(t)$ is $\mathcal{F}_t$ -measurable
(ii) $E[\xi(t+h) - \xi(t)|\mathcal{F}_t] = 0$ a.s. for all $t \geq 0, h \geq 0$ that is $\{\xi(t), \mathcal{F}_t, t \geq 0\}$ is a martingale; and
(iii) $E[(\xi(t+h) - \xi(t))^2|\mathcal{F}_t)] = h$ a.s. for all $t \geq 0$ and $h \geq 0$. Then $\{\xi(t), t \geq 0\}$ is a standard Wiener process.

Stochastic integral

Let $(\Omega, \mathcal{F}, P)$ be a probability space. We want to define a stochastic integral

$$\int_0^T f(t)dW(t)$$

for a suitable class of random functions $\{f(t), t \geq 0\}$ with respect to the Wiener process $\{W(t), t \geq 0\}$. The integral cannot be defined in the Lebesgue-Stieljes sense since the Wiener process $\{W(t), t \geq 0\}$ is of unbounded variation a.s. on any finite interval $[0,T]$. Let $\{\mathcal{F}_t\}$ be a family of sub σ -algebras of $\mathcal{F}$ satisfying

i) $t_1 < t_2 \Rightarrow \mathcal{F}_{t_1} \subset \mathcal{F}_{t_2}$

ii) $W(t)$ is, $\mathcal{F}_t$ -measurable, and

iii) $W(t+s) - W(t)$ is independent of $\mathcal{F}_t$ for any t and for every $s \geq 0$.

Let $H[0,T]$ be the class of all random functions $\{f(t), 0 \leq t \leq T\}$ such that $f(t)$ is $\mathcal{F}_t$-measurable for $0 \leq t \leq T$ and

$$\int_0^T f^2(t)dt < \infty \quad \text{a.s.}$$

Case (i) Suppose $f \in H[0,T]$ and f is a step function, that is, there exists a partition

$$0 = t_0 < t_1 < \cdots < t_m = T$$

such that

$$\begin{aligned} f(t) &= f(t_i) \text{ for } t_i \leq t < t_{i+1} \text{ for } 0 \leq i \leq m-1 \\ &= f(t_{m-1}) \text{ for } t_{m-1} \leq t \leq t_m. \end{aligned}$$

Then define

$$\int_0^T f(t)dW(t) = \sum_{k=0}^{m-1} f(t_k)[W(t_{k+1}) - W(t_k)].$$

Case (ii) Consider the class of $f \in H[0,T]$ for which

$$\int_0^T E(f^2(t))dt < \infty.$$

It can be shown that any such f can be approximated by a sequence of step functions $f_n \in H[0,T]$ such that

$$\lim_{n\to\infty} E[\int_0^T |f(t) - f_n(t)|^2 dt] = 0$$

(Liptser and Shiryayev (1977)) *Statistics of Random Processes*)

We define

$$\int_0^T f(t)dW(t) = \lim_{n\to\infty} \int_0^T f_n(t)dW(t).$$

Here the limit is taken in the sense of quadratic mean. One can show that the limit is independent of the choice of sequence of step functions.

Case (iii) Let $f \in H[0,T]$. Then there exists a sequence $g_n \in H[0,T]$ such that

$$\int_0^T Eg_n^2(t)dt < \infty$$

and

$$\int_0^T [g_n(t) - f(t)]^2 dt \xrightarrow{p} 0 \text{ as } n \to \infty.$$

Define

$$\int_0^T f(t)dW(t) = \lim_{n\to\infty} \int_0^T g_n(t)dW(t)$$

where the limit is in the sense of probability. It can be proved that the limit will be independent of the choice of $\{g_n\}$.

Properties:

(i) Suppose $f_1, f_2 \in H[0,T]$ and α_1 and α_2 are random variables such that $\alpha_1 f_1 + \alpha_2 f_2 \in H[0,T]$. Then

$$\int_0^T [\alpha_1 f_1(t) + \alpha_2 f_2(t)]dW(t) = \alpha_1 \int_0^T f_1(t)dW(t) + \alpha_2 \int_0^T f_2(t)dW(t).$$

(ii) Let $f \in H[0,T]$ for which $\int_0^T Ef^2(t)dt < \infty$. Then

$$E[\int_0^T f(t)dW(t)] = 0, \text{ and } E[\int_0^T f(t)dW(t)]^2 = \int_0^T E[f^2(t)]dt.$$

(iii) Let $f \in H[0,T]$. Then, for any $\varepsilon > 0$ and $\delta > 0$,

$$P\{|\int_0^T f(t)dW(t)| > \varepsilon\} \leq P\{\int_0^T f^2(t)dt > \delta\} + \frac{\delta}{\varepsilon^2}.$$

(iv) Let $f \in H[0,T]$ for which $\int_0^T Ef^2(t)dt < \infty$. Then

$$E[\int_\alpha^\beta f(t)dW(t)|\mathcal{F}_\alpha] = 0$$

and

$$E[(\int_\alpha^\beta f(t)dW(t))^2|\mathcal{F}_\alpha] = \int_\alpha^\beta E(f^2(t)|\mathcal{F}_\alpha)dt$$

whenever $0 \leq \alpha < \beta \leq T$.

Here $\int_\alpha^\beta f(t)dW(t)$ is defined to be $\int_0^T \chi_{[\alpha,\beta]}(t) f(t)dW(t)$ where $\chi_{[\alpha,\beta]}(t)$ is the indicator function of the interval $[\alpha, \beta]$. For $f \in H[0,T]$, define

$$I(t) = \int_0^t f(s)dW(s).$$

Then $\{I(t), \mathcal{F}_t, t \geq 0\}$ is a martingale and has continuous sample paths a.s.

Stochastic differential

Suppose the process $\{\zeta(t), 0 \le t \le T\}$ satisfies the equation

$$\zeta(t_2) - \zeta(t_1) = \int_{t_1}^{t_2} a(t)dt + \int_{t_1}^{t_2} b(t)dW(t), 0 \le t_1 \le t_2 \le T$$

where $\int_0^T |a(t)|dt < \infty$ a.s and $\int_0^T b^2(t)dt < \infty$ a.s. Then the process $\zeta(t)$ is said to have the stochastic differential

$$d\zeta(t) = a(t)dt + b(t)dW(t), 0 \le t \le T.$$

Suppose $f \in H[0,T]$ and ζ is a random variable such that $P[0 \le \zeta \le T] = 1.$ Then we define

$$\int_0^{\zeta} f(t)dW(t) = I(\zeta)$$

where $I(t)$ is as defined above. If the random variables ζ_1 and ζ_2 are such that $P[0 \le \zeta_1 \le \zeta_2 \le T] = 1,$ then, define

$$\int_{\zeta_1}^{\zeta_2} f(t)dW(t) = \int_0^{\zeta_2} f(t)dW(t) - \int_0^{\zeta_1} f(t)dW(t).$$

We choose the random variables ζ to be stopping times most often. A random variable ζ is a stopping time with respect to the family $\{\mathcal{F}_t, t \ge 0\}$ if $[\zeta \le t] \in \mathcal{F}_t$ for every $t \ge 0.$ Examples of stopping times are

$$\zeta_1 = \inf\{t \ge 0 : W(t) \ge a\} \text{ for a fixed constant } a,$$

and

$$\zeta_2 = \inf\{t \ge 0 : \int_0^t f(u)dW(u) \ge a\} \text{ for a fixed constant } a.$$

Theorem : Suppose $f \in H[0,T]$ for every $T > 0$ and

$$\int_0^{\infty} f^2(s)ds = \infty \text{ a.s.}$$

Let

$$\tau_t = \inf\{u \ge 0 : \int_0^u f^2(s)ds \ge t\}.$$

Then

$$\zeta_t = \int_0^{\tau_t} f(s)dW(s),\ t \ge 0$$

is a Wiener process.

Central Limit Theorem :(CLT) Suppose $f \in H[0,T]$ for every $T > 0$ and

$$\frac{1}{T}\int_0^T f^2(s)ds \xrightarrow{p} \sigma^2 \text{ as } T \to \infty.$$

Then

$$\frac{1}{\sqrt{T}}\int_0^T f(s)dW(s) \xrightarrow{\mathcal{L}} N(0,\sigma^2) \text{ as } T \to \infty.$$

Stochastic Differential Equations

Theorem : (Existence of a solution) Suppose there exists a constant K such that

(i) $|a(t,x)-a(t,y)|+|\sigma(t,x)-\sigma(t,y)| \leq K|x-y|, x,y \in R,$

(iii) $|a(t,x)|^2+|\sigma(t,x)|^2 \leq K^2(1+|x|^2)$, and

(iii) $\eta(0)$ is independent of the Wiener process $\{W(t), t \geq 0\}$ with $E\eta^2(0) < \infty$.

Then there exists a solution $\{\eta(t), 0 \leq t \leq T\}$ satisfying the SDE

(i) $d\eta(t) = a(t,\eta(t))dt + \sigma(t,\eta(t))dW(t), 0 \leq t \leq T,$

(ii) $\eta(t)$ is continuous a.s. on $[0,T]$ with $\eta(t) = \eta(0)$ for $t = 0$,

(iii) $\sup\limits_{0 \leq t \leq T} E\{\eta^2(t)\} < \infty$, and

(iv) $\eta(t)$ is unique in the sense if $\{\eta_1(t)\}$ and $\{\eta_2(t)\}$ are two such process satisfying (i), (ii) and (iii), then

$$P\{\sup_{0 \leq t \leq T} |\eta_1(t) - \eta_2(t)| = 0\} = 1.$$

Remarks : The coefficient $a(.,.)$ is called the *drift coefficient* and the coefficient $\sigma(.,.)$ is called the *diffusion coefficient.* The problem in statistical inference for diffusion process is the estimation of these coefficients given the process $\{\eta(t), 0 \leq t \leq T\}$.

Absolute Continuity of measures generated by diffusion process

Let $(\Omega, \mathcal{F}, P)$ be a probability space and $\{\mathcal{F}_t, 0 \leq t \leq 1\}$ be a nondecreasing family of σ-algebras contained in $\mathcal{F}$. Suppose $\{W_t, 0 \leq t \leq 1\}$ is a standard Wiener process such that W_t is $\mathcal{F}_t$-measurable. For instance, one can choose $\mathcal{F}_t = \sigma\{W_s : 0 \leq s \leq t\}$.

Let $C[0,1]$ be the space of continuous functions on $[0,1]$ endowed with supnorm. Let $\{\xi_t, 0 \leq t \leq 1\}$ be a stochastic process defined on $(\Omega, \mathcal{F}, P)$ such that ξ_t is $\mathcal{F}_t$-measurable and ξ_t continuous a.s. on $[0,1]$. Let μ_ξ denote the probability measure generated by $\{\xi_t, 0 \leq t \leq 1\}$ on $C[0,1]$ and μ_W denote the probability measure generated by the Wiener process on $C[0,1]$.

Let $\mathcal{B}$ be the Borel σ -algebra on $C[0,1]$ and $\mathcal{B}_t = \sigma\{x : x_s, s \leq t\}$. Let τ be the σ -algebra of Borel sets on $[0,1]$ independent of the future i.e., $\mathcal{B}_t$ -measurable for every $0 \leq t \leq 1$.

Definition: A continuous process $\{\xi_t, \mathcal{F}_t, 0 \leq t \leq 1\}$ defined on $(\Omega, \mathcal{F}, P)$ is called a process of *diffusion type* if there exists a $\tau \times \mathcal{B}$ -measurable function $\alpha_t(x)$ such that

$$P\{\int_0^1 |\alpha_t(\xi)|dt < \infty\} = 1$$

and for each $0 \leq t \leq 1$

$$d\xi_t = \alpha_t(\xi)dt + dW_t, \xi_0 = 0.$$

Theorem : If a process is of diffusion type, then

$$P\{\int_0^1 \alpha_t^2(\xi)dt < \infty\} = 1$$

if and only if

$$\mu_\xi << \mu_W$$

and in such a case

$$\frac{d\mu_\xi}{d\mu_W} = \exp\{\int_0^1 \alpha_t(\xi)\ \ d\xi_t - \frac{1}{2}\int_0^1 \alpha_t^2(\xi)dt\} \text{ a.s. } [P]$$

Proof: See Liptser and Shiryayev(1977). The proof depends on the Girsanov theorem stated below.

Theorem : (Girsanov) Let $\{W_t, \mathcal{F}_t, P\}$ be a standard Wiener process on a probability space $(\Omega, \mathcal{F}, P)$. Let the process $\{Y_t, \mathcal{F}_t, t \geq 0\}$ be such that

$$P\{\int_0^1 Y_t^2 dt < \infty\} = 1.$$

Let

$$\phi = \exp\{\int_0^1 Y_t dW_t - \frac{1}{2}\int_0^1 Y_t^2 dt\}.$$

If $E_P\phi = 1$, then the process $\{\xi_t, \mathcal{F}_t, \tilde{P}\}$, where

$$\xi_t = -\int_0^t Y_s ds + W_t, \quad 0 \leq t \leq 1$$

and the probability measure $\tilde{P}$ is defined by

$$\frac{d\tilde{P}}{dP} = \phi,$$

is a Wiener process relative to the probability space $(\Omega, \mathcal{F}_t, \widetilde{P})$.

Heuristics for computation of the *Radon-Nikodym* **derivative for diffusion processes**

Consider the stochastic differential equation

$$dX_t = a(X_t)dt + \sigma(X_t)dW_t,\ 0 \leq t \leq 1 \quad \text{under } H_0(\mu_0)$$

and

$$dX_t = \sigma(X_t)dW_t,\ 0 \leq t \leq 1 \quad \text{under } H_1(\mu_1).$$

Let $0 = t_0 < t_1 < \cdots < t_n < t_{n+1} = 1$ be a subdivision of $[0,1]$, and discretize the above stochastic differential equation. Then

$$X(t_{k+1}) - X(t_k) - a(X_{t_k})(t_{k+1} - t_k) \simeq N(0, \sigma^2(X_{t_k})(t_{k+1} - t_k))$$

and these increments $X(t_{k+1}) - X(t_k)$, $k = 0, 1, ., \ n-1$ can be considered independent. Hence the $\log$-likelihood ratio f_n can be written in the form

$$\begin{aligned} f_n &= -\frac{1}{2}\sum_{k=0}^{n} \frac{\{X(t_{k+1}) - X(t_k) - a(X(t_k))(t_{k+1} - t_k)\}^2}{\sigma^2(X(t_k))(t_{k+1} - t_k)} \\ &\quad + \frac{1}{2}\sum_{k=0}^{n} \frac{(X(t_{k+1}) - X(t_k))^2}{\sigma^2(X(t_k))(t_{k+1} - t_k)} \\ &= \sum_{k=0}^{n} \frac{\{X(t_{k+1}) - X(t_k)\}a(X(t_k))}{\sigma^2(X(t_k))} \\ &\quad - \frac{1}{2}\sum_{k=0}^{n} \frac{a^2(X(t_k))(t_{k+1} - t_k)^2}{\sigma^2(X(t_k))(t_{k+1} - t_k)} \\ &\simeq \int_0^1 \frac{a(X(t))}{\sigma^2(X(t))}dX(t) - \frac{1}{2}\int_0^1 \frac{a^2(X(t))}{\sigma^2(X(t))}dt. \end{aligned}$$

and

$$\frac{d\mu_0}{d\mu_1} \simeq \exp\{\int_0^1 \frac{a(X(t))}{\sigma^2(X(t))}dX(t) - \frac{1}{2}\int_0^1 \frac{a^2(X(t))}{\sigma^2(X(t))}dt\}.$$

Lecture 12

Ito's Lemma: Let $F(t,x)$ be a continuous function on $[0,T] \times R$ with continuous derivatives $\frac{\partial F}{\partial t}(t,x)$, $\frac{\partial F}{\partial x}(t,x)$, $\frac{\partial^2 F}{\partial x^2}(t,x)$ and $\{Y(t), 0 \le t \le T\}$ be a stochastic process satisfying the stochastic differential equation (SDE)

$$dY(t) = a(t)\ dt + b(t)\ dW(t), Y(0) = \eta, 0 \le t \le T.$$

Then the random process $Z(t) = F(t, Y(t))$ satisfies the SDE

$$\begin{aligned} dZ(t) &= [\frac{\partial F}{\partial t}(t,Y(t)) + \frac{\partial F}{\partial y}(t,Y(t))a(t) + \frac{1}{2}\frac{\partial^2 F}{\partial y^2}(t,Y(t))b^2(t))]dt \\ &\quad + \frac{\partial F}{\partial y}(t,Y(t))b(t)dW(t), \quad Z(0) = F(0,\eta), 0 \le t \le T. \end{aligned}$$

Heuristics : Note that

$$\begin{aligned} Z(t+h) - Z(t) &= F(t+h, Y(t+h)) - F(t, Y(t)) \\ &\simeq (t+h-t)\frac{\partial F}{\partial t}(t,Y(t)) \\ &\quad +(Y(t+h) - Y(t)))\frac{\partial F}{\partial y}(t,Y(t)) \\ &\quad +\frac{1}{2}(t+h-t)^2\frac{\partial^2 F}{\partial t^2}(t,Y(t)) \\ &\quad +\frac{1}{2}(Y(t+h)-Y(t))^2\frac{\partial^2 F}{\partial y^2}(t,Y(t)) \\ &\quad +(t+h-t)(Y(t+h)-Y(t))\frac{\partial^2 F}{\partial t \partial y}(t,Y(t)) \\ &\simeq h\frac{\partial F}{\partial t} + (Y(t+h)-Y(t))\frac{\partial F}{\partial y} \\ &\quad +\frac{1}{2}h^2\frac{\partial^2 F}{\partial t^2} + \frac{1}{2}(Y(t+h)-Y(t))^2\frac{\partial^2 F}{\partial y^2} \\ &\quad +h(Y(t+h)-Y(t))\frac{\partial^2 F}{\partial t \partial y}. \end{aligned}$$

Note that

$$Y(t+h) - Y(t) \simeq a(t)h + b(t)[W(t+h) - W(t)].$$

Hence

$$
\begin{aligned}
Z(t+h)-Z(t) &\simeq h\frac{\partial F}{\partial t}+\{a(t)h+b(t)(W(t+h)-W(t))\}\frac{\partial F}{\partial y} \\
&+\frac{1}{2}h^2\frac{\partial^2 F}{\partial t^2}+\frac{1}{2}\{a(t)h+b(t)(W(t+h)-W(t))\}^2\frac{\partial^2 F}{\partial y^2} \\
&+h\{a(t)h+b(t)(W(t+h)-W(t))\}\frac{\partial^2 F}{\partial t\partial y} \\
&\simeq h\{\frac{\partial F}{\partial t}+a(t)\frac{\partial F}{\partial y}+b(t)\frac{W(t+h)-W(t)}{h}\frac{\partial F}{\partial y}\} \\
&+\frac{1}{2}h^2\frac{\partial^2 F}{\partial t^2}+\frac{1}{2}\left\{\begin{array}{c} a^2(t)h^2+b^2(t)(W(t+h)-W(t))^2 \\ +2a(t)hb(t)W(t+h)-W(t)) \end{array}\right\}\frac{\partial^2 F}{\partial y^2} \\
&+h\{a(t)h+b(t)(W(t+h)-W(t))\}\frac{\partial^2 F}{\partial t\partial y} \\
&\simeq h\{\frac{\partial F}{\partial t}+a(t)\frac{\partial F}{\partial y}+b(t)\frac{W(t+h)-W(t)}{h}\frac{\partial F}{\partial y}\} \\
&+\frac{1}{2}h^2\frac{\partial^2 F}{\partial t^2}+\frac{1}{2}\left\{a^2(t)h^2+b^2(t)h+O_p(|h|^{\frac{3}{2}})\right\}\frac{\partial^2 F}{\partial y^2} \\
&+h\left\{a(t)h+b(t)O_p(|h|^{\frac{1}{2}})\right\}\frac{\partial^2 F}{\partial t\partial y}
\end{aligned}
$$

since $E(W(t+h)-W(t))^2=|h|$ and $E|W(t+h)-W(t)|\simeq |h|^{\frac{1}{2}}$. Hence

$$
\begin{aligned}
\frac{Z(t+h)-Z(t)}{h} &\simeq \frac{\partial F}{\partial t}+a(t)\frac{\partial F}{\partial y}+b(t)\frac{W(t+h)-W(t)}{h}\frac{\partial F}{\partial y} \\
&+\frac{1}{2}b^2(t)\frac{\partial^2 F}{\partial y^2}+O_p(|h|^{\frac{1}{2}}).
\end{aligned}
$$

Therefore

$$
dZ(t)\simeq\left[\frac{\partial F}{\partial t}+a(t)\frac{\partial F}{\partial y}\right]dt+b(t)\frac{\partial F}{\partial y}dW(t)+\frac{1}{2}b^2(t)\frac{\partial^2 F}{\partial y^2}dt.
$$

We now consider sufficient conditions under which a solution of a SDE is an ergodic process.

Theorem : (Maruyama and Tanaka (1957)) Consider the SDE

$$
dX(t)=a(X(t))dt+b(X(t))dW(t), X(0)=X_0, t\geq 0.
$$

Define

$$
\phi(x)=2\int_0^x\frac{a(y)}{b^2(y)}dy.
$$

Suppose

$$
g=\int_{-\infty}^{\infty}\frac{e^{\phi(x)}}{b^2(x)}dx<\infty.
$$

Define $\mu(x) = \frac{1}{g}\int_{-\infty}^{x} \frac{e^{\phi(y)}}{b^2(y)}dy, -\infty < x < \infty$. Then the process is ergodic with stationary distribution having distribution function $\mu(.)$ and the strong law of large numbers holds, that is, if f is a function such that

$$\int_{-\infty}^{\infty} f(x)\mu(dx) < \infty,$$

then

$$\frac{1}{T}\int_0^T f(X(t))dt \to \int_{-\infty}^{\infty} f(x)\mu(dx) \text{ a.s as } T \to \infty.$$

(For proof, see Gikhman and Skorokhod: *Stochastic Differential Equations*).

Example : Suppose

$$dX(t) = -\theta X(t)dt + dW(t), X(0) = X_0, 0 \le t \le T$$

where $\theta \in \Theta \subset R$. Let $F(t,x) = e^{\theta t}x$. Then

$$\frac{\partial F}{\partial t} = \theta e^{\theta t}x, \frac{\partial F}{\partial x} = e^{\theta t}, \frac{\partial^2 F}{\partial x^2} = 0.$$

Hence, by the Ito's Lemma,

$$\begin{aligned} d(F(t,X(t)) &= [\theta e^{\theta t}X(t) + e^{\theta t}(-\theta X(t))]dt + e^{\theta t}dW(t) \\ &= e^{\theta t}dW(t). \end{aligned}$$

Therefore

$$d(e^{\theta t}X(t)) = e^{\theta t}dW(t)$$

which implies that

$$e^{\theta t}X(t) - X(0) = \int_0^t e^{\theta s}dW(s)$$

or equivalently

$$(\star)X(t) = \int_0^t e^{-\theta(t-s)}dW(s) + X(0)e^{-\theta t}.$$

If $\theta > 0$, then the process $\{X(t)\}$ is ergodic by the above theorem (Maruyama and Tanaka, *Mem. Fac. Kyushu Uni. 11* (1957) 117-141. Some properties of one-dimensional diffusion processes) and the ergodic theorem holds i.e., for any measurable function f integrable with respect to the stationary measure μ,

$$\lim_{T\to\infty}\frac{1}{T}\int_0^T f(X(t))dt = \int_{-\infty}^{\infty} f(x)\mu(dx) \text{ a.s.}$$

Suppose the process $\{X(t), 0 \leq t \leq T\}$ is observed. Let P_θ be the probability measure generated by the process on $C[0,T]$ and P_W be the measure generated by the Wiener process. Then

$$L_T(\theta) \equiv \frac{dP_\theta}{dP_W} = \exp\left\{\int_0^T -\theta X(t)dX(t) - \frac{1}{2}\int_0^T \theta^2 X^2(t)dt\right\}.$$

Note that the MLE $\hat{\theta}_T$ of θ is given by

$$\begin{aligned}\hat{\theta}_T &= \frac{-\int_0^T X(t)dX(t)}{\int_0^T X^2(t)dt} \\ &= -\left\{\frac{X^2(T) - X^2(0) - T}{2\int_0^T X^2(t)dt}\right\}.\end{aligned}$$

(It can be shown that $\int_0^T X(t)dX(t) = \frac{X^2(T)-X^2(0)-T}{2}$ by applying the Ito's lemma to the function $F(t,x) = x^2$). Note that

$$\begin{aligned}V_T(\theta) = \frac{\partial \log L_T(\theta)}{\partial\theta} &= -\int_0^T X(t)dX(t) - \theta\int_0^T X^2(t)dt \\ &= -\int_0^T X(t)[dX(t) + \theta X(t)\ dt] \\ &= -\int_0^T X(t)dW(t)\end{aligned}$$

is the score function and $\{V_t(\theta), \mathcal{F}_t, 0 \leq t \leq T\}$ is a zero mean martingale. Let θ be the true parameter. Then

$$\begin{aligned}\hat{\theta}_T - \theta &= \frac{-\int_0^T X(t)dX(t)}{\int_0^T X^2(t)dt} - \theta \\ &= \frac{-\int_0^T X(t)dX(t) - \theta\int_0^T X^2(t)dt}{\int_0^T X^2(t)dt} = \frac{-\int_0^T X(t)dW(t)}{\int_0^T X^2(t)dt}\end{aligned}$$

Hence

$$\hat{\theta}_T - \theta = \frac{V_T(\theta)}{I_T(\theta)} \text{ where } I_T(\theta) = \int_0^T X^2(t)dt.$$

In other words,

$$V_T(\theta) = I_T(\theta)(\hat{\theta}_T - \theta).$$

Suppose the process is ergodic ($\theta > 0$). Then

$$\frac{1}{T}\int_0^T X^2(t)dt \overset{a.s}{\to} \int_{-\infty}^{\infty} x^2\mu(dx) = \sigma^2 \text{ (say)}$$

where μ is the stationary distribution of the process which is the normal distribution with mean zero and variance $(2\theta)^{-1}$. Hence, by the CLT for stochastic integrals, it follows that

$$\frac{1}{\sqrt{T}}\int_0^T X(t)dW(t) \xrightarrow{\mathcal{L}} N(0,\sigma^2), \sigma^2 = (2\theta)^{-1}$$

which implies that

$$\sqrt{T}(\hat{\theta}_T - \theta) \xrightarrow{\mathcal{L}} N(0, 2\theta) \text{ as } T \to \infty.$$

Suppose $\theta < 0$. Note that

$$\{e^{\theta t}X(t) - X(0), \mathcal{F}_t, t \geq 0\}$$

is a zero mean martingale which is L_2 -bounded. Hence, by the martingale convergence theorem, we note that

$$e^{\theta t}X(t) - X(0) \to Z \text{ a.s as } t \to \infty$$

for some random variables $Z < \infty$ a.s. and

$$e^{2\theta t}X^2(t) \to (Z + X(0))^2 \text{ a.s as } t \to \infty.$$

Apply an integral version of the Toeplitz lemma. We have

$$(\star\star) \qquad e^{2\theta t}I_t(\theta) = e^{2\theta t}\int_0^t X^2(s)ds \to -\frac{1}{2\theta}(Z + X_0)^2 \text{ a.s as } t \to \infty.$$

Hence $I_t(\theta) \to \infty$ a.s as $t \to \infty$. By the martingale central limit theorem, it follows that

$$I_T^{\frac{1}{2}}(\theta)(\hat{\theta}_T - \theta) \xrightarrow{\mathcal{L}} N(0,1) \text{ as } T \to \infty$$

Note that

$$(\hat{\theta}_T - \theta)e^{-\theta T} = (e^{2\theta T}\int_0^T X^2(s)ds)^{-1}\{-e^{\theta T}\int_0^T X(s)dW(s)\}.$$

Check that $Z \sim N(0, -\frac{1}{2\theta})$ from ($\star$) on page 70 and

$$(-2\theta)^{-1}e^{-\theta T}(\hat{\theta}_T - \theta) \xrightarrow{\mathcal{L}} N(0,1) \text{ as } T \to \infty$$

from ($\star\star$) since $e^{2\theta t}E(I_t(\theta)) \to \frac{-1}{2\theta}E(Z + X(0))^2$ as $t \to \infty$.

If $\theta = 0$, then $(\hat{\theta}_T - 0) = \mathcal{L}\left(-\frac{\int_0^T W(t)dW(t)}{\int_0^T W^2(t)dt}\right) = \mathcal{L}\left(-\frac{W^2(T)-T}{2\int_0^T W^2(s)ds}\right)$. In this case, the random variable $\hat{\theta}_T$ does not have an asymptotically normal distribution.

Remarks on the structure of continuous parameter martingales: Let $(\Omega, \mathcal{F}, P)$ be a probability space and $\{\mathcal{F}_t, t \geq 0\}$ be a right continuous nondecreasing family of sub σ -algebras of $\mathcal{F}$ such that $\mathcal{F}_0$ is complete with respect to the probability measure P. Suppose $\{V_t, \mathcal{F}_t, t \geq 0\}$ is a square integrable martingale with mean zero and that the process $\{V_t, t \geq 0\}$ has right continuous sample paths almost surely. Then V_t is $\mathcal{F}_t$ -measurable, $E[V_t] = 0, E[V_t^2] < \infty$ and $E[V_t|\mathcal{F}_s] = V_s$ a.s for $0 \leq s \leq t$. Then it is known that there exists a right continuous increasing process $\{I_t, t \geq 0\}$ such that I_t is $\mathcal{F}_t$ -measurable and

$$E[(V_t - V_s)^2|\mathcal{F}_s) = E(I_t - I_s|\mathcal{F}_s) \text{ a.s } , 0 \leq s \leq t \qquad (*)$$

(cf. Meyer (1962)). The process $\{I_t, t \geq 0\}$ is the continuous analogue of the conditional variance $I_n = \sum_{j=1}^n E(X_j^2|\mathcal{F}_{j-1})$ for a discrete parameter square integrable martingale $\{S_n = \sum_{j=1}^n X_j, n \geq 1\}$. In analogy with the definition of I_n, one can formally define

$$I_t = \int_0^t E([dV_s]^2|\mathcal{F}_s)$$

and this can be used as a check for computing I_t. Suppose there exists a procee $\{\zeta_t, t \geq 0\}$ such that ζ_t is $\mathcal{F}_t$ -measurable for which

$$I_t = \int_0^t \zeta_u^2 du \text{ a.s.} \qquad (**)$$

Theorem 1:(SLLN) If $\{V_t, \mathcal{F}_t, t \geq 0\}$ satisfies (*) and the condition (**) holds, then

$$\frac{V_t}{I_t} \to 0 \text{ a.s. on } [I_t \to \infty].$$

Theorem 2:(Kunita and Watanabe (1967)) If $\{V_t, \mathcal{F}_t, t \geq 0\}$ satisfies (*) and has continuous sample paths almost surely, then there exists a standard Wiener process $\{W_t, t \geq 0\}$ such that

$$V_t = W_{I_t} \text{ a.s } , t \geq 0.$$

Theorem 3: Suppose the conditions stated in Theorems 1 and 2 hold and there exists a function $m_t \uparrow \infty$ as $t \to \infty$ such that

$$\frac{I_t}{m_t} \xrightarrow{p} \eta^2$$

where $P(\eta^2 > 0) > 0$. Then

$$V_t I_t^{-1/2} \xrightarrow{\mathcal{L}} N(0,1)$$

as $t \to \infty$ and the convergence holds with respect to any probability measure μ on $(\Omega, \mathcal{F})$ which is absolutely continuous with respect to the conditional probability measure $P_B(.) = P(.|B)$ where $B = [\eta^2 > 0]$.

Lecture 13

Estimation from discrete sampling

Let us again consider the process

$$dX(t) = \theta X(t)dt + dW(t), t \geq 0, X_0 = 0.$$

We have seen that the MLE of θ is given by

$$\hat{\theta}_T = \frac{\int_0^T X(t)dX(t)}{\int_0^T X^2(t)dt}.$$

when a continuous sample path $\{X(t), 0 \leq t \leq T\}$ is available. Suppose the process is observed at the time points $0 = t_0 < t_1 < \ldots < t_N = T$ say. In order to estimate the parameter θ, one can either consider the likelihood function of the Markov chain $\{X_{t_i}, 0 \leq i \leq N\}$ and then estimate by the maximum likelihood method provided the transition function can be explicitly computed or discretize the likelihood estimator from the continuous sample version or apply other methods of estimation such as conditional least squares.

Le Breton (1976)

Suppose we approximate

$$\int_0^T X(t)dX(t)$$

by

$$\sum_{i=0}^{N}(X(t_i) - X(t_{i-1}))X(t_{i-1})$$

and

$$\int_0^T X^2(t)dt$$

by

$$\sum_{i=1}^{N} X^2(t_{i+1})(t_i - t_{i-1}).$$

Then the estimate $\hat{\theta}_T$ can be approximated by

$$\overset{\sim}{\theta}_{NT} = \frac{\sum_{i=1}^{N}(X(t_i) - X(t_{i-1}))X(t_{i-1})}{\sum_{i=1}^{N} X^2(t_i)(t_i - t_{i-1})}.$$

Let

$$\delta_N = \max_{1 \leq i \leq N} |t_i - t_{i-1}|.$$

Theorem : (Le Breton (1975)) Suppose $\delta_N \to 0$ as $N \to \infty$. Then

(i) $\widetilde{\theta}_{N,T} \xrightarrow{p} \hat{\theta}_T$ as $N \to \infty$,

(ii) $\delta_N^{-\frac{1}{2}}(\widetilde{\theta}_{N,T} - \hat{\theta}_T) = O_p(1)$.

In general, suppose we consider the SDE

$$dX_t = a(\theta, X_t)dt + dW_t, \quad X_0 = x_0, t \geq 0.$$

Consider

$$\sum_{i=1}^{N}[X(t_i) - X(t_{i-1}) - a(\theta, X(t_{i-1}))(t_i - t_{i-1})]^2$$

and choose θ minimizing this expression. It can be shown that the estimator so obtained is consistent if

$$\frac{T}{N} \to 0 \text{ as } N \to \infty \quad \text{(Dorgovcev (1976))}$$

and asymptotically normal if

$$\frac{T}{\sqrt{N}} \to 0 \text{ as } N \to \infty \quad \text{(Prakasa Rao (1983))}.$$

Kasonga (1988) suggested the following approach. Let $U_k(\theta, t)$ be the solution of the ordinary differential equation

$$\frac{dx_t}{dt} = a(\theta, x_t) \text{ on } [t_{k-1}, t_k], \quad x_{t_{k-1}} = X_{t_{k-1}}.$$

Let $Q(\theta) = \sum_{k=1}^{N} |X(t_k) - U_k(\theta, t_k)^2|$ with

$$U_k(\theta, t) = X(t_{k-1}) + \int_{t_{k-1}}^{t} a(\theta, U_k(\theta, s))ds \quad \text{for } t_{k-1} \leq t \leq t_k$$

Choose θ to minimize $Q(\theta)$. Let $\theta^{\star}_{N,T}$ be such an estimator.

Theorem :(Kasonga (1988)) Suppose that for every $\theta_1 \neq \theta_2$,

$$p - \lim_{N\to\infty} \frac{1}{N} \sum_{k=1}^{N} |U_k(\theta_1, t_k) - U_k(\theta_2, t_k)|^2 > 0$$

and $\delta_N = \max_{1\leq i\leq N} |t_i - t_{i-1}| \to 0$ as $N \to \infty$. Then $\theta^{\star}_{N,T} \xrightarrow{p} \theta$ as $N \to \infty$ and $T \to \infty$ when θ is the true parameter.

Remarks : Consider the SDE

$$dX_t = \theta X_t dt + \sigma dW_t, t \geq 0, X_0 = 0.$$

It is known that

$$\lim_{N\to\infty}\sum_{i=1}^{2^N}[W(\frac{it}{2^N})-W(\frac{(i-1)t}{2^N})]^2=t \text{ a.s. (Doob (1953), p.395).}$$

Applying this result, it can be shown that

$$\lim_{N\to\infty}\sum_{i=1}^{2^N}[X(\frac{it}{2^N})-X(\frac{(i-1)t}{2^N})]^2=\sigma^2 t \text{ a.s. (Basawa and Prakasa Rao (1980), p.242).}$$

Parametric estimation for linear SDE

Consider the SDE

$$d\boldsymbol{X}_t=\theta\boldsymbol{X}_t dt+\boldsymbol{G}\,d\boldsymbol{W}_t, t\geq 0,\ \boldsymbol{X}_0=0$$

where the process $\boldsymbol{X}_t$ is an n-dimensional vector-valued process, $\theta\in\Theta,\Theta$ is a subset of the space of square matrices of order $n\times n$, $\boldsymbol{G}\in\zeta$ where ζ is a subset of the space of nonsingular matrices of order $n\times n$ and $\{\boldsymbol{W}_t, t\geq 0\}$ is an n-dimensional stochastic process with independent standard Wiener processes as its components. Let $\mu^T_{\theta,\boldsymbol{G}}$ be the probability measure induced by the process $\{\boldsymbol{X}_t, 0\leq t\leq T\}$ on the space $C([0,T],R^n)$, the space of continuous functions from $[0,T]$ to R^n. Using Girsanov's theorem, it can be shown that

$$\frac{d\mu^T_{\theta,\boldsymbol{G}}}{d\mu^T_{0,\boldsymbol{G}}}=\exp\{\int_0^T<\theta\boldsymbol{X}_t,(\boldsymbol{G}\boldsymbol{G}')^{-1}d\boldsymbol{X}_t>-\frac{1}{2}\int_0^T<\theta\boldsymbol{X}_t,(\boldsymbol{G}\boldsymbol{G}')^{-1}\theta\boldsymbol{X}_t>dt\}$$

where $<.,.>$ denotes the inner product in R^n and $\boldsymbol{M}'$ denotes the transpose of the matrix $\boldsymbol{M}$. Maximization of the Radon-Nikodym derivative given above with respect to the parameter θ leads to a system of linear equations which can be solved to obtain the MLE $\hat{\theta}^T_{\boldsymbol{G}}$. Furthermore, for any $(\theta,\boldsymbol{G})\in\Theta\times\zeta$ and for $0\leq t\leq T$,

$$\lim_{N\to\infty}\sum_{i=1}^{2^N}(\boldsymbol{X}_{it2^{-N}}-\boldsymbol{X}_{(i-1)t2^{-N}})(\boldsymbol{X}_{it2^{-N}}-\boldsymbol{X}_{(i-1)t2^{-N}})'=\boldsymbol{G}\boldsymbol{G}' \text{ a.s.}$$

(Ref: Basawa and Prakasa Rao (1980), p.212).

Remarks : If the true value θ_0 of the parameter θ is a real stable matrix, that is, the eigenvalues of the matrix θ have negative real parts, then the MLE $\hat{\theta}^T_{\boldsymbol{G}}$ is consistent and asymptotically normal. In fact,

$$T^{1/2}(\hat{\theta}^T_{\boldsymbol{G}}-\theta_0)\xrightarrow{\mathcal{L}}\boldsymbol{K}^{\theta_0} \text{ as } T\to\infty$$

where $\boldsymbol{K}^{\theta_0} = ((K_{ij}^{\theta_0}))$ is a Gaussian matrix with mean zero and covariance given by

$$E_{\theta_0}(K_{ij}^{\theta_0} K_{kl}^{\theta_0}) = (\boldsymbol{G}\ \boldsymbol{G}')_{ik}(\boldsymbol{Q}_{\theta_0}^{-1})_{jl}$$

and $\boldsymbol{Q}_{theta_0}$ is a positive definite matrix satisfying the relation

$$\theta_0 \boldsymbol{Q}_{\theta_0} + \boldsymbol{Q}_{\theta_0}\theta' = -\boldsymbol{G}\ \boldsymbol{G}'.$$

Remarks : Most of the results discussed above can be extended to stochastic differential equations of the type

$$d\boldsymbol{X}_t = \theta A(t, \boldsymbol{x})dt + \boldsymbol{G}\, d\boldsymbol{W}_t, t \geq 0,\ \boldsymbol{X}_0 = 0.$$

Sequential estimation for linear SDE

Consider a SDE of the form

$$d\xi(t) = \lambda\ A(t,\xi)dt + dW_t, t \geq 0, \xi_0 = 0,$$

where the unknown parameter is $\lambda, -\infty < \lambda < \infty$ and $A(t,\xi)$ is $\mathcal{F}_t$ -measurable for every $t \geq 0$. Further suppose that, for every $x(.) \in C[0,\infty), x(0) = 0,$ there exists $\epsilon = \epsilon(x) > 0$ such that

$$\int_0^{\epsilon(x)} A^2(t,\xi)dt < \infty$$

and for every λ and for every $t \geq 0$,

$$P_\lambda\{\int_0^t A^2(t,\xi)dt < \infty\} = 1.(*)$$

Here P_λ is the probability measure generated by the process $\{\xi_t, t \geq 0\}$ when λ is the true parameter. The measures P_λ and P_0 are equivalent under the condition (*). Note that the probability measure P_0 is the Wiener measure. Let P_λ^t denote the probability measure generated by the process $\{\xi(u), 0 \leq u \leq t\}$ over the space $C[0,t]$. Observe that

$$\frac{dP_\lambda^t}{dP_0^t} = \exp\{\lambda \int_0^t A(s,\xi)d\xi_s - \frac{1}{2}\lambda^2 \int_0^t A^2(s,\xi)ds\}.$$

It is now easy to check that the MLE of the parameter λ, given the observation $\{\xi(s), 0 \leq s \leq T\}$, is

$$\hat{\lambda}_T(\xi) = \frac{\int_0^T A(s,\xi)d\xi_s}{\int_0^T A^2(s,\xi)ds}.$$

Observe that

$$
\begin{aligned}
E_\lambda[\hat{\lambda}_T(\xi)] &= E_\lambda[\frac{\int_0^T A(s,\xi)d\xi_s}{\int_0^T A^2(s,\xi)ds}] \\
&= E_\lambda[\frac{\lambda \int_0^T A^2(s,\xi)ds + \int_0^T A(s,\xi)dW_s}{\int_0^T A^2(s,\xi)ds}] \\
&= \lambda + E_\lambda[\frac{\int_0^T A(s,\xi)dW_s}{\int_0^T A^2(s,\xi)ds}].
\end{aligned}
$$

Suppose that

$$
P_\lambda\{\int_0^\infty A^2(s,\xi)ds = \infty\} = 1, -\infty < \lambda < \infty.
$$

For any $H \geq 0$, define

$$
\tau(H) = \inf\{t \geq 0 : \int_0^t A^2(s,\xi)ds = H\}.
$$

Define

$$
\begin{aligned}
\hat{\lambda}(H) &= \hat{\lambda}_{\tau(H)} = \frac{\int_0^{\tau(H)} A(s,\xi)d\xi_s}{\int_0^{\tau(H)} A^2(s,\xi)ds} \\
&= \frac{1}{H}\int_0^{\tau(H)} A(s,\xi)d\xi_s.
\end{aligned}
$$

The estimator $\hat{\lambda}(H)$ is called a *sequential maximum likelihood estimator* of the parameter λ. Note that

$$
\begin{aligned}
\hat{\lambda}(H) &= \frac{1}{H}\int_0^{\tau(H)} A(s,\xi)d\xi_s \\
&= \frac{1}{H}\{\lambda \int_0^{\tau(H)} A^2(s,\xi)ds + \int_0^{\tau(H)} A(s,\xi)dW_s\} \\
&= \lambda + \frac{1}{H}\int_0^{\tau(H)} A(s,\xi)dW_s.
\end{aligned}
$$

Hence the distribution of the estimator $\hat{\lambda}(H)$ is $N(\lambda, \frac{1}{H})$ from the properties of stochastic integrals with respect to a standard Wiener process.

Cramer-Rao inequality

Let us consider a sequential plan $(\tau(\xi), \hat{\lambda}_\tau(\xi))$ for estimating a function $h(\lambda)$ such that

$$
E_\lambda[\hat{\lambda}_\tau(\xi)] = h(\lambda).
$$

Note that $\tau(\xi)$ is the stopping time of the sequential plan $(\tau(\xi), \hat{\lambda}_\tau(\xi))$. Suppose that $h(\lambda)$ is differentiable and that differentiation with respect to λ under the expectation operator is permissible in the above equation. Further suppose that

$$E_\lambda\{\int_0^{\tau(\xi)} A^2(t,\xi)dt\} < \infty.$$

Theorem:(Cramer-Rao inequality) Under the conditions stated above

$$Var_\lambda(\hat{\lambda}_\tau(\xi)) \geq \frac{[h'(\lambda)]^2}{E_\lambda\{\int_0^{\tau(\xi)} A^2(t,\xi)dt\}}.$$

Proof: Let P_λ be the probability measure generated by the process $\{\xi(s), 0 \leq s \leq t\}$ corresponding to the parameter λ and $P_\lambda^{\tau(\xi)}$ be the probabilty measure generated by the process $\{\xi(t), 0 \leq t \leq \tau(\xi)\}$. Applying Sudakov's lemma (cf. Basawa and Prakasa Rao (1980)), it can be shown that

$$\frac{dP_\lambda^{\tau(\xi)}}{dP_{\lambda_0}^{\tau(\xi)}}$$

exists and

$$\frac{dP_\lambda^{\tau(\xi)}}{dP_{\lambda_0}^{\tau(\xi)}} = \exp\{(\lambda - \lambda_0)\int_0^{\tau(\xi)} A(t,\xi)d\xi(t) - \frac{1}{2}(\lambda^2 - \lambda_0^2)\int_0^{\tau(\xi)} A^2(t,\xi)dt\}.$$

Note that $E_\lambda[\hat{\lambda}_\tau(\xi)] = h(\lambda)$ and hence

$$\int \hat{\lambda}_\tau(\xi)dP_\lambda = h(\lambda)$$

which can also be written in the form

$$\int \hat{\lambda}_\tau(\xi)\frac{dP_\lambda}{dP_{\lambda_0}}dP_{\lambda_0} = h(\lambda).$$

Differentiating under the integral sign with respect to λ , we get that

$$\int \hat{\lambda}_\tau(\xi)\frac{d}{d\lambda}(\frac{dP_\lambda}{dP_{\lambda_0}})dP_{\lambda_0} = h'(\lambda).$$

Hence

$$\int \hat{\lambda}_\tau(\xi)(\frac{dP_\lambda}{dP_{\lambda_0}})(\int_0^{\tau(\xi)} A(t,\xi)d\xi(t) - \lambda\int_0^{\tau(\xi)} A^2(t,\xi)dt)dP_{\lambda_0} = h'(\lambda).$$

Therefore

$$E_\lambda[\hat{\lambda}_\tau(\xi)(\int_0^{\tau(\xi)} A(t,\xi)d\xi(t) - \lambda\int_0^{\tau(\xi)} A^2(t,\xi)dt)] = h'(\lambda).$$

Observe that

$$\int_0^T A(t,\xi)d\xi(t) = \lambda \int_0^T A^2(t,\xi)dt + \int_0^T A(t,\xi)dW_t$$

and hence

$$E_\lambda[\int_0^{\tau(\xi)} A(t,\xi)d\xi(t) - \lambda \int_0^{\tau(\xi)} A^2(t,\xi)dt] = E_\lambda[\int_0^{\tau(\xi)} A(t,\xi)dW_t] = 0.$$

The above relations imply that

$$E_\lambda[(\hat{\lambda}_\tau(\xi) - h(\lambda))(\int_0^{\tau(\xi)} A(t,\xi)dW_t)] = h'(\lambda).$$

Applying the Cauchy-Schwarz inequality, we have

$$\begin{aligned} [h'(\lambda]^2 &\leq Var(\hat{\lambda}_\tau(\xi))E_\lambda[\int_0^{\tau(\xi)} A(t,\xi)dW_t]^2 \\ &= Var(\hat{\lambda}_\tau(\xi))E_\lambda[\int_0^{\tau(\xi)} A^2(t,\xi)dt]. \end{aligned}$$

Hence

$$Var(\hat{\lambda}_\tau(\xi)) \geq \frac{[h'(\lambda]^2}{E_\lambda[\int_0^{\tau(\xi)} A^2(t,\xi)dt]}.$$

In particular, if $h(\lambda) \equiv \lambda$, then

$$Var(\hat{\lambda}_\tau(\xi)) \geq \frac{1}{E_\lambda[\int_0^{\tau(\xi)} A^2(t,\xi)dt]}.$$

Definition: A sequential plan $(\tau(\xi), \hat{\lambda}_\tau(\xi))$ is said to be *efficient* if the variance of the corresponding estimator $\hat{\lambda}_\tau(\xi)$ attains the Cramer-Rao lower bound.

Observe that, for the sequential plan defined by the stopping time $\tau(H)$,

$$Var(\hat{\lambda}_{\tau(H)}) = \frac{1}{H}$$

which is the Cramer-Rao lowerbound for the variance of unbiased estimators of λ. Hence the estimator $\hat{\lambda}_{\tau(H)}$ is an efficient estimator for estimating the parameter λ.

MLE for the drift parameter for the diffusion process

Suppose $(\Omega, \mathcal{F}, P)$ is a probability space and $\{X_t, t \geq 0\}$ be a stochastic process defined on it satisfying the SDE

$$dX_t = a(t, X_t, \theta)dt + dW_t, X_0 = 0, t \geq 0, \theta \in \Theta \subset R.$$

The problem is to estimate the parameter θ based on the observation $\{X_s, 0 \leq s \leq T\}$. We assume that $(A_0) P_{\theta_1} \neq P_{\theta_2}$ (Identifiability condition) whenever $\theta_1 \neq \theta_2 \in \Theta$, and $(A_1) P_\theta(\int_0^T a^2(t, X_t, \theta)dt < \infty) = 1, \theta \in \Theta, T \geq 0$.

Let P_θ^T denote the probability measure generated by the process $\{X_s, 0 \leq s \leq T\}$ and P_W^T denote the probability measure generated by the standard Wiener process $\{W_s, 0 \leq s \leq T\}$. Then

$$\frac{dP_\theta^T}{dP_W^T} = \exp\{\int_0^T a(t, X_t, \theta)dX_t - \frac{1}{2}\int_0^T a^2(t, X_t, \theta)dt\}.$$

A maximum likelihood estimator (MLE) $\hat{\theta}_T(X^T)$ maximizes the likelihood function $L_T(\theta) = \frac{dP_\theta^T}{dP_W^T}$. If Θ is compact and $L_T(\theta)$ is continous in θ, then there exists measurable MLE (cf. Schemetterer (1974), Prakasa Rao (1987)). We assume the existence of a measurable MLE in the following discussion. Let

$$F(t, x, \theta) = \int_0^x a(t, y, \theta)dy. \tag{1}$$

(A_2) (i)Suppose the function $a(t, x, \theta)$ is continuous in x and the function $F(t, x, \theta)$ is jointly continuous in (t, x) with partial derivatives $F_x, F_t,$ and F_{xx}.

Observe that $F_x = a$ and $F_{xx} = a_x$. Applying the Ito's lemma, we have

$$dF(t, X_t, \theta) = [F_t(t, X_t, \theta) + \frac{1}{2}a_x(t, X_t, \theta)]dt + a(t, X_t, \theta)dX_t.$$

Hence

$$\int_0^T a(t, X_t, \theta)dX_t = F(T, X_T, \theta) - \int_0^T f(t, X_t, \theta)dt$$

where

$$f(t, x, \theta) = F_t(t, x, \theta) + \frac{1}{2}a_x(t, x, \theta).$$

Therefore

$$\ell_T(\theta) = \log L_T(\theta) = F(t, X_t, \theta) - \int_0^T [f(t, X_t, \theta) + \frac{1}{2}a^2(t, X_t, \theta)]dt. \tag{2}$$

(A_2) (ii) Suppose the function $\ell_T(\theta) = \log L_T(\theta)$ is twice differentiable in θ.

Observe that

$$\begin{aligned} \ell_T'(\theta) &= \int_0^T a'(t, X_t, \theta)(dX_t - a(t, X_t, \theta)dt) \\ &= \int_0^T a'(t, X_t, \theta)dW_t^\theta \end{aligned}$$

where

$$W_t^\theta = X_t - \int_0^t a(s, X_s, \theta)ds$$

is a Wiener process under the parameter $\theta.$. Similarly

$$\begin{aligned}
\ell_T''(\theta) &= \int_0^T a''dX_t - \int_0^T (aa'' + (a')^2)dt \\
&= \int_0^T a''(dX_t - adt) - \int_0^T (a')^2 dt \\
&= \int_0^T a'' dW_t^\theta - \int_0^T (a')^2 dt.
\end{aligned}$$

(A_2) (iii)Suppose that the function $\ell_T''(\theta)$ is continuous in a neighbourhood V_θ of θ for every $\theta \in \Theta$ and

$$E_\theta[\int_0^T (a'(t, X_t, \theta))^2 dt] < \infty, E_\theta[\int_0^T (a''(t, X_t, \theta))^2 dt] < \infty.$$

Further suppose that

(A_3) for every θ, there exists a neighbourhood V_θ of θ in Θ such that

$$P_\theta(\int_0^\infty (a(t, X_t, \theta') - a(t, X_t, \theta)^2 dt = \infty) = 1$$

for every $\theta' \in V_\theta - \{\theta\}$.

Let

$$I_T(\theta) = \int_0^T (a'(t, X_t, \theta))^2 dt$$

and

$$Y_T(\theta) = \int_0^T (a''(t, X_t, \theta))^2 dt.$$

(A_4) Suppose that there exista a function $m_t \uparrow \infty$ such that

$$\frac{I_T(\theta)}{m_T} \xrightarrow{p} \eta^2(\theta)$$

and

$$\frac{Y_T(\theta)}{m_T} \xrightarrow{p} \zeta^2(\theta)$$

under P_θ -measure as $T \to \infty$ where $P_\theta(\eta^2(\theta) > 0) > 0$.

Theorem : Under the conditions $(A_0) - (A_4)$ stated above, there exists a solution of the likelihood equation $\ell_T'(\theta) = 0$ which is strongly consistent as $T \to \infty$. Furthermore

$$(I_T(\theta))^{1/2}(\hat{\theta}_T - \theta) \xrightarrow{\mathcal{L}} N(0, 1)$$

as $T \to \infty$ conditionally with respect to any probability measure $\mu << P_\theta^A$ where $P_\theta^A(.) = P_\theta(.|A)$ and $A = [\eta^2(\theta) > 0]$.

Proof : For detailed proof, see Prakasa Rao (1999),p.16. We sketch it. Let $\delta > 0$ such that θ and $\theta+\delta$ belong to Θ. Then

$$\begin{aligned}
\ell_T(\theta+\delta) - \ell_T(\theta) &= [\int_0^T a(t,X_t,\theta+\delta)dX_t - \frac{1}{2}\int_0^T a^2(t,X_t,\theta+\delta)dt] \\
&\quad -[\int_0^T a(t,X_t,\theta)dX_t - \frac{1}{2}\int_0^T a^2(t,X_t,\theta)dt] \\
&= \int_0^T [a(t,X_t,\theta+\delta) - a(t,X_t,\theta)]dX_t \\
&\quad -\frac{1}{2}\int_0^T [a^2(t,X_t,\theta+\delta) - a^2(t,X_t,\theta)]dt \\
&= \int_0^T A_t^{\theta+\delta}dX_t - \frac{1}{2}\int_0^T [a^2(t,X_t,\theta+\delta) - a^2(t,X_t,\theta)]dt
\end{aligned}$$

where

$$A_t^{\theta+\delta} = a(t,X_t,\theta+\delta) - a(t,X_t,\theta).$$

It is easy to check that

$$\ell_T(\theta+\delta) - \ell_T(\theta) = \int_0^T A_t^{\theta+\delta}dW_t^{\theta} - \frac{1}{2}\int_0^T (A_t^{\theta+\delta})^2 dt.$$

Let

$$K_T = \int_0^T (A_t^{\theta+\delta})^2 dt.$$

Then

$$\frac{\ell_T(\theta+\delta) - \ell_T(\theta)}{K_T} = \frac{\int_0^T A_t^{\theta+\delta}dW_T^{\theta}}{\int_0^T (A_t^{\theta+\delta})^2 dt} - \frac{1}{2}.$$

Applying Lepingle's strong law of large numbers (cf. Prakasa Rao(1999)), it follows that

$$\frac{\int_0^T A_t^{\theta+\delta}dW_T^{\theta}}{\int_0^T (A_t^{\theta+\delta})^2 dt} \stackrel{a.s.}{\to} 0$$

as $T \to \infty$ since

$$\int_0^T (A_t^{\theta+\delta})^2 dt \stackrel{a.s.}{\to} \infty$$

as $T \to \infty$ by the condition (A_3). Hence, for every θ and δ and for almost every $\omega \in \Omega$, , there exists T_0 depending on θ, δ and ω such that for every $T \geq T_0$,

$$\ell_T(\theta+\delta) < \ell_T(\theta). \tag{3}$$

Similarly we can show that

$$\ell_T(\theta-\delta) < \ell_T(\theta). \tag{4}$$

for sufficiently large T. Since the function $\ell_T(\theta)$ is continuous on the closed interval $[\theta-\delta,\theta+\delta]$, it has a local maximum and the maximum is attained at some point $\hat{\theta}_T$ in the open interval $(\theta-\delta,\theta+\delta)$ in view of inequalities (3) and (4). Furthermore $\ell_T'(\hat{\theta}_T)=0$. This proves that

$$\hat{\theta}_T \stackrel{a.s.}{\rightarrow} \theta$$

as $T\to\infty$ under P_θ -measure. This proves the existence and strong consistency of a maximum likelihood estimator.

Applying Taylor's expansion to the function $\ell_T'(\theta)$ at $\hat{\theta}_T$, , we get that

$$\ell_T'(\theta)=\ell_T'(\hat{\theta}_T)+(\theta-\hat{\theta}_T)\ell_T''(\theta_T^*)$$

where $|\theta_T^*-\theta|\le|\hat{\theta}_T-\theta|$. Hence

$$\begin{aligned}\frac{\ell_T'(\theta)}{\sqrt{I_T(\theta)}} &= \frac{(\theta-\hat{\theta}_T)\ell_T''(\theta_T^*)}{\sqrt{I_T(\theta)}}\\ &\simeq \frac{(\theta-\hat{\theta}_T)\ell_T''(\theta)}{\sqrt{I_T(\theta)}}\end{aligned}$$

as $T\to\infty$ since $\hat{\theta}_T^* \stackrel{a.s.}{\rightarrow} \theta$ and $I_T(\theta)\stackrel{a.s}{\rightarrow}\infty$ as $T\to\infty$ and $\ell_T''(\theta)$ is continuous. Let $\mathcal{F}_T$ be the sub- σ -algebra generated by the process $\{X_s, 0\le s\le T\}$. Note that the process $\{\ell_T'(\theta),\mathcal{F}_T,T\ge 0\}$ is a martingale and, by the earlier remarks,

$$\frac{\ell_T'(\theta)}{\sqrt{I_T(\theta)}}\stackrel{\mathcal{L}}{\rightarrow}N(0,1)$$

as $T\to\infty$ under P_θ^A -measure. Hence

$$\frac{(\theta-\hat{\theta}_T)\ell_T''(\theta)}{\sqrt{I_T(\theta)}}\stackrel{\mathcal{L}}{\rightarrow}N(0,1).$$

Observe that

$$\frac{\ell_T''(\theta)}{I_T(\theta)}=\frac{\int_0^T a''(t,X_t,\theta)dW_t^\theta-\frac{1}{2}\int_0^T(a'(t,X_t,\theta))^2dt}{I_T(\theta)}\stackrel{a.s}{\rightarrow}-1$$

as $T\to\infty$ under P_θ^A -measure (cf. Feigin (1976)). In particular,

$$\sqrt{I_T(\theta)}(\hat{\theta}_T-\theta)\stackrel{\mathcal{L}}{\rightarrow}N(0,1)$$

as $T\to\infty$ under P_θ^A -measure. This result proves the asymptotic normality of the MLE under random norming.

Example : Consider the SDE

$$dX_t = \theta t X_t dt + dW_t, X_0 = 0, t \geq 0.$$

Check that

$$X_t = e^{\theta t^2/2} \int_0^t e^{-\theta s^2/2} dW_s, t \geq 0$$

and the MLE is strongly consistent and asymptotically normal after random normalization.

Remarks: For the vector parameter case, see Prakasa Rao(1999), p. 20.

In order to find "efficient" estimators as in the classical problems of estimation in the finite dimensional case, we now obtain analogue of Cramer-Rao lower bound and discuss the concept of locally asymptotically normal (LAN) families of distributions.

Cramer-Rao lower bound

Consider the SDE

$$dX(t) = a(\theta, t, X)dt + dW_t, X(0) = X_0, t \geq 0, \theta \in \Theta \subset R.$$

Suppose that

$$P_\theta(\int_0^T a^2(\theta, t, X)dt < \infty) = 1$$

and $a(\theta, t, x)$ is differentiable with respect to θ. Let

$$I(\theta_1, \theta_2) = E_{\theta_1}[\int_0^T a_\theta^2(\theta_2, t, X)dt].$$

Here a_θ denotes denote the partial derivative of the function $a(\theta, t, x)$ with respect to θ.

Theorem: Suppose that $I(\theta_1, \theta_2) > 0$ for all $\theta_1, \theta_2 \in \Theta$ and $I(\theta, \theta)$ is continuous in $\theta.$. Let θ_T^* be any estimator of the parameter θ, based on the observation $\boldsymbol{X}_T = \{X_s, 0 \leq s \leq T\}$, such that $E_\theta(\theta_T^* - \theta)^2$ is bounded over compact subsets of $\Theta.$. Let $b(\theta) = E_\theta(\theta_T^* - \theta)$. Then $b(\theta)$ is differentiable almost everywhere and

$$E_\theta(\theta_T^* - \theta)^2 \geq \frac{(1 + b'(\theta))^2}{I(\theta, \theta)} + b^2(\theta)$$

where $b'(\theta)$ denotes the derivative of $b(\theta)$ with respect to θ whenever it exists.

For proof, see Prakasa Rao (1999), p. 28.

Local Asymptotic Normality (LAN):

Let $(\Omega, \mathcal{F}, P)$ be a probability space and for $\epsilon \in (0,1]$, let $\mathcal{F}^{(\epsilon)} = \{\mathcal{F}_t^{(\epsilon)}, 0 \leq t \leq 1\}$ be a fitration, that is, a family of nondecreasing family of sub σ -algebras contained in $\mathcal{F}$. Let $\boldsymbol{X}_\epsilon = \{X_\epsilon(t), 0 \leq t \leq T_\epsilon\}$ be a diffusion process satisfying the SDE

$$dX_\epsilon(t) = a_\epsilon(\theta, t, \boldsymbol{X}_\epsilon)dt + dW_\epsilon(t), X_\epsilon(0) = \eta_\epsilon, 0 \leq t \leq T_\epsilon$$

where η_ϵ is an $\mathcal{F}_0^{(\epsilon)}$ -measurable random variable and $\theta \in \Theta$ open in R. Let $P_\theta^{(\epsilon)}$ be the probability measure generated by the process $\boldsymbol{X}_\epsilon$. Suppose that

$$P_\theta^{(\epsilon)}(\int_0^{T_\epsilon} a_\epsilon^2(\theta, t, X_\epsilon)dt < \infty) = 1, \theta \in \Theta, 0 < \epsilon \leq 1.$$

Let $\theta_0 \in \Theta$. Suppose further $\phi_\epsilon(\theta_0) \to 0$ as $\epsilon \to 0$. It can be shown that the measures $P_{\theta_0+\phi_\epsilon(\theta_0)u}^{(\epsilon)}$ and $P_{\theta_0}^{(\epsilon)}$ are absolutely continous with respect to each other in a neighbourhood of θ_0. Let

$$Z_\epsilon(u) = \frac{dP_{\theta_0+\phi_\epsilon(\theta_0)u}^{(\epsilon)}}{dP_{\theta_0}^{(\epsilon)}}(\boldsymbol{X}_\epsilon).$$

Definition : A family of probability measures $\{P_\theta^{(\epsilon)}, \theta \in \Theta\}$ is said to be *locally asymptotically normal* (LAN) at $\theta_0 \in \Theta$ if

$$\log Z_\epsilon(u) = u\Delta_\epsilon(\theta_0, \boldsymbol{X}_\epsilon) - \frac{1}{2}u^2 + \psi_\epsilon(\theta_0, u, \boldsymbol{X}_\epsilon)$$

where

$$\Delta_\epsilon(\theta_0, \boldsymbol{X}_\epsilon) \xrightarrow{\mathcal{L}} N(0,1)$$

and

$$\psi_\epsilon(\theta_0, u, \boldsymbol{X}_\epsilon) \xrightarrow{p} 0$$

as $\epsilon \to 0$ under $P_{\theta_0}^{(\epsilon)}$ -measure.

Remarks : The function $\phi_\epsilon(\theta_0)$ is called the normalization. Local asymptotic normality of the family of probability measures $\{P_\theta^{(\epsilon)}, \theta \in \Theta\}$ implies that the likelihood ratio process

$$\frac{dP_\theta^{(\epsilon)}}{dP_{\theta_0}^{(\epsilon)}}(\boldsymbol{X}_\epsilon)$$

has the properties of the process

$$Z(u) = \exp\{u\zeta - \frac{1}{2}u^2\}, -\infty < u < \infty$$

where ζ is N(0,1) whenever θ is close to θ_0 and for ϵ small. Typically, the normalization $\phi_\epsilon(\theta) = (I_\epsilon(\theta))^{-1/2}$ where $I_\epsilon(\theta)$ is the Fisher information. Under some conditions, it can be shown that the family of probability measures $\{P_\theta^{(\epsilon)}, \theta \in \Theta\}$ is LAN (cf. Theorem 2.2.17, p.32, Prakasa Rao (1999)).

Hajek-Lecam inequality

Suppose the family of probability measures $\{P_\theta^{(\epsilon)}, \theta \in \Theta\}$ is LAN with normalizing function $\phi_\epsilon(\theta)$. Let $\ell(.)$ be a symmetric function, continuous at zero, such that the set $\{x : \ell(x) < c\}$ is convex for all $c > 0$. Further suppose that for any $h > 0$,

$$\ell(x) < e^{hx^2}$$

for $|x|$ large. Then, for every $\gamma \in (0,1)$,

$$\liminf_{\epsilon \to 0} \inf_{\theta_\epsilon^*} \sup_{|\theta - y| < \phi_\epsilon{}^\gamma(\theta)} E_y[\ell(\frac{\theta_\epsilon^* - y}{\phi_\epsilon(\theta)})] \geq E[\ell(\xi)] \qquad (*)$$

where the random variable ξ has the standard normal distribution.

For a proof of this result, see Kutoyants (1984). If $\ell(x) = x^2$, then the inequality (*) reduces to

$$\liminf_{\epsilon \to 0} \inf_{\theta_\epsilon^*} \sup_{|\theta - y| < \phi_\epsilon{}^\gamma(\theta)} E_y[\frac{\theta_\epsilon^* - y}{\phi_\epsilon(\theta)}]^2 \geq 1.$$

Definition: An estimator θ_ϵ^* is said to be *asymptotically efficient* if

$$\lim_{\epsilon \to 0} \sup_{|\theta - y| < \phi_\epsilon{}^\gamma(\theta)} E_y[\frac{\theta_\epsilon^* - y}{\phi_\epsilon(\theta)}]^2 = 1.$$

Example : Consider the SDE

$$dX(t) = -\theta X(t)dt + dW_t, X(0) = 0, \theta \in (\alpha, \beta), \alpha > 0.$$

Then the family of probability measures $\{P_\theta^T, \theta \in \Theta\}$ is LAN with the normalizing function $\phi_T(\theta) = \sqrt{2\theta} T^{-1/2}$ as $T \to \infty$. If $\theta \in (\alpha, \beta), \beta < 0$, then the family of probability measures $\{P_\theta^T, \theta \in \Theta\}$ is LAN as $T \to \infty$ with the normalizing function $\phi_T(\theta) = 2\theta e^{\theta T}$.

Parametric estimation for diffusion type processes from sampled data

Consider the SDE

$$dX_t = a(X_t, \theta)dt + \sigma(X_t)dW_t, t \geq 0.$$

We now describe some methods of estimation of the parameter when the process X_t is sampled at discret time points at equal time intervals. For detailed exposition, see Prakasa Rao (1999).

Estimation based on discretization by the Euler method

Suppose the drift and diffusion are constant over the interval $[t, t+\Delta t)$. Then

$$X_{t+\Delta t} - X_t = a(X_t, \theta)\Delta t + \sigma(X_t)(W_{t+\Delta t} - W_t).$$

This discretized process is considered as a local approximation to the original process. Note that

$$\sigma(X_t)(W_{t+\Delta t} - W_t)$$

has normal distribution with mean zero and variance $\sigma^2(X_t)\Delta t$ and the transition density function of the discretized process is

$$p(X_{t+\Delta t}|X_t = x_t) = (2\pi\sigma^2(x_t)\Delta t)^{-1/2} \exp\{-\frac{(X_{t+\Delta t} - x_t - a(x_t, \theta)\Delta t)^2}{2\sigma^2(x_t)\Delta t}\}.$$

Suppose we observe the process $\{X_t, t \geq 0\}$ at the points $t + i\Delta t, 0 \leq i \leq n$. Let $Z_i = X_{t+i\Delta t}$. Then the joint probability density function of the random vector $(Z_0, \ldots, Z_N)$ is

$$p(z_0, \ldots, z_N) = \Pi_{i=1}^{N} p(z_i|z_{i-1})p(z_0)$$

and the parameters θ, σ can be estimated by the method of maximum likelihood.

Estimation based on local linearization method of Shoji-Ozaki

Consider the SDE

$$dx_t = a(X_t)dt + \sigma dW_t, t \geq 0.$$

Suppose the diffusion parameter σ is a constant and the drift function $a(.)$ is possibly nonlinear and differentiable. We try to approximate the above SDE by a linear SDE. Consider the ordinary differential equation

$$\frac{dx_t}{dt} = a(x_t).$$

Suppose the function x_t is differentiable twice with respect to t. Then

$$\frac{d^2x_t}{dt^2} = a'(x_t)\frac{dx_t}{dt}.$$

Suppose $a'(x)$ is constant over the interval $[t, t+\Delta t)$. Let $u \in [t, t+\Delta t)$. . Then

$$\frac{dx_t}{dt}|_{t=u} = \frac{dx_t}{dt}e^{a'(x_t)(u-t)}$$

and

$$x_{t+\Delta t} = x_t + \frac{a(x_t)}{a'(x_t)}[e^{a'(x_t)\Delta t} - 1].$$

Suppose we approximate the drift function $a(x)$ by a linear function Lx on $[t, t+\Delta t)$. Then we have the SDE

$$dX_t = LX_t dt + \sigma dW_t$$

where L is a constant on the interval $[t, t+\Delta t)$. Apllying Ito's lemma, we get that

$$X_{t+\Delta t} = X_t e^{L\Delta t} + \sigma \int_t^{t+\Delta t} e^{L(t+\Delta t-u)} dW_u. \qquad (*)$$

Let us choose L such that the conditional mean $E[X_{t+\Delta t}|X_t]$ coincides with the mean of the process given by (*). Hence

$$X_t e^{L\Delta t} = X_t + \frac{a(X_t)}{a'(X_t)}[e^{a'(X_t)\Delta t} - 1]$$

or

$$L = \frac{1}{\Delta t}\log[1 + \frac{a(X_t)}{X_t a'(X_t)}(e^{a'(X_t)\Delta t} - 1)].$$

Observe that the constant L depends on t. Denote it by L_t. The discretized process by the local linearization method is as follows:

$$X_{t+\Delta t} = X_t e^{L_t\Delta t} + \sigma \int_t^{t+\Delta t} e^{L_t(t+\Delta t-u)} dW_u.$$

Since the random variable

$$\int_t^{t+\Delta t} e^{L_t(t+\Delta t-u)} dW_u$$

has the normal distribution with mean zero and variance $\frac{e^{2L_t\Delta t}-1}{2L_t}$, we can now write the transition density function of the discretized observations $Y_i = X_{t+i\Delta t}$ given Y_{i-1} for $0 \le i \le N$ and compute the likelihood function. MLE of θ and σ can now be obtained.

Estimation via martingale estimating functions

Consider the SDE

$$dX_t = a(X_t,\theta)dt + \sigma(X_t,\theta)dW_t, X(0) = X_0, t \ge 0, \theta \in \Theta \subset R.$$

(i)Let us first consider the case when the diffusion function $\sigma(x,\theta)$ *does not depend* on the parameter θ. This is the case in all the earlier discussions on estimation of the drift parameter θ. If the process $\{X_s, 0 \le s \le t\}$ is observed completely, then the likelihood function $L_t(\theta)$ based on the observations is

$$\begin{aligned} \ell_t(\theta) &= \log L_t(\theta) \\ &= \int_0^t \frac{a(X_s,\theta)}{\sigma^2(X_s)} dX_s - \frac{1}{2}\int_0^t \frac{a^2(X_s,\theta)}{\sigma^2(X_s)} ds. \end{aligned}$$

Suppose now the process is observed at times $i\Delta, 0 \leq i \leq n$. We approximate the integrals in the above expression by Riemann-type sums to obtain an approximate log-likelihood function. It is given by

$$\tilde{\ell}_n(\theta) = \sum_{i=1}^{n} \frac{a(X_{(i-1)\Delta}, \theta)}{\sigma^2(X_{(i-1)\Delta})}(X_{i\Delta} - X_{(i-1)\Delta}) - \frac{1}{2}\sum_{i=1}^{n} \frac{a^2(X_{(i-1)\Delta}, \theta)}{\sigma^2(X_{(i-1)\Delta})}\Delta.$$

Suppose the function $a(x, \theta)$ is differentiable with respect to θ. Then

$$\tilde{\ell}'_n(\theta) = \sum_{i=1}^{n} \frac{a'(X_{(i-1)\Delta}, \theta)}{\sigma^2(X_{(i-1)\Delta})}(X_{i\Delta} - X_{(i-1)\Delta}) - \Delta\sum_{i=1}^{n} \frac{a(X_{(i-1)\Delta}, \theta)}{\sigma^2(X_{(i-1)\Delta})}a'(X_{(i-1)\Delta}, \theta).$$

The process $\{\tilde{\ell}'_n(\theta)\}$ is a zero mean martingale with respect to the fitration $\{\mathcal{F}_i\}$ with $\mathcal{F}_i$ generated by the set $\{X_0, X_\Delta, \ldots, X_{i\Delta}\}$. Solving the equation

$$\tilde{\ell}'_n(\theta) = 0,$$

which is called a martingale estimating equation, we can estimate the parameter θ.

Let us now consider the case when the diffusion $\sigma(x, \theta)$ depends on θ. Let us consider analogue of the function $\tilde{\ell}'_n(\theta)$ given by

$$J_n(\theta) = \sum_{i=1}^{n} \frac{a'(X_{(i-1)\Delta}, \theta)}{\sigma^2(X_{(i-1)\Delta}, \theta)}(X_{i\Delta} - X_{(i-1)\Delta}) - \Delta\sum_{i=1}^{n} \frac{a(X_{(i-1)\Delta}, \theta)}{\sigma^2(X_{i-1)\Delta}, \theta)}a'(X_{(i-1)\Delta}, \theta).$$

This process is *not* a martingale with respect to the filtration $\{\mathcal{F}_i\}$. Define

$$G_n(\theta) = J_n(\theta) - \sum_{i=1}^{n} E_\theta[J_i(\theta) - J_{i-1}(\theta)|\mathcal{F}_{i-1}].$$

The process $\{G_n(\theta)\}$ is a zero mean martingale with respect to the fitration $\{\mathcal{F}_i\}$ with $\mathcal{F}_i$ generated by the set $\{X_0, X_\Delta, \ldots, X_{i\Delta}\}$. Solving the martingale estimating equation

$$G_n(\theta) = 0$$

, we can estimate the parameter θ whether the function σ is a function of θ or otherwise.

The following list of references contains bibliographic details of some books and some important review papers published in the area of "Statistical Inference for Stochastic Processes" but are not cited in the text.

References:

Aalen, O.O. (1975) *Statistical Inference for a family of Counting Processes*, Ph.D. Thesis, University of California, Berkeley.

Aubry, C. (1997) *Estimation parametrique par la methode de la distance minimale pour des processus de Poisson et de diffusion.* Ph.D. Thesis, Universite du Maine, Le Mans.

Andersen, P.K., Borgan, Φ. , Gill, R,D. and Keiding, N. (1993) *Statistical Methods for Counting Processes*, Springer, New York.

Arato, M. (1982)*Linear Stochastic Systems with Constant Coefficients; a Statistical Approach*: Lecture Notes in Control and Information Sciences, **45**, Springer, Berlin.

Bar-Shalom, Y. (1971) On the asymptotic properties of maximum likelihood estimate obtained from dependent observations, *J. Roy. Statist. Soc. Ser. B*, **33**, 72-77.

Basawa, I.V. and Prabhu, N.U. (1994) *Statistical Inference in Stochastic Processes*, Special Issue of *Journal of Statistical Planning and Inference*, **39**, No. 2, pp. 135-352.

Basawa, I.V. and Prakasa Rao, B.L.S. (1980) "Statistical Inference for Stochastic Processes", Academic Press, London.

Basawa, I.V. and Prakasa Rao, B.L.S. (1980) Asymptotic inference for stochastic processes, *Stoch. Proc. Appl.*, **10**, 221-254.

Basawa, I.V. and Scott, D.J. (1983) *Asymptotic Optimal Inference for Non-ergodic Models*, Lecture Notes in Statistics, **17**, Springer, Heidelberg.

Baxter, G. (1956) A strong limit theorem for Gaussian processes, *Proc. Amer. Math. Soc.*, **7**, 522-527.

Bhat, B.R. (1974) On the method of maximum likelihood for dependent observations, *J. Roy. Statist. Soc. Ser. B*, **36**, 48-53.

Bhat, B.R. (1996) Tests based on estimating functions, In *Stochastic Processes and Statistical Inference*, Ed. B.L.S. Prakasa Rao and B.R. Bhat, New Age International, New Delhi, pp. 20-38.

Billingsley, P. (1961) *Statistical Inference for Markov processes*, University of Chicago Press, Chicago.

Bishwal, J.P.N. (2000) *Asymptotic Theory of Estimation of the Drift Parameter in Diffusion Processes*, Ph.D. Thesis, Sambalpur University, Sambalpur, India.

Borwanker, J.D. Kallianpur, G. and Prakasa Rao, B.L.S. (1971) The Bernstein-von Mises theorem for Markov Processes, *Ann. Math. Statist.*, **42**, 1241-1253.

Bose, A. and Politis, D.N. (1996) A review of the bootstrap for dependent samples, In *Stochastic Processes and Statistical Inference*, Ed. B.L.S. Prakasa Rao and B.R. Bhat, New Age International, New Delhi, pp. 73-89.

Brody, E. (1971) An elementary proof of the Gaussian dichotomy theorem, *Z. Wahrs.* , **20**, 217-226.

Brown, B.M. (1971) Martingale central limit theorems, *Ann. Math. Statist.*, **42**, 59-66.

Bosq, D. (1998) *Nonparametric Statistics for Stochastic Processes*, Lecture Notes in Statistics, **110**, Springer, New York.

Bosq, D. (2012) *Statistique Mathematique et Statistique des Processus*, hermes-science publications, Lavoisier, Cachan.

Brillinger, D. (1975) Statistical inference for stationary point processes, In *Stochastic Processes and Related Topics*, Ed. M.L. Puri, Academic Press, New York, pp. 55-99.

Cox, D.R. and Lewis, P.A.W. (1966) *The Statistical Analysis of Series of Events*, Methuen and Barnes and Nobel, New York.

Cressie, N. (1991) *Statistics of Spatial Data*, Wiley, New York.

Dalalyan, A. (2001) *Estimation non-parametrique pour les processus de diffusion ergodiques*, Ph.D. Thesis, Universite du Maine, Le Mans.

Dewan, Isha and Prakasa Rao, B.L.S. (2001) Associated sequences and related inference problems, In *Handbook of Statistics: Stochastic Processes;Theory and Methods*, **19**, Ed. D.N.Shanbhag and C.R.Rao , Elsevier Science B.V., Amsterdam, pp. 693-731.

Dion, J.-P. (1974) *Estimation des probabilites initiales et de la moyenne d'un processus de Galton-Watson*, Ph.D. Thesis, Universite de Montreal, Montreal.

Dion, J.-P. and Keiding, N. (1978) Statistical inference in branching processes, In *Branching Processes*, Ed. A. Joffe and P. Ney, Marcel Dekker, New York, pp. 105-140.

Doob, J.L. (1953) *Stochastic Processes*, Wiley, New York.

Dorogovcev, A. Ja. (1976) The consistency of an estimate of a parameter of stochastic differential equation, *Theory Prob. Math. Statist.*, **10**, 73-82.

Foutz, R. (1974) *Studies in Large sample Theory*, Ph.D. Thesis, Ohio State University, Columbus.

Feigin, P.D. (1975) *Maximum likelihood estimation for continuous time stochastic processes - A Martingale approach*, Ph.D. Thesis, Australian National University, Canberra.

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

Feldman, J. (1958) Equivalence and perpendicularity of Gaussian processes, *Pacific J. Math.*, **8** , 699-708, correction, ibid. **9**, 1295-1296.

Feller, W. (1971) *An Introduction to Probability Theory and its Applications*, Vol.II, 2nd ed., Wiley, New York.

Fleisher, I. and Kooharian, A. (1958) On the statistical treatment of stochastic processes, *Ann. Math. Statist.*, **29**, 544-549.

Fleming, T.R. and Harrington, D.P. (1991) *Counting Processes and Survival analysis*, Wiley, New York.

Gallant, A.R. and Tauchen, G. (1996) *A Unified Theory of Estimation and Inference for Nonlinear Dynamic Models*, Basil Blackwell, Oxford.

Girsanov, I.V. (1960) On transforming a certain class of stochastic processes by absolutely continuous substitution of measures, *Theory Prob. Appl.*, **5**, 285-301.

Grenander, U. (1950) Stochastic processes and statistical inference, *Arkiv. fur Mathematik*, **1**, 195-227.

Grenander, U. (1968) Eight lectures on statistical inference in stochastic processes, Tech. Report No.2, Division of Appl. Math., Brown University, Providence, Rhode Island.

Grenander, U. (1981) *Abstract Inference*, Wiley, New York.

Guttorp, P. (1991) *Statistical Inference for Branching Processes*, Wiley, New York.

Hall, P. and Heyde, C.C. (1980) *Martingale Limit Theory and its Application*, Academic Press, London.

Hajek , J. (1958) On a property of normal distribution of any stochastic processes, *Chech. Math. J.*, **8**, 610-618.

Heyde, C.C. (1974) On estimating the variance of the offspring distribution in a simple branching process, *Adv. Appl. Probab.*, **6**, 421-433.

Heyde, C.C. (1997) *Quasi-Likelihood and its Applications: A General Approach to Optimal Parameter estimation*, Springer, New York.

Ibragimov, I.A. (1963) A central limit theorem for a class of dependent random variables, *Theory Prob. Appl.*, **8**, 83-89.Brown,

Jacobsen, M. (1982) *Statistical Analysis of Counting Processes*, Lecture Notes in Statistics No. **12**, Springer, New York.

Kalman, R.E. (1960) A new approach to linear filtering and prediction theory, *J. Basic Engg.*, **82**, 35-45.

Kalman, R.E. and Bucy, R.S. (1961) New results in linear filtering and prediction theory, *J. Basic. Engg.*, **83**, 95-108.

Kakutani, S. (1948) On the equivalence of infinite product measures, *Ann. Math.*, **49**, 214-224.b

Karhunen, K. (1947) Uber lineare methoden in der wahrscheinlichkeitsrechnung, *Ann. Acad. Sci. Finn. a1*, **37**, 1-79.

Karr, A. (1991) *Point Processes and their Statistical Inference*, Marcel Dekker, New York.

Kasonga, R. (1988) The consistency of a nonlinear least squares estimator for diffusion processes, *Stoch. Proc. Appl.* , **30**, 263-275.

Klimko, L.A. and Nelson, P.I. (1978) On conditional least squares estimation for stochastic processes, *Ann. Statist.*, **6**, 629-642.

Krickeberg, K. (1980) *Statistical Problems on Point Processes*, Banach Center Publications No. **6**, pp.197-223.

Krickeberg, K. (1982) Processus ponctuels en statistique, In *Lecture Notes Math.* , **929**, Springer, Berlin, pp. 205-313.

Kuchler, U.and Sorensen, M. (1997) *Exponential Families of Stochastic Processes*, Springer, New York.

Kunita, H. and Watanabe, S. (1967) On square integrable martingales, *Nagoya Math. J.*, **30**, 209-245.

Kutoyants, Yu. A. (1984) *Parameter Estimation for Stochastic Processes*, Translated and Ed. B.L.S. Prakasa Rao, Heldermann, Berlin.

Kutoyants, Yu. A. (1994) *Identification of Dynamical Systems with Small Noise*, Kluwer, Dordrecht.

Kutoyants, Yu. A. (1998) *Statistical Inference for Spatial Poisson Processes*, Lecture Notes in Statistics, **134**,Springer, New York.

Kutoyants, Yu. A. (2004) *Statistical Inference for Ergodic Diffusion Processes*, Springer, London.

Le Breton, A. (1976) On continuous and discrete sampling for parameter estimation in diffusion type process, In *Mathematical Programming Studies*, **5**, 124-144.

Lewis, P.A.W. (1972) *Stochastic Point Processes: Statistical Analysis, Theory and Applications*, Wiley, New York.

Linkov, Y.N. (2001) *Asymptotic Methods in the Statistics of Stochastic Processes*, American Mathematical Society, Providence, Rhode Island.

Liptser, R.S. and Shiryayev, A.N. (1977) *Statistics of Random Processes: General Theory*, Springer, New York.

Liptser, R.S. and Shiryayev, A.N. (1978) *Statistics of Random Processes: Applications*, Springer, New York.

Loeve, M. (1946) Fonctions aleatoires de second ordre, *C.R. Acad. Sci. Paris*, **222**.

Loeve, M. (1977) *Probability Theory I*, 4th ed., Springer, Berlin.

Maruyama, G. and Tanaka, H. (1957) Some properties of one-dimensional diffusion processes, *Mem. Fac. Kyushu Univ.*, **11**, 117-141.

Meyer, P. (1962) A decomposition theorem for supermartingales, *Illinois J. Math.*, **6**, 193-205.

Naik-Nimbalkar, U.V. (1996) In *Stochastic Processes and Statistical Inference*, Ed. B.L.S. Prakasa Rao and B.R. Bhat, New Age International, New Delhi, pp.52-72.

Negri, I. (1998) *Efficacite globale de la fonction de repartition empirique dans le cas d'un processus de diffusion ergodique*, Ph.D. Thesis, Universite du Maine, Le Mans.

Norman, M.F. (1971) Statistical inference with dependent observations. Extensions of classical procedures, :*J. Mathematical Psychology*, **8**, 444-451.

Novikov, A.A. and Shiryayev, A.N. (1994) *Statistics and control of random processes*, Proceedings of Steklov Institute of Mathematics, **202**, Amer. Math. Soc. Providence, Rhode Island, USA.

Prabhu, N.U. (1988)*Statistical Inference from Stochastic Processes*, Contemporary Mathematics, **80**, American Mathematical Society, Providence, Rhode Island.

Prabhu. N.U. and Basawa, I.V. (1991) *Statistical Inference in Stochastic Processes*, Marcel Dekker, New York.

Prakasa Rao, B.L.S. (1972) Maximum likelihood estimation for Markov processes, *Ann. Inst. Statist. Math*, **24**, 333-345.

Prakasa Rao, B.L.S. (1974) Statistical inference for stochastic processes, Tech. Report CRM-465, Centre de Recherches Mathematiques, Universite de Montreal.

Prakasa Rao, B.L.S. (1983) Asymptotic theory for nonlinear least squares estimator for diffusion processes, *Math. Oper. Stat. Series Statistik*, **14**, 195-209.

Prakasa Rao, B.L.S. (1987) *Asymptotic Theory of Statistical Inference*, Wiley, New York.

Prakasa Rao, B.L.S. (1988) Statistical inference from sampled data for stochastic processes, In *Contemporary Mathematics*, **80**, American Mathematical Society, Providence, Rhode Island, pp. 249-284.

Prakasa Rao, B.L.S. (1990) Nonparametric density estimation for stochastic processes from sampled data, *Publ. Inst. Stat. Univ. de Paris.*, **35**, 51-83.

Prakasa Rao, B.L.S. (1991) Asymptotic theory of weighted maximum likelihood estimation for growth models, In *Statistical Inference in Stochastic Processes*, Ed. N.U. Prabhu and I.V. Basawa, Marcel Dekker, New York, pp. 183-208.

Prakasa Rao, B.L.S. (1996) Optimal asymptotic tests of composite hypotheses for continuous time stochastic processes, *Sankhya Ser. A*, **58**, 8-24.

Prakasa Rao, B.L.S. (1996) Nonparametric approach to time series analysis, In *Stochastic Processes and Statistical Inference*, Ed. B.L.S. Prakasa Rao and B.R. Bhat, New Age International, New Delhi, pp.73-89.

Prakasa Rao, B.L.S. (1999) *Statistical Inference for Diffusion Type Processes*, Kendall's Library of Statistics No.8, Arnold, London and Oxford University Press, New York.

Prakasa Rao, B.L.S. (1999) *Semimartingales and their Statistical Inference*, Chapman and Hall/ CRC Press, Boca Raton., Florida.

Prakasa Rao, B.L.S. (2001) Statistical inference for stochastic partial differential equations, In *Selected Proceedings of the Symposium on Inference for Stochastic Processes*, Ed. I.V. Basawa, C.C.Heyde, and R.L.Taylor, IMS Monograph Series, **37**, pp.47-70.

Prakasa Rao, B.L.S. (2001) Nonparametric inference for parabolic stochastic partial differential equations, *Random Operators and Stochastic Equations*, **9**, 329-338.

Prakasa Rao, B.L.S. (2002) Nonparametric inference for a class of stochastic partial differential equations based on discrete observations, *Sankhya Ser.A*, **64**, 1-15.

Prakasa Rao, B.L.S. (2002) On some problems of estimation for some stochastic partial differential equations, In *Uncertainty and Optimality*, Ed. J.C.Misra (2002) World Scientific, Singapore, pp. 71-154.

Prakasa Rao, B.L.S. (2003) Parametric estimation for linear stochastic differential equations driven by fractional Brownian motion, *Random Operators and Stochastic Equations*, **11**, 229-242.

Prakasa Rao, B.L.S. (2004) Self-similar processes, fractional Brownian motion and statistical inference, In *Festschrift for Herman Rubin* , Ed. A. Das Gupta, Institute of Mathematical Statistics, Lecture Notes and Monograph Series, **45**, 98-125.

Prakasa Rao, B.L.S. (2009) Conditional independence, conditional mixing and conditional association, *Ann. Inst. Statist. Math.*, **61**, pp. 441-460.

Prakasa Rao, B.L.S. (2010) *Statistical Inference for Fractional Diffusion Processes*, Wiley, London.

Prakasa Rao, B.L.S. (2012) *Associated Sequences, Demimartingales and Nonparametric Inference*, Birkhauser, Springer, Basel.

Prakasa Rao, B.L.S. and Bhat, B.R. (1996) *Stochastic Processes and Statistical Inference*, New Age International, New Delhi.

Prakasa Rao, B.L.S. and Prasad, M.S. (1976) Maximum likelihood estimation for dependent random variables, *J. Indian Statist. Assoc.*, **14**, 75-79.

Prasad, M. S. (1971) *Some Contribution to the theory of Maximum likelihood Estimation for Dependent Random Variables*, Ph.D. Thesis, Indian Institute of Technology, Kanpur.

Rajarshi, M.B. (1996) Resampling methods for stochastic processes, In *Stochastic Processes and Statistical Inference*, Ed. B.L.S. Prakasa Rao and B.R. Bhat, New Age International, New Delhi, pp.90-120.

Rao, M.M. (2000) *Stochastic Processes: Inference Theory*, Kluwer, Dordrecht.

Renyi, A. (1963) On stable sequences of events, *Sankhya Series A*, **25**, 293-302.

Revesz, P. 91968) *The Laws of Large Numbers*, Academic Press, New York.

Rippley, B.D. (1988) *Statistical Inference for Spatial Point Processes*, Cambridge University Press, Cambridge, UK.

Sagdar, D. (1974) On an approximate test of hypotheses about the correlation function of a Gaussian random process, *Theor. Probab. Math. Stat.*, **2**, 231-238.

Sarma, Y.R. (1976) Sur les tests et sur l'estimation de parametres pour certains processus stochastiques stationnaires, *Publ. Inst. Statist. Univ. Paris*, **17**, 1-124.

Sagirow, P. (1970) *Stochastic Methods in the Dynamics of Satellites*, CISM Courses and Lectures, **57**, Springer, Berlin.

Schmetterer, L. (1974) *Introduction to Mathematical Statistics*, Springer, Berlin.

Silvey, S.D. (1961) A note on the maximum likelihood in the case of dependent observations, *J. Roy. Statist. Soc. ser. B*, **23**, 444-452.

Striebel, C.T. (1959) Densities for stochastic processes, *Ann. Math. Statist.*,**30**, 559-567.

Swensen , A. (1980) *Asymptotic Inference for a Class of Stochastic Processes*, Ph.D. Thesis, University of California, Berkeley.

Wald. A. (1948) Asymptotic properties of the maximum likelihood estimate of an unknown parameter of a discrete stochastic process, *Ann. Math. Statist.*, **19**, 40-46.

Winnicki, J. (1988) Estimation theory for the branching process with immigration, In Contemporary Mathematics, **80**, pp. 301-322.

Woerner, J. (2001) *Statistical Analysis for Discretely Observed Levy Process*, Ph.D. Thesis, Albert-Ludwig-Universitat, Freiburg.

Yanev, N.M. (1975) On the statistics of branching processes, *Theory Prob. Appl.*, **20**, 612-622.